Institute of Mathematical Statistics

LECTURE NOTES – MONOGRAPH SERIES

Inequalities in Statistics and Probability

Edited by
Y. L. Tong

Volume 5

Shanti S. Gupta, Series Editor

Institute of Mathematical Statistics

LECTURE NOTES - MONOGRAPH SERIES
Shanti S. Gupta, Series Editor

Volume 5

Inequalities in Statistics and Probability

Proceedings of the Symposium on Inequalities in Statistics and Probability, October 27-30, 1982, Lincoln, Nebraska

Edited by
Y. L. Tong
University of Nebraska-Lincoln

With the cooperation of
Ingram Olkin, *Stanford University*
Michael D. Perlman, *University of Washington*
Frank Proschan, *Florida State University*
C.R. Rao, *University of Pittsburgh*

Institute of Mathematical Statistics
Hayward, California

Institute of Mathematical Statistics
Lecture Notes – Monograph Series

Series Editor, Shanti S. Gupta, Purdue University

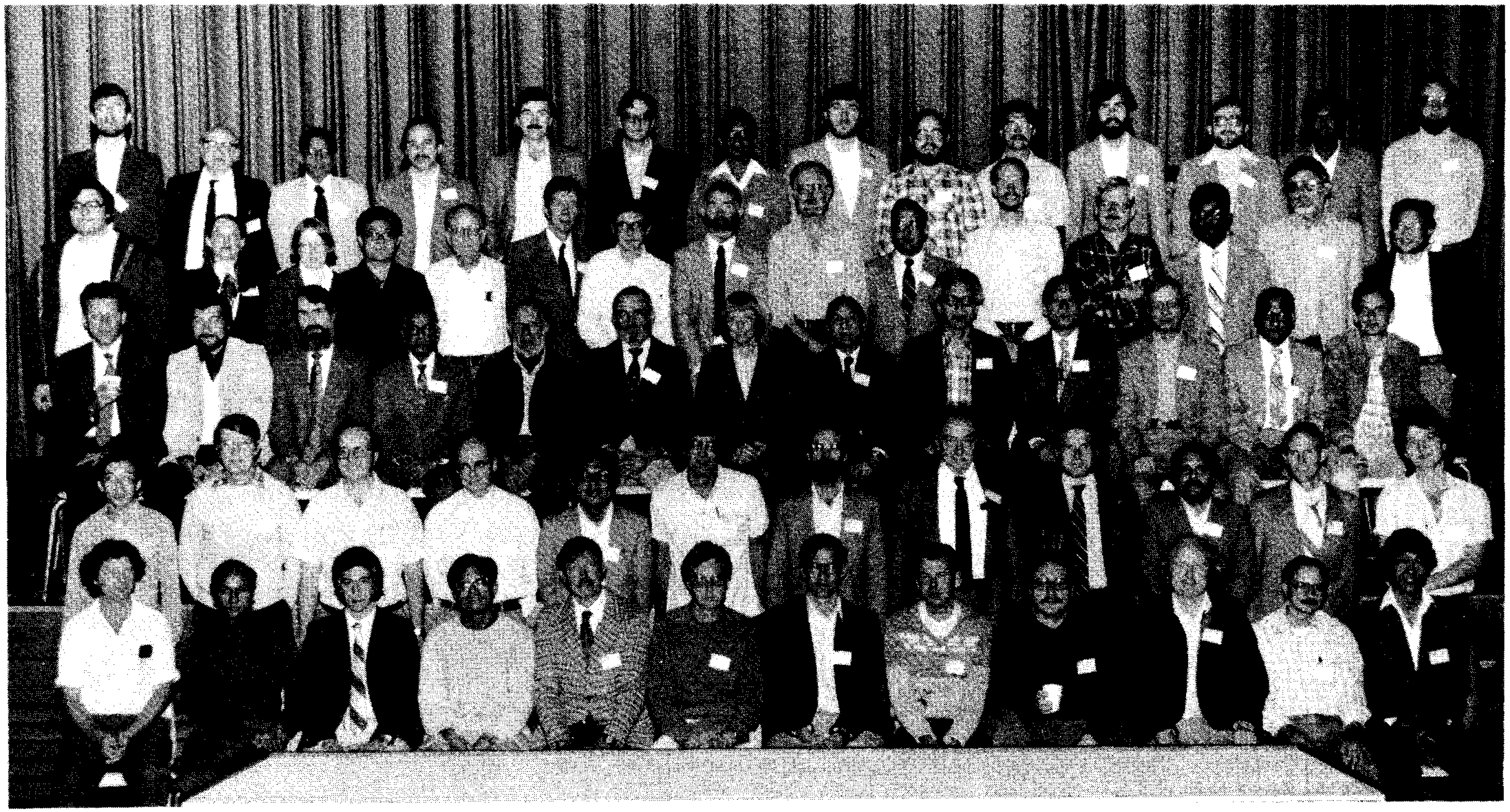
The production of the IMS Lecture Notes — Monograph Series is managed by the IMS Business Office: Bruce E. Trumbo, Treasurer, and Jose L. Gonzalez, Business Manager.

Library of Congress Cataloging in Publication Data

Symposium on Inequalities in Statistics and Probability
(1982 : Lincoln, Neb.)
Inequalities in statistics and probability.

(Lecture notes-monograph series / Institute of Mathematical Statistics ; v. 5)
Includes bibliographies.
1. Mathematical statistics—Congresses. 2. Probabilities—Congresses.
3. Inequalities (Mathematics)—Congresses. I. Tong, Y. L. (Yung Liang), 1935-
II. Title. III. Series: Lecture notes-monograph series ; v. 5.
QA276.A1S87 1982 519.5 83-12975
ISBN 0-940600-04-8

AMS 1980 subject classifications. 60E15, 26D15, 60E99, 62E99



Symposium on Inequalities in Statistics and Probability

Lincoln, Nebraska · October 27–30, 1982

PREFACE

The Symposium on Inequalities in Statistics and Probability was held at the University of Nebraska-Lincoln during October 27–30, 1982. It was sponsored by the National Science Foundation, the Office of Naval Research, and the University of Nebraska-Lincoln. The organizer of the Symposium was Y. L. Tong, and the Advisory Board consisted of Ingram Olkin, Michael D. Perlman, Frank Proschan, and C. R. Rao. The general study of inequalities has experienced a remarkable growth during the past decade, and is currently expanding at an even faster rate. In view of this growth, the purpose of the Symposium was to provide an opportunity for research workers in inequalities-related areas to meet together, to exchange and share ideas, and to present their latest developments.

There were 15 sessions consisting of 28 invited papers and 7 contributed papers. The topics of the sessions were: (a) Matrix Related Inequalities, (b) Multivariate Majorization and Dilations, (c) Stochastic Optimization and Rearrangement Inequalities, (d) Probabilities of Geometric Regions, (e) Moment and Markov Inequalities, (f) Perspectives on Inequalities and Entropy Functions, (g) Association and FKG Inequalities, (h) Inequalities for Selecting and Ordering Populations, (i) Trends and Order Restrictions, (j) Inequalities in Multivariate Analysis (two sessions), (k) Convex and Stochastic Orderings, (l) Inequalities in Reliability, (m) Contributed Papers (two sessions). In addition there was a general problem session. More than 70 research workers attended and participated in the conference.

Almost all of the papers included in this volume were presented at the conference. The contents of a few papers differ slightly from the conference presentations due to arrangements already made for prior publications. The original Symposium program can be found at the end of this volume.

ACKNOWLEDGMENTS

I would like to gratefully acknowledge the support of the National Science Foundation, the Office of Naval Research, and the University of Nebraska-Lincoln for sponsoring the Symposium on Inequalities in Statistics and Probability. I thank the Institute of Mathematical Statistics for publishing this volume. The typesetting work was done by Pied Typer, and was supported by the University of Nebraska-Lincoln. My thanks are also extended to the Symposium participants, and the authors and referees of the papers; it is their efforts and noble contributions which have made this volume a valuable addition.

I wish to express my special gratitude to Professors Ingram Olkin, who originally suggested the Symposium, Michael D. Perlman, Frank Proschan, and C. R. Rao. Without their continuing support, cooperation, and encouragement this volume would not have become a reality.

Y. L. Tong

CONTRIBUTORS

Roger L. Berger, Department of Statistics, North Carolina State University, Box 8203, Raleigh, NC 27695

John R. Birge, Department of Industrial and Operations Engineering, University of Michigan, Ann Arbor, MI 48109

Henry W. Block, Department of Mathematics and Statistics, University of Pittsburgh, Pittsburgh, PA 15260

R. Bohrer, Department of Mathematics, University of Illinois, Urbana, IL 61801

Pinyuen Chen, Department of Mathematics, Syracuse University, Syracruse, NY 13210

Joel E. Cohen, Box 20, The Rockefeller University, 1230 York Avenue, New York, NY 10021

David C. Cox, Battelle Memorial Institute, Washington Operations, 2030 M Street N.W., Washington, D.C. 20036

Catherine D'Abadie, AT&T Technologies, Inc., Gateway Two, Newark, NJ 07102

Somesh Das Gupta, Indian Statistical Institute, 203 B.T. Road, Calcutta, 700 035, India

S. W. Dharmadhikari, Department of Mathematics, Southern Illinois University, Carbondale, IL 62901

Richard L. Dykstra, Department of Statistics and Actuarial Science, University of Iowa, Iowa City, IA 52242

Morris L. Eaton, School of Statistics, University of Minnesota, Minneapolis, MN 55455

A. M. Fink, Department of Mathematics, Iowa State University, Ames, IA 50011

Marshall Freimer, The Graduate School of Management, University of Rochester, Rochester, NY 14627

Shanti S. Gupta, Department of Statistics, Purdue University, West Lafayette, IN 47907

Deng-Yuan Huang, Institute of Mathematics, National Taiwan Normal University, Taipei, Taiwan, China

D. R. Jensen, Department of Statistics, Virginia Polytechnic Institute and State University, Blacksburg, VA 24061

Kumar Joag-Dev, Department of Mathematics, University of Illinois, Urbana, IL 61801

Max Jodeit, Jr., School of Mathematics, University of Minnesota, Minneapolis, MN 55455

Samuel Karlin, Department of Mathematics, Stanford University, Stanford, CA 94305

J. H. B. Kemperman, Department of Mathematics, University of Rochester, Rochester, NY 14627

Tze Leung Lai, Department of Mathematical Statistics, Columbia University, New York, NY 10027

Rhonda Magel, Division of Mathematical Sciences, North Dakota State University, Fargo, ND 58105

CONTRIBUTORS

Albert W. Marshall, Department of Statistics, University of British Columbia, Vancouver, British Columbia, Canada V6T 1W5

Govind S. Mudholkar, Department of Statistics, University of Rochester, Rochester, NY 14627

Charles M. Newman, Department of Mathematics, University of Arizona, Tucson, AZ 85721

S. Panchapakesan, Department of Mathematics, Southern Illinois University, Carbondale, IL 62901

Michael D. Perlman, Department of Statistics, University of Washington, Seattle, WA 98195

Frank Proschan, Department of Statistics, Florida State University, Tallahassee, FL 32306

C. Radhakrishna Rao, Department of Mathematics and Statistics, University of Pittsburgh, Pittsburgh, PA 15260

Yosef Rinott, Department of Statistics, The Hebrew University of Jerusalem, Jerusalem 91905, Israel

Tim Robertson, Department of Statistics and Actuarial Science, University of Iowa, Iowa City, IA 52242

S. K. Sarkar, Department of Statistics, Temple University, Philadelphia, PA 19122

Thomas H. Savits, Department of Mathematics and Statistics, University of Pittsburgh, Pittsburgh, PA 15260

Moshe Shaked, Department of Mathematics, University of Arizona, Tucson, AZ 85721

L. A. Shepp, Bell Laboratories, Murray Hill, NJ 07974

Milton Sobel, Department of Mathematics, University of California at Santa Barbara, Santa Barbara, CA 93106

Wagner de Souza Borges, Instituto de Mathematica e Estatistica, Universidade de São Paulo, caixa postal 20570 ag. Iguatemi, 05508 São Paulo, SP, Brazil

Y. L. Tong, Department of Mathematics and Statistics, University of Nebraska, Lincoln, NE 68588

Richard A. Vitale, Department of Mathematics, Claremont Graduate School, Claremont, CA 91711

Ching Zong Wei, Department of Mathematics, University of Maryland, College Park, MD 20742

Roger J.-B. Wets, Department of Mathematics, University of Kentucky, Lexington, KY 40546

F. T. Wright, Department of Mathematics and Statistics, University of Missouri-Rolla, Rolla, MO 65401

H. P. Wynn, Department of Mathematics, Imperial College, Queen's Gate, London SW7 2BZ, England

CONTENTS

CONTENTS

Inequalities in Statistics and Probability
IMS Lecture Notes–Monograph Series Vol. 5 (1984), 1–3

INTRODUCTION[1]

BY Y. L. TONG
University of Nebraska

As noted by Pólya (1967), "Inequalities play a role in most branches of mathematics and have widely different applications." This is certainly true in statistics and probability. From the viewpoint of applications, inequalities have become a useful tool in estimation and hypothesis-testing problems (such as for yielding bounds on the variances of estimators and on the probability contents of confidence regions, and for establishing monotonicity properties of the power of certain tests), in multivariate analysis, in reliability theory, and so forth. Perhaps the usefulness of inequalities can be best illustrated by the following situation in reliability theory: Under certain circumstances it is desirable or necessary to determine whether or not the reliability of a system meets a given specification. The evaluation of the true reliability of a complex system is not always feasible. But if an inequality in the form of a lower bound on the system reliability can be easily obtained, and if the lower bound already meets the specification, then one knows for sure that the system meets or exceeds the specification.

On the other hand, the theory of inequalities in statistics and probability has intrinsic interest and importance and need not rely only on applications. For deriving such inequalities one usually studies a problem from several different approaches, such as monotonicity properties via concepts of stochastic ordering of random variables and distributions, positive or negative dependence and/or association properties via a mixture of distributions or monotone transformations of random variables, Schur concavity and the notion of majorization via the diversity of the components of a vector (or matrix), and so on. Thus the study of inequalities *per se* can provide a better understanding of the interrelationships among the random variables and their transformations, and may reveal new information concerning the complicated structure of probability distributions. This in turn provides new insights, ideas, and approaches for solving a variety of problems in statistics and probability.

The general study of the theory of inequalities in statistics and probability is, of course, closely related to the developments of inequalities in mathematics. As Mitrinović pointed out (1970, p. v), although "the theory of inequalities (in mathematics) began its development from (the days of) C. F. Gauss, A. L. Cauchy, and P. L. Cebysev," it is "the classical work *Inequalities* by G. H. Hardy, J. E. Littlewood and G. Pólya (1934, 1952) . . . which transformed the field of inequalities from a collection of isolated formulas into a systematic discipline." After the publication of the second editon of their book in 1952, there have been several other volumes on mathematical inequalities, such

---

[1] This research was partially supported by NSF Grant MCS81–00775, A01.

AMS 1980 subject classifications. 60E15, 26D15, 62H99.
Key words and phrases: Inequalities in statistics and probability.

as Beckenbach and Bellman (1965) and Mitrinović (1970). The latest addition, the book by Marshall and Olkin (1979), contains an up-to-date treatment of the theory of majorization inequalities and its applications in linear algebra, geometry, as well as statistics and probability, and is also highly influenced by Hardy, Littlewood and Pólya (1934, 1952). Among the conference proceedings, there have been three volumes edited by Shisha (1967, 1970, 1972).

Among the books and monographs related to inequalities in statistics and probability, some chapters of the two volumes by Karlin and Studden (1966) and Karlin (1968) involve such inequalities, mainly for density functions which are totally positive. The book by Barlow and Proschan (1975) contains probability inequalities and their applications in reliability theory. The monograph by Marshall and Olkin (1979) concerns mainly inequalities in statistics and probability, and partial orderings for probability distributions, via the theory of majorization. The book by Tong (1980) deals with probability inequalities in multivariate distributions via dependence, association, and mixture of random variables and distributions, via monotonicity and diversity of the parameter vectors and other related concepts, and includes statistical applications. Each of these books contains a complete bibliography which gives sources for the research papers published in this area. Among the latest review articles, the review written by Kemperman (1981) contains a detailed description of majorization and the concept of dilation of probability measures; the paper by Eaton (1982) gives an up-to-date review of certain types of probability inequalities, and contains a few references which appeared after the publications of Marshall and Olkin (1979) and Tong (1980).

The present volume represents the most recent developments in the area of inequalities in statistics and probability. It contains 30 research and expository papers which have been grouped according to the following topics:

(a) Inequalities Via Partial Orderings: D'Abadie and Proschan (pp. 4–12), Eaton (pp. 13–25), Jensen (pp. 26–34), Karlin and Rinott (pp. 35–40).

(b) Convex and Matrix-Related Inequalities: Cohen (pp. 41–53), Das Gupta and Sarkar (pp. 54–58), Freimer and Mudholkar (pp. 59–67), Rao (pp. 68–77).

(c) Probabilistic and Distribution-Free Inequalities: Cox (pp. 78–83), Kemperman (pp. 84–103), Marshall (pp. 104–108), Vitale (pp. 109–111), Vitale (pp. 112–114).

(d) Dependence-Related Inequalities: Fink and Jodeit (pp. 115–120), Joag-Dev, Shepp and Vitale (pp. 121–126), Newman (pp. 127–140), Shaked and Tong (pp. 141–149).

(e) Inequalities in Regression and Multivariate Analysis: Bohrer and Wynn (pp. 150–155), Dharmadhikari and Joag-Dev (pp. 156–164), Lai and Wei (pp. 165–172), Perlman (pp. 173–177).

(f) Inequalities in Stochastic Optimization and Reliability: Birge and Wets (pp. 178–186), Block and Souza Borges (pp. 187–192), Savits (pp. 193–198).

(g) Inequalities in Selecting and Ordering Populations: Berger and Proschan (pp. 199–205), Chen and Sobel (pp. 206–210), Gupta, Huang and Panchapakesan (pp. 211–227).

(h) Trends and Order Restrictions: Dykstra (pp. 228–235), Magel and Wright (pp. 236–243), Robertson and Wright (pp. 244–250).

REFERENCES

BARLOW, R. E. and PROSCHAN, F. (1975). *Statistical Theory of Reliability and Life Testing*. Holt, Rinehart and Winston, New York.

BECKENBACH, E. F. and BELLMAN, R. (1965). *Inequalities*. Springer-Verlag, Berlin and New York.

EATON, M. L. (1982). A review of selected topics in multivariate probability inequalities. *Ann. Statist. 10* 11–43.

HARDY, G. H., LITTLEWOOD, J. E., and PÓLYA, G. (1934, 1952). *Inequalities,* 1st ed., 2nd ed. Cambridge University Press, London and New York.

KARLIN, S. (1968). *Total Positivity, Vol. 1*. Stanford University Press, Stanford, CA.

KARLIN, S. and STUDDEN, W. J. (1966). *Tchebycheff Systems: With Applications in Analysis and Statistics*. Interscience, New York.

KEMPERMAN, J. H. B. (1981). Book review of Inequalities: Theory of majorization and its applications by Albert W. Marshall and Ingram Olkin; Probability inequalities in multivariate distributions by Y. L. Tong. *Bull. (New Ser.) Amer. Math. Soc. 5* 319–324.

MARSHALL, A. W. and OLKIN, I. (1979). *Inequalities: Theory of Majorization and Its Applications*. Academic Press, New York.

MITRINOVIĆ, D. S. (1970). *Analytic Inequalities*. Springer-Verlag, Berlin and New York.

PÓLYA, G. (1967). Inequalities and the principle of nonsufficient reason. in *Inequalities,* O. Shisha ed., p. 1, Academic Press, New York.

SHISHA, O. (ed.) (1967, 1970, 1972). *Inequalities, Inequalities II, III*. Academic Press, New York.

TONG, Y. L. (1980). *Probability Inequalities in Multivariate Distributions*. Academic Press, New York.

Inequalities in Statistics and Probability
IMS Lecture Notes–Monograph Series Vol. 5 (1984), 4–12

STOCHASTIC VERSIONS OF REARRANGEMENT INEQUALITIES[1]

BY CATHERINE D'ABADIE and FRANK PROSCHAN
AT&T Technologies, Inc. and Florida State University

This paper develops a unified way of obtaining stochastic versions of deterministic rearrangement inequalities. Rearrangement inequalities compare the value of a function of vector arguments with the value of the same function after the components of the vectors have been rearranged. The classical example of a rearrangement inequality is the well-known inequality of Hardy, Littlewood, and Pólya for sums of products. They show that if $a_1 \geqslant \dots \geqslant a_n$ and $b_1 \geqslant \dots \geqslant b_n$ are positive numbers, then for every permutation $(\pi(1), \dots, \pi(n))$ of $(1, \dots, n)$ the inequalities $\Sigma_{i=1}^n a_i b_i \geqslant \Sigma_{i=1}^n a_i b_{\pi(i)} \geqslant \Sigma_{i=1}^n a_i b_{n-i+1}$ hold.

The function Σxy is an example from a class of functions called arrangement increasing functions for which such rearrangement inequalities hold. Given two nonnegative random vectors $\mathbf{X}$ and $\mathbf{Y}$ with joint density $f(\mathbf{x},\mathbf{y})$ we determine conditions on f for the stochastic rearrangment inequalities $g(X_1, \dots, X_n; Y_1, \dots, Y_n) \overset{st}{\geqslant} g(X_1, \dots, X_n; Y_{\pi(1)}, \dots, Y_{\pi(n)}) \overset{st}{\geqslant} g(X_1, \dots, X_n; Y_n, \dots, Y_1)$ to hold for every permutation π and arrangement increasing function g. We present a number of examples of densities which satisfy the condition.

1. Introduction. The development of stochastic versions of deterministic concepts arising in mathematics has, in the past, led to important new results in probabiliy and statistics. The subject of this paper is in this spirit.

Specifically, we obtain *stochastic versions of rearrangement inequalities*. Rearrangement inequalities compare the value of a function of vector arguments with the value of the same function after the components of the vectors have been rearranged.

The classical example of a rearrangement inequality involving a function of two vector arguments is the well-known inequality of Hardy, Littlewood, and Pólya (1952) for sums of products. For vectors $\mathbf{a} = (a_1, \dots, a_n)$ and $\mathbf{b} = (b_1, \dots, b_n)$ of positive numbers, Hardy, Littlewood, and Pólya show that the function $f(\mathbf{a},\mathbf{b}) = \Sigma_{i=1}^n a_i b_i$ takes its largest value when the components of each $\mathbf{a}$ and $\mathbf{b}$ are arranged in increasing (or, equivalently, decreasing) order, and that f takes its smallest value when the components of one of the vectors are arranged in increasing order and those of the other vector are arranged in decreasing order. In symbols they show that if $a_1 \geqslant a_2 \geqslant \dots \geqslant a_n$ and $b_1 \geqslant b_2 \geqslant \dots \geqslant b_n$ (after relabelling, say) for every permutation $(\pi(1), \dots, \pi(n))$ of $(1, \dots, n)$, then the inequalities:

$$\Sigma_{i=1}^n a_i b_i \geqslant \Sigma_{i=1}^n a_i b_{\pi(i)} \geqslant \Sigma_{i=1}^n a_i b_{n-i+1} \tag{1.1}$$

hold.

The original idea which motivated this work was to obtain a stochastic version of the inequalities in (1.1). More explicitly, given two random vectors $\mathbf{X} \equiv (X_1, \dots, X_n)$ and $\mathbf{Y} \equiv (Y_1, \dots, Y_n)$, we wished to determine conditions to impose on $\mathbf{X}$ and $\mathbf{Y}$ to yield for every permutation π the stochastic inequalities:

$$\Sigma_{i=1}^n X_i Y_i \overset{st}{\geqslant} \Sigma_{i=1}^n X_i Y_{\pi(i)} \overset{st}{\geqslant} \Sigma_{i=1}^n X_i Y_{n-i+1}, \tag{1.2}$$

where $X \overset{st}{\geqslant} Y$ means $P(X \geqslant t) \geqslant P(Y \geqslant t)$ for all t.

[1] The research of both authors was supported under Contract AFOSR 82–K–0007 with the Air Force Office of Scientific Research, AFSC, U.S.A.F.

AMS 1980 subject classifications. 60C05.

Key words and phrases: Rearrangement inequalities, stochastic rearrangement inequalities, arrangement increasing, Hardy-Littlewood-Pólya inequality, permutations, preservation of properties, positive set function, monotone power function.

As it happens, it is not hard to obtain such a stochastic version of the Hardy–Littlewood–Pólya inequality. Sufficient conditions on $\mathbf{X}$ and $\mathbf{Y}$ for (1.2) to hold can be easily stated. Suppose that $\mathbf{X}$ and $\mathbf{Y}$ are nonnegative random vectors having a joint density $f(\mathbf{x},\mathbf{y})$. For a given vector $\mathbf{x}$ we write $\mathbf{x} \geq^{t_{ij}} \mathbf{x}'$ if $i < j$, $x_i \leq x_j$, and $\mathbf{x}'$ is obtained from $\mathbf{x}$ by interchanging x_i and x_j and leaving the other components fixed. Then inequality (1.2) holds for $\mathbf{X}$ and $\mathbf{Y}$ if for all pairs i, j, $1 \leq i < j \leq n$, f satisfies:

$$f(\mathbf{x},\mathbf{y}) + f(\mathbf{x}',\mathbf{y}') - f(\mathbf{x}',\mathbf{y}) - f(\mathbf{x},\mathbf{y}') \geq 0, \tag{1.3}$$

where $\mathbf{x} \geq^{t_{ij}} \mathbf{x}'$ and $\mathbf{y} \geq^{t_{ij}} \mathbf{y}'$.

Since the work of Hardy, Littlewood, and Pólya (1952), papers on inequalities involving rearrangements of vectors in $\mathscr{R}^n$ have appeared widely in the literature. Marshall and Olkin (1979) present a unified approach to the study of deterministic rearrangement inequalities.

We develop a theory which offers a *unified approach* to the task of obtaining stochastic versions of rearrangement inequalities. Our work generalizes that of previous authors in that we obtain their deterministic inequalities as special cases. In this paper *we present an overview of the theory we develop and some applications to statistics.*

2. Deterministic Rearrangement Inequalities. A deterministic rearrangement inequality compares the value of a function of vector arguments with the value of the same function after the components of the vectors have been rearranged. In the case of two vectors, rearrangement inequalities have the form

$$f(\overset{\uparrow}{\mathbf{x}},\overset{\uparrow}{\mathbf{y}}) = f(\overset{\downarrow}{\mathbf{x}},\overset{\downarrow}{\mathbf{y}}) \geq f(\mathbf{x},\mathbf{y}) \geq f(\overset{\uparrow}{\mathbf{x}},\overset{\downarrow}{\mathbf{y}}) = f(\overset{\downarrow}{\mathbf{x}},\overset{\uparrow}{\mathbf{y}}), \tag{2.1}$$

where $\overset{\uparrow}{\mathbf{z}}$ ($\overset{\downarrow}{\mathbf{z}}$) denotes the vector whose components are those of $\mathbf{z}$ arranged in increasing (decreasing) order.

The classical rearrangement inequality is the inequality of Hardy, Littlewood, and Pólya (1952) where $f(\mathbf{x},\mathbf{y}) = \Sigma x_i y_i$. As we have noted above the inequality states that if $x_1 \geq x_2 \geq \ldots \geq x_n$ and $y_1 \geq y_2 \geq \ldots \geq y_n$ are nonnegative numbers, then for every permutation π of the subscripts of $\mathbf{y}$, the inequalities

$$\Sigma_{i=1}^n x_i y_i \geq \Sigma_{i=1}^n x_i y_{\pi(i)} \geq \Sigma_{i=1}^n x_i y_{n-1+1}$$

hold.

Rearrangement inequalities involving functions of vectors in $\mathscr{R}^n$ have been widely studied in the literature. Jurkat and Ryser (1966) obtained rearrangement inequalities for functions of $\min(x,y)$. They show that for nonnegative n-tuples $\overset{\downarrow}{\mathbf{x}}$ and $\overset{\downarrow}{\mathbf{y}}$

$$\Pi_{i=1}^n \min(x_i,y_i) \geq \Pi_{i=1}^n \min(x_i,y_{\pi(i)}) \geq \Pi_{i=1}^n \min(x_i,y_{n-i+1})$$

and

$$\Sigma_{i=1}^n \min(x_i,y_i) \geq \Sigma_{i=1}^n \min(x_i,y_{\pi(i)}) \geq \Sigma_{i=1}^n \min(x_i,y_{n-1+1})$$

for all permutations π.

Minc (1971) obtained similar rearrangement inequalities for products and sums of $\max(x,y)$.

Rearrangement inequalities also hold for a number of well-known test statistics. An example is Pearson's product moment correlation coefficient given by

$$r(\mathbf{x},\mathbf{y}) = \Sigma_{i,j}(x_i - x_j)(y_i - y_j)\,[\Sigma_{i,j}(x_i - x_j)^2\,\Sigma_{i,j}(y_i - y_j)^2]^{-1/2}$$

Spearman's ρ and Kendall's correlation coefficient τ also yield rearrangement inequalities.

Rearrangement inequalities can be obtained for Blomquist's quadrant test. Blomquist (1950) proposed the following test for positive association:

$$\beta(\mathbf{x},\mathbf{y}) = \Sigma_{i=1}^{n}[a_1(x_i)b_1(y_i) + a_2(x_i)b_2(y_i)],$$

where

$$a(x_i) = 0, \text{ if } x_i \leqslant x_{\text{med}}$$
$$= 1, \text{ if } x_i > x_{\text{med}},$$

$b(y_i)$ is similary defined, $a_2 = 1 - a_1$, and $b_2 = 1 - b_1$.

3. The Arrangement Ordering and Arrangement Increasing Functions. All rearrangement inequalities, such as the ones just described, are examples of functions which are increasing in a partial ordering on $\mathscr{R}^n \times \mathscr{R}^n$. This partial ordering, implicit in the work of Hollander, Proschan, and Sethuraman (1977), is defined in Marshall and Olkin (1979). They refer to the partial ordering as the arrangement ordering. Using this ordering they obtain refinements of rearrangement inequalities involving many more comparisons than given in the examples in the previous section.

To define the arrangement ordering we need some terminology and notation.

Let S_n denote the group of all permutations of $\{1,2, \ldots , n\}$. An element of S_n will be denoted by $\pi \equiv (\pi(1), \ldots , \pi(n))$. Let π and π' be elements of S_n. We say that π' is a *simple transposition* of π if there exist positive integers $1 \leqslant i < j \leqslant n$ such that $\pi(i) = \pi'(j) < \pi'(i) = \pi(j)$ and $\pi(k) = \pi'(k)$ for $k \neq i, j$. We write this as $\pi <^{t_{ij}} \pi'$. For π,π' in S_n we say that π' is a *transposition* of π, written $\pi \overset{t}{\leqslant} \pi$, if $\pi = \pi'$ or if π' can be obtained from π by a sequence of simple transpositions.

For a vector $\mathbf{x}$ in $\mathscr{R}^n$, we define $\mathbf{x}\pi$ to be the vector $(x_{\pi(1)}, \ldots , x_{\pi(n)})$. Recall that we denote by $\overset{\uparrow}{\mathbf{x}}$ the vector obtained from $\mathbf{x}$ by arranging the components of $\mathbf{x}$ in increasing order. We say that $\mathbf{x}'$ is a *transposition* of $\mathbf{x}$ if $\mathbf{x} = \overset{\uparrow}{\mathbf{x}}\pi$, $\mathbf{x}' = \overset{\uparrow}{x}\pi'$ where $\pi \overset{t}{\geqslant} \pi'$. We write $\mathbf{x} \overset{t}{\geqslant} \mathbf{x}'$. We note that this defines a partial ordering of $\mathscr{R}^n$. This partial ordering has been studied by Savage (1957), Lehmann (1966), and Hollander, Proschan, and Sethuraman (1977), among others.

Let $(\mathbf{x},\mathbf{y}) \in \mathscr{R}^n \times \mathscr{R}^n$. The *orbit* of $(\mathbf{x},\mathbf{y})$ is the set $0_{\mathbf{x},\mathbf{y}} = \{(\mathbf{x}\pi,\mathbf{y}\sigma)\colon \pi,\sigma \in S_n\}$. For a vector $\mathbf{x} \in \mathscr{R}^n$ the orbit of $\mathbf{x}$ is defined similarly.

Definition 3.1. Let $(\mathbf{x},\mathbf{y})$ and $(\mathbf{x}',\mathbf{y}')$ be two elements of $\mathscr{R}^n \times \mathscr{R}^n$ belonging to the same orbit. We say that $(\mathbf{x},\mathbf{y})$ is *more similarly arranged* than $(\mathbf{x}',\mathbf{y}')$ if there exist $\pi,\sigma \in S_n$ such that $\mathbf{x}\pi = \mathbf{x}'\sigma = \overset{\uparrow}{\mathbf{x}}$ and $\mathbf{y}\pi \overset{t}{\geqslant} \mathbf{y}'\sigma$. We write $(\mathbf{x},\mathbf{y}) \overset{a}{\geqslant} (\mathbf{x}',\mathbf{y}')$.

This partial ordering of $\mathscr{R}^n \times \mathscr{R}^n$ is referred to as the *arrangement ordering*. We write $(\mathbf{x},\mathbf{y}) \overset{a}{=} (\mathbf{x}',\mathbf{y}')$ if $(\mathbf{x},\mathbf{y}) \overset{a}{\geqslant} (\mathbf{x}',\mathbf{y}')$ and $(\mathbf{x}',\mathbf{y}') \overset{a}{\geqslant} (\mathbf{x},\mathbf{y})$.

Figure 3.1 illustrates the arrangement ordering when $\overset{\uparrow}{\mathbf{x}} = (.5,1,3)$ and $\overset{\uparrow}{\mathbf{y}} = (2,3.5,4)$. An arrow in the diagram from an element $(\overset{\uparrow}{\mathbf{x}},\mathbf{y})$ to an element $(\overset{\uparrow}{\mathbf{x}},\mathbf{y}')$ means that $(\overset{\uparrow}{\mathbf{x}},\mathbf{y}) \overset{a}{\geqslant} (\overset{\uparrow}{\mathbf{x}},\mathbf{y}')$.

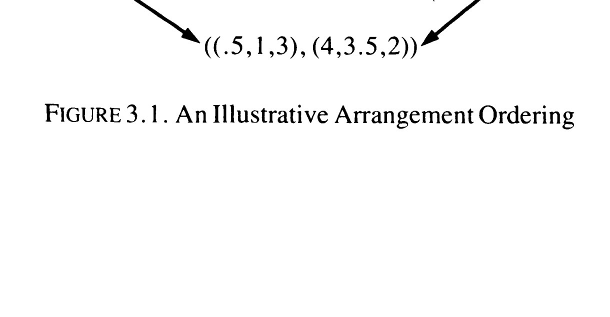


FIGURE 3.1. An Illustrative Arrangement Ordering

Remark. Let $(\mathbf{x},\mathbf{y})$ denote the largest element of its orbit in the arrangement ordering, that is, $(\mathbf{x},\mathbf{y}) \overset{a}{\geqq} (\mathbf{x}\pi,\mathbf{y}\sigma)$ for all $\pi,\sigma \in S_n$. Then it is easy to see that $(x_i-x_j)(y_i-y_j) \geqslant 0$ for all pairs i,j. In this case we say that $\mathbf{x}$ and $\mathbf{y}$ are *similarly arranged.* (Hardy, Littlewood, and Pólya (1952) use the expression "similarly ordered".) We write $\mathbf{x} \overset{s}{=} \mathbf{y}$.

Functions which are order-preserving with respect to the arrangement ordering were introduced by Hollander, Proschan, and Sethuraman (1977).

Definition 3.2. A function f from $\mathcal{R}^n \times \mathcal{R}^n$ into R is said to be *arrangement increasing* if $(\mathbf{x},\mathbf{y}) \overset{a}{\geqq} (\mathbf{x}',\mathbf{y}')$ implies $f(\mathbf{x},\mathbf{y}) \geqslant f(\mathbf{x}',\mathbf{y}')$ for all $(\mathbf{x},\mathbf{y}) \in \mathcal{R}^n \times \mathcal{R}^n$.

Functions which are arrangement increasing play an important role in the theory we develop. Their properties and many useful applications were first studied by Hollander, Proschan, and Sethuraman (1977). In their 1977 paper they gave an alternative definition of an arrangement increasing function which they call a function "decreasing in transposition". The present name is due to Marshall and Olkin.

PROPOSITION 3.3. (Marshall and Olkin (1979).) *A function f from $\mathcal{R}^n \times \mathcal{R}^n$ into $\mathcal{R}$ is arrangement increasing if and only if* (i) $f(\mathbf{x},\mathbf{y}) = f(\mathbf{x}\pi,y\pi)$ *for* $(\mathbf{x},\mathbf{y}) \in \mathcal{R}^n \times \mathcal{R}^n$, $\pi \in S_n$, *and* (ii) $f(\vec{\mathbf{x}},\mathbf{y}) \geqslant f(\vec{\mathbf{x}},\mathbf{y}')$, *where* $\mathbf{y} \overset{t}{\geqslant} \mathbf{y}'$.

A function satisfying (i) of Proposition 3.3 is called *permutation invariant.*

Hollander, Proschan, and Sethuraman (1977) give many examples of arrangement increasing functions including a number of well-known densities in statistics. Some of these examples are presented in Section 5.

4. Stochastic Rearrangement Inequalities. In this section we obtain stochastic versions of the rearrangement inequalities of Section 2. Specifically, we show the following. Let $\mathbf{X}$ and $\mathbf{Y}$ be nonnegative random n-vectors with joint density f satisfying the conditions of Def. 4.1 below. Then for any arrangement increasing function g and permutation π we have

$$g(X_1, \ldots ,X_n; Y_1, \ldots , Y_n) \overset{st}{\geqslant} g(X_1, \ldots ,X_n; Y_{\pi(1)}, \ldots , Y_{\pi(n)}) \overset{st}{\geqslant} g(X_1, \ldots ,X_n; Y_n, \ldots , Y_1).$$

These stochastic rearrangement inequalities follow as a corollary to our main result presented in Theorem 4.3. The condition we need on the joint density of $\mathbf{X}$ and $\mathbf{Y}$ to obtain stochastic rearrangement inequalities is defined as follows.

Definition 4.1. A function f from $\mathcal{R}^n \times \mathcal{R}^n$ into $\mathcal{R}$ is called a *positive set function in arrangement* (PSA) if $\mathbf{x} \geqslant^{t_{ij}} \mathbf{x}'$ and $\mathbf{y} \geqslant^{t_{ij}} \mathbf{y}'$ for any pair $i<j$ imply

$$f(\mathbf{x},\mathbf{y}) - f(\mathbf{x}',\mathbf{y}) - f(\mathbf{x},\mathbf{y}') + f(\mathbf{x}',\mathbf{y}') \geqslant 0.$$

We note that a function f is arrangement increasing if and only if f is PSA and permutation invariant.

Before we state the main theorem, we introduce the notion of an arrangement preserving kernel.

Definition 4.2. A function K from $(\mathcal{R}^n \times \mathcal{R}^n) \times (\mathcal{R}^n \times \mathcal{R}^n)$ into $\mathcal{R}$ is called an *arrangement preserving (AP) kernel* if: (i) $K(\mathbf{u},\mathbf{x}; \mathbf{v},\mathbf{y})$ is permutation invariant in $(\mathbf{u},\mathbf{x})$ and in $(\mathbf{v},\mathbf{y})$, and (ii) For all $\mathbf{u},\mathbf{v} \in \mathcal{R}^n$, $K(\mathbf{u},\mathbf{x}; \mathbf{v},\mathbf{y})$ is PSA in $(\mathbf{x},\mathbf{y})$.

In our main result Theorem 4.3 below we state that the arrangement increasing property is preserved under an integral transform defined by an arrangement preserving kernel.

THEOREM 4.3. *Let $f(\mathbf{x},\mathbf{y})$ be arrangement increasing and let $K(\mathbf{u},\mathbf{x}; \mathbf{v},\mathbf{y})$ be an arrangement preserving kernel. Then under mild conditions on the measure m and the assumption that the integral exists finitely, the function*

$$g(\mathbf{u},\mathbf{v}) = \int\int f(\mathbf{x},\mathbf{y})K(\mathbf{u},\mathbf{x};\mathbf{v},\mathbf{y})m(d\mathbf{x},d\mathbf{y})$$

is arrangement increasing.

A corollary to Theorem 4.3 yields stochastic versions of deterministic rearrangement inequalities.

COROLLARY 4.4. *Let* $(\mathbf{X},\mathbf{Y})$ *have a PSA density. Then for all arrangement increasing functions f and all permutations* π *we have*

$$f(X_1, \ldots, X_n; Y_1, \ldots, Y_n) \overset{st}{\geqslant} f(X_1, \ldots, X_n; Y_{\pi(1)}, \ldots, Y_{\pi(n)}) \overset{st}{\geqslant} f(X_1, \ldots, X_n; Y_n, \ldots, Y_1).$$

Since the function $g(\mathbf{x},\mathbf{y}) = \Sigma x_i y_i$ is arrangement increasing we have as a consequence of Corollary 4.4 a stochastic version of the Hardy, Littlewood, and Pólya inequality, namely that

$$\Sigma X_i Y_i \overset{st}{\geqslant} \Sigma X_i Y_{\pi(i)} \overset{st}{\geqslant} \Sigma X_i Y_{n-i+1}.$$

A similar result holds for all the other rearrangement inequalities in Section 2.

5. Examples of PSA Functions and AP Kernels. The results in the previous section allow us to obtain stochastic versions of rearrangement inequalities for a large class of random vectors which contains those pairs $(\mathbf{X},\mathbf{Y})$ having PSA and AP densities. The purpose of this section is to show that many multivariate densities of interest in statistical practice fall into these two classes of functions.

A function ϕ is called a *positive set function* if

$$\phi(x_1,y_1) - \phi(x_1,y_2) - \phi(x_2,y_1) + \phi(x_2,y_2) \geqslant 0 \text{ for all } x_1 < x_2 \text{ and } y_1 < y_2.$$

Positive set functions can be used to construct AP kernels as we state in Theorem 5.1.

THEOREM 5.1. *Let* ϕ *be a positive set function and let* g_1 *and* g_2 *be arrangement increasing. Then* $\phi(g_1(\mathbf{u},\mathbf{x}), g_2(\mathbf{v},\mathbf{y}))$ *is an AP kernel.*

Some examples of positive set functions are (i) $\phi(x,y) = xy$, (ii) $\phi(x,y) = F(x,y)$ where F is a c.d.f., and (iii) $\phi(x,y) = h(x-y)$ where h is concave.

As a consequence of Theorem 5.1 and the fact that the product xy is a positive set function we have the following important example showing how to construct AP kernels

Example 5.2. Let $\mathbf{X}$ and $\mathbf{Y}$ be independent random vectors each having arrangement increasing density. Then the joint density of $(\mathbf{X},\mathbf{Y})$ is an AP kernel.

The following examples of AI densities can be used to construct AP densities. (See Hollander, Proschan, and Sethuraman (1977) for proofs.)

5.3.a. Multinomial: $g_1(\mathbf{u},\mathbf{x}) = N!\,\Pi_{i=1}^n(u_i^{x_i}/x_i!)$, where $0 < u_i < 1$, $x_i = 0,1,2, \ldots$, $i = 1, \ldots, n$, $\Sigma_{i=1}^n u_i = 1$, and $\Sigma_{i=1}^n x_i = N$.

5.3.b. Negative multinomial:

$$g_2(\mathbf{u},\mathbf{x}) = (\Gamma(N))^{-1}\Gamma(N + \Sigma_{i=1}^n x_i)(1 + \Sigma_{i=1}^n u_i)^{-N-\Sigma_{i=1}^n x_i}\Pi_{i=1}^n(u_i^{x_i}/x_i!),$$

where $u_i > 0, x_i = 0,1, \ldots, i = 1, \ldots, n$, and $N > 0$.

5.3.c. Multivariate hypergeometric: $g_3(\mathbf{u},\mathbf{x}) = \Pi_{i=1}^n\binom{u_i}{x_i} \Big/ \binom{\Sigma_{i=1}^n u_i}{N}$, where $u_i > 0$, $x_i = 0,1, \ldots, \Sigma_{i=1}^n x_i = N < \Sigma_{i=1}^n u_i$.

5.3.d. Dirichlet: $g_4(\mathbf{u},\mathbf{x}) = (\Gamma(\theta)\Pi_{i=1}^n G(u_i))^{-1}\Gamma(\theta + \Sigma_{i=1}^n u_i)(1 - \Sigma_{i=1}^n x_i)^{\theta-1}\Pi_{i=1}^n x_i^{u_i-1}$, where $u_i > 0, x_i \geqslant 0, i = 1, \ldots, n$. $\Sigma_{i=1}^n x_i \leqslant 1$, and $\theta > 0$.

5.3.e. Inverted Dirichlet:

$g_5(\mathbf{u},\mathbf{x}) = (\Gamma(\theta)\Sigma_{i=1}^n\Gamma(u_i))^{-1}\Gamma(\theta + \Sigma_{i=1}^n u_i)\Pi_{i=1}^n x_i^{u_i-1}/((1 + \Sigma_{i=1}^n x_i)^\theta + \Sigma_{i=1}^n x_i)$,
where $u_i > 0, x_i \geq 0, i = 1, \ldots, n$, *and* $\theta > 0$.

5.3.f. Negative multivariate hypergeometric:

$$g_6(\mathbf{u},\mathbf{x}) = (\Pi_{i=1}^n x_i!\Gamma(N + \Sigma_{i=1}^n u_i))^{-1}N!\,(\Sigma_{i=1}^n u_i)\Pi_{i=1}^n(\Gamma(x_i+u_i)/\Gamma(u_i)),$$

where $u_i > 0, x_i = 0,1, \ldots, N, \Sigma_{i=1}^n x_i = N$, and $N = 1,2, \ldots$.

5.3.g. Dirichlet compound negative multinomial: $g_7(\mathbf{u},\mathbf{x}) = (\Pi_{i=1}^n x_i!\Gamma(N)\Gamma(\theta)\Gamma(N + \theta + \Sigma_{i=1}^n u_i + \Sigma_{i=1}^n x_i))^{-1}\,\Gamma(N - \Sigma_{i=1}^n x_i)\,\Gamma(\theta + \Sigma_{i=1}^n u_i)\,\Gamma(N + \theta)\,\Pi_{i=1}^n(\Gamma(x_i + u_i)/\Gamma(u_i))$, where $u_i > 0, x_i = 0,1, \ldots, i = 1, \ldots, n, \theta > 0$, and $N = 1,2, \ldots$.

5.3.h. Multivariate logarithmic series distribution:

$$g_8(\mathbf{u},\mathbf{x}) = (\log(1 + \Sigma_{i=1}^n u_i))^{-1}(\Sigma_{i=1}^n x_i - 1)!\,(1 + \Sigma_{i=1}^n u_i)^{-\Sigma_{i=1}^n x_i}\Pi_{i=1}^n(u_i^{x_i}/x_i!),$$

where $u_i > 0, x_i = 0,1, \ldots, i = 1, \ldots, n$.

5.3.i. Multivariate F distribution:

$$g_9(\mathbf{u},\mathbf{x}) = (2\Pi_{i=0}^n\Gamma(\lambda_i)(\lambda_0 + \Sigma_{i=1}^n\lambda_i x_i)^\lambda)^{-1}\,\Gamma(u)\Pi_{i=0}^n(2u_i)^{u_i}\Pi_{i=1}^n x_i^{u_i-1},$$

where $u_i > 0, i = 0,1, \ldots, n, u = \Sigma_{i=0}^n\lambda_i, x_i \geq 0, i = 0,1, \ldots, n$.

5.3.j. Multivariate Pareto distribution: $g_{10}(\mathbf{u},\mathbf{x}) = (\Pi_{i=1}^n u_i)^{-1}(\Sigma_{i=1}^n\lambda_i^{-1}x_i - n + 1)^{-(a+n)}$, where $x_i > u_i > 0, i = 1, \ldots, n$, and $a > 0$.

5.3.k. Multivariate normal distribution with common variance and common covariance: $g_{11}(\mathbf{u},\mathbf{x}) = (2\pi)^{n/2}|\Sigma|^{-1/2}\exp(-½(\mathbf{x}-\mathbf{u})\Sigma^{-1}(\mathbf{x}-\mathbf{u})^{-1})$, where Σ is the positive definite covariance matrix with elements σ^2 along the main diagonal and elements $\rho\sigma^2$ elsewhere, $\rho > -1/(n-1)$.

AP densities can also be constructed from independent TP_2 densities using the next result. Recall that f is TP_2 if

$$f(x_1,y_1)f(x_2,y_2) \geq f(x_1,y_2)f(x_2,y_1) \text{ for all } x_1 < x_2 \text{ and } y_1 < y_2.$$

THEOREM 5.4. *Let f be TP_2. Then $\Pi f(\theta_i,y_i)$ is arrangement increasing.*

Some examples of TP_2 densities are (i) normal with variance 1, (ii) exponential, and (iii) Poisson.

In the next example we use a result of Karlin for TP_2 densities to obtain more PSA densities.

Suppose that components have lifelengths X with TP_2 density $g(\theta,x)$ and Y with TP_2 density $f(\theta,y)$. Further suppose that X and Y are independent and that θ depends on the environment with distribution $\pi(\theta)$. Then Karlin (1968) has shown that the joint distribution of X,Y given by $K(x,y) = \int f(\theta,y)g(\theta,y)d\pi(\theta)$ is TP_2. From this result we get the following:

THEOREM 5.5. *Let $(X_1,Y_1), \ldots, (X_n,Y_n)$ be independent with TP_2 density. Then the joint density of $(\mathbf{X},\mathbf{Y}) = (X_1, \ldots, X_n; Y_1, \ldots, Y_n)$ is arrangement increasing and hence PSA.*

In the next two results we state that certain operations on pairs of random vectors which are PSA (AP) preserve the PSA (AP) property.

A vector $\mathbf{x} \in \mathcal{R}^n$ *majorizes* $\mathbf{y} \in \mathcal{R}^n$ if $\Sigma_{i=1}^k x_{[i]} \geq \Sigma_{i=1}^k y_{[i]}$ for $k = 1, \ldots, n$ and $\Sigma_{i=1}^n x_i = \Sigma_{i=1}^n y_i$ where $x_{[1]} \geq \ldots \geq x_{[n]}$. A function f is *Schur-concave if* $f(\mathbf{x}) \leq f(\mathbf{y})$ whenever $\mathbf{x}$ majorizes $\mathbf{y}$. In Theorem 5.6 we show that Schur-concave densities can be used to construct PSA densities.

THEOREM 5.6. *Let* $(\mathbf{X},\mathbf{Y})$ *have a PSA (AP) density. Let* $\mathbf{W}$ *and* $\mathbf{Z}$ *be mutually independent and independent of* $(\mathbf{X},\mathbf{Y})$ *each having a Schur-concave density. Then* $(\mathbf{X} + \mathbf{W},\mathbf{Y} + \mathbf{Z})$ *has a PSA (AP) density.*

Marshall and Olkin (1974) give a number of interesting examples of Schur-concave densities. We list some of them in Example 5.7 below.

Example 5.7. The following multivariate densities are Schur-concave:

5.7.a. Multivariate normal: Let $X_1, \ldots, X_n$ be exchangeable and jointly normally distributed. Then the joint density of $\mathbf{X}$ is Schur-concave.

5.7.b. Multivariate "t" distribution: Let $U_1, \ldots, U_n$ be exchangeable and jointly normally distributed. Let Z^2 be chi-square distributed for $Z \geqslant 0$. Then $\mathbf{X} = (U_1/Z, \ldots, U_m/Z)$ has a Schur-concave density.

5.7.c. Multivariate beta distribution: Let $U_1, \ldots, U_n$ be independent, identically distributed chi-square random variables and let Z be a chi-square random variable, indpendent of $U_1, \ldots, U_n$. Then $\mathbf{X} = (U_1/(\Sigma U_i+Z), \ldots, U_m/(\Sigma U_i+Z)$ has a Schur-concave density.

5.7.d. Multivariate "F" distribution: Let $U_1, \ldots, U_n$ each have a chi-square distribution with $r \geqslant 2$ degress of freedom, and let Z have a chi-square distribution with s degrees of freedom. Then $\mathbf{X} = (U_1/Z, \ldots, U_n/Z)$ has a Schur-concave density.

For random variables $X_1, \ldots, X_n$, denote by R_i the rank of X_i among $X_1, \ldots, X_n$. The random vector $\mathbf{R} \equiv (R_1, \ldots, R_n)$ is called the *rank order* of $(X_1, \ldots, X_n)$. We have the following useful result for the rank orders of PSA (AP) random vectors $(\mathbf{X},\mathbf{Y})$.

THEOREM 5.8. *Let* $(\mathbf{X},\mathbf{Y})$ *be PSA (AP). Let* $\mathbf{R}$ *be the rank order of* $\mathbf{X}$ *and let* $\mathbf{S}$ *be the rank order of* $\mathbf{Y}$. *Then the random vectors* $(\mathbf{R},\mathbf{S})$ *are PSA (AP).*

6. Applications to Statistics. The theory of stochastic rearrangement inequalities has applications in a number of areas in statistics. In this section we present two of these applications.

In the first example we show how the stochastic version of the Hardy, Littlewood, and Pólya inequality may be applied to reliability theory. This generalizes a result of Derman, Lieberman, and Ross (1972) in the case of two vectors.

Application 6.1. Suppose that we have two stockpiles of n components each, stockpile one of type 1 components, stockpile two of type 2 components. From these stockpiles we are to construct n systems, each composed of a component of type 1 and a component of type 2 arranged in series. A component i of type j has a random reliability p_i^j, $j = 1,2$; $i = 1, \ldots, n$. We assume that $\mathbf{P}^1 \equiv (P_1^1, \ldots, P_n^1)$ and $\mathbf{P}^2 \equiv (P_1^2, \ldots, P_n^2)$ are independent, each having an AI density with parameters $\alpha_1 \leqslant \ldots \leqslant \alpha_n$ and $\beta_1 \leqslant \ldots \leqslant \beta_n$, respectively. Then, as we have seen in Section 4, $(\mathbf{P}^1,\mathbf{P}^2)$ is AP.

For the assembly which pairs the i-th component of type 1 with the $\pi(i)$-th component of type 2, the average reliability of the n system is $1/n \, \Sigma_{i=1}^n p_i^1 p_{\pi(i)}^2$.

Thus by the stochastic Hardy, Littlewood, and Pólya inequality, the optimal assembly, in terms of average reliability of the n systems, is achieved when the i-th component of type 1 is paired with the i-th component of type 2. □

Let $(\mathbf{X},\mathbf{Y})$ be AP with parameters (α,β) Let α_0 be a fixed vector of $\mathcal{R}^n$ in the orbit of α. The theory we have developed can be used to study the problem of testing the hypothesis

(6.1) $$H_0: \beta \overset{s}{=} \alpha_0 \quad \text{against } H_a: \beta \overset{s}{\neq} \alpha_0.$$

Let f be an AI function and define the test T_f by

(6.2) $$T_f(\mathbf{x},\mathbf{y}) = \begin{cases} 1, & \text{if } f(\mathbf{x},\mathbf{y}) < \nu_\alpha \\ \gamma, & \text{if } f(\mathbf{x},\mathbf{y}) = \nu_\alpha \\ 0, & \text{otherwise.} \end{cases}$$

The null hypothesis is rejected with probability $T_f(\mathbf{x},\mathbf{y})$ if $(\mathbf{x},\mathbf{y})$ is observed. Note that the numbers ν_α and $(0 < \gamma < 1)$ are determined to give size α to the test.

Let $B_{T_f}(\alpha,\beta)$ be the power function of the above test against alternatives (α,β), that is,

$$B_{T_f}(\alpha,\beta) = ET_f(X(\alpha),Y(\beta)).$$

We shall need the following definition (see Barlow, Bartholomew, Bremner, and Brunk (1972), Chapter 6).

Definition 6.2. Let $(\alpha_0,\beta_0) \in \mathcal{R}^n \times \mathcal{R}^n$ be given. A test T has *isotonic power* against alternative $(\alpha,\beta) \overset{a}{\leq} (\alpha_0,\beta_0)$ (with respect to the ordering "$\overset{a}{\leq}$") if for any (α_1,β_1) and (α_2,β_2) in $\mathcal{R}^n \times \mathcal{R}^n$ such that

$$(\alpha_2,\beta_2) \overset{a}{<} (\alpha_1,\beta_1) \overset{a}{<} (\alpha_0,\beta_0),$$

we have

$$B_{T_f}(\alpha_2,\beta_2) \geqslant B_{T_f}(\alpha_1,\beta_1).$$

Remark 6.3. It is a consequence of Definition 6.2 that any test T which is isotonic with respect to the "$\overset{a}{\leq}$" ordering is unbiased for testing

(6.3) $$H_0: (\alpha_0,\beta) \overset{a}{=} (\alpha_0,\beta_0) \quad \text{against}$$
$$H_a: (\alpha_0,\beta) \overset{a}{<} (\alpha_0,\beta_0), (\alpha_0,\beta) \overset{a}{\neq} (\alpha_0,\beta_0).$$

Note that by the remark in Section 3, the hypotheses in (6.3) are equivalent to those in (6.1).

It follows from Theorem 4.3, that tests of the form given in (6.2) are isotonic with respect to the arrangement ordering and, consequently, by Remark 6.3 that such tests will be unbiased for testing H_0 against H_a. We state this formally in Theorem 6.4.

THEOREM 6.4. *Let $(\mathbf{X},\mathbf{Y})$ be SSA like (α_o,β). Consider testing the hypothesis*

$$H_0: (\alpha_0,\beta) \overset{a}{=} (\alpha_0,\beta_0)$$

against

$$H_a: (\alpha_0,\beta) \overset{a}{<} (\alpha_0,\beta), (\alpha_0,\beta) \overset{a}{\neq} (\alpha_0,\beta_0).$$

Let f be an AI function and let T_f be the test given in (6.2). Then the test T_f has isotonic power against alternatives $(\alpha_0,\beta) \overset{a}{\leq} (\alpha_0,\beta_0)$. Consequently, a test based on T_f is unbiased for testing H_0: $\alpha_0 \overset{s}{=} \beta$ against H_a: $\alpha_0 \overset{s}{\neq} \beta$.

A number of well-known statistics, including those given in Section 2 such as Spearman's ρ and Kendall's τ are AI functions and hence can be used to test the hypotheses in (6.1).

Remark 6.5. In Theorem 5.8 we showed that if $(\mathbf{X},\mathbf{Y})$ is PSA (AP) then its rank order $(\mathbf{R},\mathbf{S})$ is PSA (AP). Thus Theorem 6.4 also holds for test statistics T_f based on the rank order of $(\mathbf{X},\mathbf{Y})$. A useful application of the above remark arises in testing for the existence of positive dependence between two time series. An example is described below.

Application 6.6. Studies of air pollution have shown that automobile exhaust is the major source of lead elemental air pollution in many urban areas. It is believed that automobile exhaust is also the major source of bromine pollution in the atmosphere. For a particular city, we wish to determine whether automobile exhaust is the predominant source of both of these two pollutants or, alternatively, whether other sources are responsible for bromine pollution. Suppose that λ_i, the concentration of lead at time i, $i = 1, \ldots, n$, is known. Let $\lambda_0 = (\lambda_1, \ldots, \lambda_n)$.

To help in distinguishing between the two alternative hypotheses, we test H_0: $\lambda_0 \overset{s}{=} \beta$ against H_a: $\lambda_0 \overset{s}{\neq} \beta$, where β_i is the true concentration of bromine at time i, $i = 1, \ldots, n$, and $\beta = (\beta_1, \ldots, \beta_n)$. Rejection of H_0 would indicate that sources other than automobile exhaust contribute to the bromine pollution.

Observations **L** on lead and **B** on bromine are assumed to be governed by a joint AP density with parameters (λ_0, β). By Theorem 6.4 we conclude that a test using *an AI test statistic based on the ranks of* **L** *and the ranks of* **B** *is isotonic and is consequently unbiased against* H_a.

Remark 6.7. Suppose that the measurements **L** and **B** are subject to errors **X** and **Y** with $\mathbf{X} \sim \text{MVN}(\mathbf{0}, \Sigma(\rho_1))$ and $\mathbf{Y} \sim \text{MVN}(\mathbf{0}, \Sigma(\rho_2))$, where

$$\Sigma(\rho) = \begin{pmatrix} \sigma^2 & & \rho\sigma^2 \\ & \ddots & \\ \rho\sigma^2 & & \sigma^2 \end{pmatrix}$$

for $0 \leq \rho \leq 1$. Since the density of each of **X** and **Y** is Schur-concave by Theorem 5.6, $(\mathbf{L}+\mathbf{X}, \mathbf{B}+\mathbf{Y})$ is PSA (AP) and, as before, a test using an AI test statistic based on the ranks of $\mathbf{L}+\mathbf{X}$ and of $\mathbf{B}+\mathbf{Y}$ is isotonic.

REFERENCES

BARLOW, R.E., BARTHOLOMEW, D.J., BREMNER, J.M., and BRUNK, H.D. (1972). *Statistical Inference Under Order Restrictions*. Wiley, New York.

BLOMQUIST, N. (1950). On a measure of dependence between two random variables. *Ann. Math. Statist. 21* 593–600.

DERMAN, C., LIEBERMAN, G.J., and ROSS, S.M. (1972). On optimal assembly of systems. *Nav. Res. Logist. Quart. 19* 569–573.

HARDY, G.H., LITTLEWOOD, J.E., and PÓLYA, G. (1952). *Inequalities*. 2nd. ed. Cambridge University Press, London and New York.

HOLLANDER, M., PROSCHAN, F., and SETHURAMAN, J. (1977). Functions decreasing in transposition and their applications in ranking problems. *Ann. Statist. 5* 722–733.

JURKAT, F.B., and RYSER, H.J. (1966). Matrix factorization of determinants and permanents. *J. Algebra. 15* 60–77.

KARLIN, S. (1968). *Total Positivity*. Stanford University Press, Stanford, CA.

LEHMANN, E.L. (1966). Some concepts of dependence. *Ann. Math. Statist. 37* 1137–1153.

MARSHALL, A. and OLKIN, I. (1979). *Inequalities: Theory of Majorization and Its Applications*. Academic Press, New York.

MINC, H. (1971). Rearrangements. *Trans. Amer. Math. Soc. 159* 497–504.

SAVAGE, I.R. (1957). Contributions to the theory of rank order statistics—the "trend" case. *Ann. Math. Statist. 28* 968–977.

Inequalities in Statistics and Probability
IMS Lecture Notes–Monograph Series Vol. 5 (1984), 13–25

ON GROUP INDUCED ORDERINGS, MONOTONE FUNCTIONS, AND CONVOLUTION THEOREMS[1]

BY MORRIS L. EATON
University of Minnesota

Orderings defined by compact groups of linear transformations acting on vector spaces are studied. In some cases, these orderings induce orderings on convex cones similar to those defined by reflection groups. In these cases the monotone functions can be conveniently characterized. Convolution theorems for monotone functions are discussed.

1. Introduction. Majorization, as defined by Hardy, Littlewood and Pólya (1934), has been an extremely important notion in the theory and applications of many types of inequalities. The recent work of Marshall and Olkin (1979) contains an extensive discussion of majorization and its application to many branches of mathematics including probability and statistics. Although not essential for understanding this paper, the reader may find it useful to glance through Part I of Marshall and Olkin (1979).

To motivate the situation to be considered here, first recall the permutation group definition of majorization (see Rado (1952)). Let $\mathcal{P}_n$ be the group of $n \times n$ permutation matrices acting on $\mathcal{R}^n$. For $x,y \in \mathcal{R}^n$, x is majorized by y (written as $x \leqslant y$) means that x is in the convex hull of the set $\{gy|g\epsilon\mathcal{P}_n\}$ (the $\mathcal{P}_n$-orbit of y). A careful study of the pre-order $\leqslant$ (using the terminology in Marshall and Olkin (1979), p. 13) has resulted in a useful and important characterization of the real valued functions f which are decreasing or increasing in the pre-order of majorization (see Schur (1923), Ostrowski (1952)). A recent result of Marshall and Olkin (1974), which has had applications in probability and statistics, shows that the convolution of two decreasing (in the pre-order of majorization) functions is again a decreasing function.

In this paper, we begin a systematic study of pre-orderings defined on vector spaces which arise in much the same way that majorization arises. Let G be any closed group of $n \times n$ orthogonal matrices. Using G, rather than $\mathcal{P}_n$, define a pre-order on $\mathcal{R}^n$ as follows: $x \leqslant y$ iff x is in the convex hull of $\{gy|g\epsilon G\}$. The examples in the next section show that there are a number of groups G which give useful and interesting orderings. Based on the known majorization results, it seems rather natural to ask for conditions on G for which (i) it is possible to characterize the class of decreasing real valued functions on $\mathcal{R}^n$. (ii) the convolution result of Marshall and Olkin (1974) continues to hold.

This paper is mainly concerned with (i), but (ii) is discussed rather incompletely. Here is a brief outline of the paper. In Section 2, group induced orderings are defined on inner product spaces. The geometry which prevails in the permutation group case is described and is shown to hold in a number of interesting cases. It is this geometry which is used in Section 3 to give a characterization of the decreasing functions. The results of Marshall, Walkup and Wets (1967) on cone orderings are used extensively in Section 3. In Section

[1] This work was supported in part by National Science Foundation Grant MCS 81-00762.
AMS 1980 subject classifications. Primary, 52A40; secondary, 60E15, 06A10.
Key words and phrases: group induced orderings, reflection groups, monotone functions, Schur convexity, cone orderings, convolution theorems.

4, the convolution type results are discussed with special attention being given to a necessary condition for the Marshall and Olkin (1974) result to hold. It is shown that this necessary condition does not hold for any finite rotation group acting on $\mathcal{R}^2$.

2. Group Induced Orderings. To set notation, let $(V,(\cdot,\cdot))$ be a finite dimensional real inner product space and let G be a closed subgroup of the group of orthogonal transformations $\mathcal{O}(V)$, of $(V,(\cdot,\cdot))$. The compact group G defines a pre-order on V as follows:

(2.1) For $x,y \in V$, write $x \leqslant y$ to mean x is in the convex hull of $\{gy|g\epsilon G\}$.

Thus, $x \leqslant y$ means that x is in the convex hull of the G-orbit of y. To say $\leqslant$ is a pre-order means that for all $x,y,z\epsilon V$, (i) $x \leqslant y$ and $y \leqslant z$ implies $x \leqslant z$ and (ii) $x \leqslant x$.

These two conditions are easily checked. The dependence of $\leqslant$ on G will usually be supressed as G will remain fixed through much of our discussion.

Definition 1. *A function $f:V \to \mathcal{R}^1$ is decreasing (increasing) if $x \leqslant y$ implies $f(x) \geqslant f(y)$ $(f(x) \leqslant f(y))$. A set $B \subseteq V$ is called* monotone *if the indicator of B, say I_B, is decreasing.*

Our first task is to give an analytic rather than geometric description of $\leqslant$. To this end, recall the following.

PROPOSITION 1. *Let A be a non-empty subset of V and let C be the closed convex set generated by A. Then, $x\epsilon C$ iff. for all $u\epsilon V$,*

(2.2) $$(u,x) \leqslant \sup_{z\epsilon A}(u,z).$$

Proof. Without loss of generality, $C \neq V$ since otherwise the right hand side of (2.2) is $+\infty$ and the assertion is trivial. If $x = \Sigma\, \alpha_i z_i$ with $z_i \epsilon A$, $0 \leqslant \alpha_i \leqslant 1$, and $\Sigma\alpha_i = 1$, then $(u,x) = \Sigma\alpha_i(u,z_i) \leqslant \sup_{z\epsilon A}(u,z)$ so (2.2) holds for convex combinations of elements of A. But, every point in C is the limit of such convex combinations so continuity implies that (2.2) holds for all $x\epsilon C$. Conversely, assume (2.2) holds and write C as the intersection of all the closed half spaces which contains it—say $C = \bigcap_\alpha H_\alpha$ where $H_\alpha = \{y|(h_\alpha, y) \leqslant k_\alpha\}$ with $\|h_\alpha\| = 1$ and $k_\alpha \epsilon \mathcal{R}^1$. Since $A \subseteq C \subseteq H_\alpha$ for all α, we have $(h_\alpha, z) \leqslant k_\alpha$ for all $z\epsilon A$. If x satisfies (2.2), then $(h_\alpha, x) \leqslant \sup_{z\epsilon A}(h_\alpha, z) \leqslant k_\alpha$ so $x\epsilon H_\alpha$ for all α. Hence $x\epsilon \bigcap_\alpha H_\alpha = C$. □

Given $y\epsilon V$, let $C(y)$ denote the convex hull of $\{gy|g\epsilon G\}$. The compactness of G implies $C(y)$ is compact. Since $x \leqslant y$ means $x \in C(y)$, Proposition 1 with $A = \{gy|g\epsilon G\}$ shows that $x \leqslant y$ iff for all $u\epsilon V$

(2.3) $$(u,x) \leqslant \sup_{g\epsilon G}(u,gy).$$

For $u,y\epsilon V$, consider

(2.4) $$m(u,y) \equiv \sup_{g\epsilon G}(u,gy).$$

defined on $V \times V$. The following properties of m are easily verified.

(2.5)
(i) $m(c_1u,c_2y) = c_1c_2m(u,y)$ for $c_1,c_2 \geqslant 0$
(ii) $m(g_1u,g_2y) = m(u,y)$ for $g_1,g_2\epsilon G$
(iii) $m(u,y) = m(y,u)$
(iv) $m(u,\cdot)$ is convex on V

That the pre-order $\leqslant$ is completely determined by m is the content of

PROPOSITION 2. *For $x,y \in V$, $x \leqslant y$ iff.*

(2.6) $$m(u,x) \leqslant m(u,y) \quad \textit{for all } u\epsilon V.$$

Proof. If $x \leq y$, then (2.3) shows that for all $u \in V$,

$$(u,x) \leq m(u,y). \tag{2.7}$$

Since $m(gu,y) = m(u,y)$, (2.7) implies that for $g \in G$,

$$(gu,x) \leq m(u,y), \tag{2.8}$$

so $m(u,x) = \sup_{g \in G}(gu,x) \leq m(u,y)$. If (2.6) holds, the inequality $(u,x) \leq m(u,x)$ together with (2.3) shows that $x \leq y$. □

In a number of important examples, Proposition 2 can be used to provide a useful analytic characterization of $\leq$. First, (2.5) (ii) shows that m is determined by its values on the quotient space V/G. In other words, m is a function of a maximal invariant under the action of G on V. Let τ be such a maximal invariant (see Lehmann (1959), Ch. 6). *Assume* that $\tau(x) \in \{gx | g \in G\}$, and let $\mathcal{T} \subseteq V$ be the range of τ. Thus, $\tau(x) = \tau(gx)$ for all $x \in V$, $g \in G$ and $\tau(x_1) = \tau(x_2)$ implies that $x_1 = gx_2$ for some $g \in G$. From (2.5) (ii), we see that

$$m(u,y) = m(\tau(u), \tau(y)) \tag{2.9}$$

for all $u,y \in V$. This implies

PROPOSITION 3. *For $x,y \in V$, $x \leq y$ iff $m(\tau(u), \tau(x)) \leq m(\tau(u), \tau(y))$ for all $u \in V$.*

Proof: This is immediate from Proposition 2 and (2.9). □

For all of the interesting examples that I know, there is a natural choice for τ which results in $\mathcal{T}$ being a convex cone (such $\mathcal{T}$'s are often called fundamental regions—see Benson and Grove (1971), p. 27). The following assumption is to hold for the remainder of this section:

(A.1) The maximal invariant τ has a range $\mathcal{T} \subseteq V$ which is a convex cone, and $\tau(x) \in \{gx | g \in G\}$.

The key to analyzing a number of important examples is being able to calculate the restriction of m to $\mathcal{T}$. Many of these examples are special cases of the following result.

PROPOSITION 4. *For $\beta, \gamma \in \mathcal{T}$, suppose that $m(\beta, \gamma) = (\beta, \gamma)$—that is, assume m restricted to $\mathcal{T} \times \mathcal{T}$ is just the inner product on $V \times V$. Then $x \leq y$ iff*

$$(\beta, \tau(x)) \leq (\beta, \tau(y)) \quad \text{for all } \beta \in \mathcal{T} \tag{2.10}$$

Proof. This is an immediate consequence of Proposition 3 and the assumption that m restricted to $\mathcal{T} \times \mathcal{T}$ is just the inner product. □

Recall that a subset T of $\mathcal{T}$ *spans* $\mathcal{T}$ *positively* if every element of $\mathcal{T}$ can be written as a positive linear combination of a finite number of elements of T. Further, T is called a *frame* if T spans $\mathcal{T}$ positively, but no proper subset of T does. The following result is clear.

COROLLARY 1. *Under the assumption of Proposition 4, if T spans $\mathcal{T}$ positively, then $x \leq y$ iff.*

$$(t, \tau(x)) \leq (t, \tau(y)) \quad \textit{for all } t \in T. \tag{2.11}$$

Before discussing a characterization of the decreasing functions, we first introduce the examples alluded to above. At this point it is appropriate to mention the recent work of Jensen (1984) whose examples coincide with some here. Jensen considers orderings (sometimes pre-orders, lattice orders, etc.) on a set (corresponding to our $\mathcal{T}$) and then lifts the ordering to the whole space via an invariance requirement. Aside from applications, Jensen's main concern is the effect of the lifting but he does not attempt to identify general situations where the lifted ordering is equivalent to the type of group induced ordering discussed above. However, the overlap of Jensen's and our examples show that closely related ideas generated the two works. The important special case treated by Proposition 4 and Corollary 1 is not discussed in Jensen. Under a rather weak assumption, this case leads to a complete description of the G-decreasing functions (see Section 3).

The first three examples here are also discussed briefly in Jensen (1984).

Example 2.1. Take $V = \mathcal{R}^n$ with the usual inner product and let $\mathcal{D}_n$ be the group of coordinate sign changes. Elements of $\mathcal{D}_n$ can be represented as $n \times n$ diagonal matrices whose diagonal elements are $\pm$ 1. Let $\mathcal{F} = \{x | x_i \geq 0, i = 1, \ldots, n\}$ so a frame for $\mathcal{F}$ is $T = \{\epsilon_1, \ldots, \epsilon_n\}$ where ϵ_i is the i-th unit vector in $\mathcal{R}^n$. A convenient choice for τ is $(\tau(x))_i = |x_i|$, $i = 1, \ldots n$; $\tau(x)$ is the vector of absolute values of the coordinates of $x \in \mathcal{R}^n$. For $\beta, \gamma \in \mathcal{F}$, $m(\beta, \gamma) = \sup_{g \epsilon G} \beta' g \gamma = \beta' \gamma$. The definition of m gives the first equality while the second follows from the non-negativity of the coordinates of β and γ. Thus Corollary 1 is applicable and yields $x \leqslant y$ iff $\epsilon_i' \tau(x) \leqslant \epsilon_i' \tau(y)$, $i = 1, \ldots, n$ which is equivalent to $|x_i| \leqslant |y_i|$, $i = 1, \ldots, n$.

Example 2.2. Again take $V = \mathcal{R}^n$ with the usual inner product and take G to be the group $\mathcal{P}_n$ of permutations acting on $\mathcal{R}^n$. Let

$$\mathcal{F} = \{x | x_1 \geqslant \ldots \geqslant x_n, \mathbf{x} \in \mathcal{R}^n\}$$

and let e_i be the vector whose first i coordinates are 1 and the rest of the coordinates are 0, $i = 1, \ldots, n$. It is not hard to show that $T = \{e_1, \ldots, e_n, -e_n\}$ is a frame for $\mathcal{F}$. A classical rearrangement result due to Hardy, Littlewood and Pólya (1952, p. 261) shows that $m(\beta, \gamma) = \beta' \gamma$ for $\beta, \gamma \in \mathcal{F}$. Let $\tau(x)$ be the vector of the ordered values of x so $\tau(x) \in \mathcal{F}$. These ordered values are denoted by $x_{(i)}$, $i = 1, \ldots, n$ so $x_{(1)} \geqslant x_{(2)} \geqslant \ldots \geqslant x_{(n)}$. A direct application of Corollary 1 shows that $x \leqslant y$ iff $e_i' \tau(x) \leqslant e_i' \tau(y)$, $i = 1, \ldots, n-1$, and $e_n' \tau(x) = e_n' \tau(y)$. Thus, $x \leqslant y$ iff.

$$(*) \qquad \Sigma_1^k x_{(i)} \leqslant \Sigma_1^k y_{(i)}, k = 1, \ldots, n-1 \text{ and } \Sigma_1^n x_i = \Sigma_1^n y_i.$$

Of course, this is the traditional ordering of majorization discussed at length in Marshall and Olkin (1979). For this example, that (*) is equivalent to saying x is in the convex hull of $\{gy | g \in \mathcal{P}_n\}$ was observed by Rado (1952).

Example 2.3. We use the notation established in Examples 2.1 and 2.2. Take $V = \mathcal{R}^n$ and take G to be the group generated by $\mathcal{D}_n \cup \mathcal{P}_n$. Take $\mathcal{F}$ to be

$$\mathcal{F} = \{\mathbf{x} | x_1 \geqslant x_2 \geqslant \ldots \geqslant x_n \geqslant 0\}$$

and note that $T = \{e_1, \ldots, e_n\}$ is a frame for $\mathcal{F}$. For $x \in \mathcal{R}^n$, let $|x|_{(i)}$ denote the i-th largest value of $\{|x_j|, j = 1, \ldots, n\}$, and let $\tau(x) \in \mathcal{F}$ be the vector with i-th coordinate $|x|_{(i)}$. Then τ is a maximal invariant for this example. Combining the results of Examples 2.1 and 2.2 shows that $m(\beta, \gamma) = \beta' \gamma$ for $\beta, \gamma \in \mathcal{F}$. Corollary 1 shows that $x \leqslant y$ iff $e'_i \tau(x) \leq e'_i \tau(y)$ for $i = 1, \ldots, n$ which is equivalent to $\Sigma_1^k |x|_{(i)} \leqslant \Sigma_1^k |y|_{(i)}$, $k = 1, \ldots, n$. This is usually called the sub-majorization ordering although terminology is not consistent in this case (see Marshall and Olkin (1979)).

Before discussing the next three examples, some notation is required. Given a real symmetric $p \times p$ matrix $\mathbf{x}$ let $\mu_1(\mathbf{x}) \geqslant \ldots \geqslant \mu_p(\mathbf{x})$ denote the p ordered eigenvalues of $\mathbf{x}$. Given an $n \times p$ real matrix $\mathbf{x}$, let $\lambda_1(\mathbf{x}) \geqslant \ldots \geqslant \lambda_p(\mathbf{x}) \geqslant 0$ denote the singular values of $\mathbf{x}$ (if $n < p$, then necessarily the last $p - n$ singular values are zero). Thus, $\lambda_i(\mathbf{x}) = (\mu_i(\mathbf{x}'\mathbf{x}))^{1/2}$ where $\mathbf{x}'$ is the transpose of $\mathbf{x}$. A useful result due to von Neumann (1937) and Fan (1951) is

THEOREM 1. *Let* $\mathbf{A}$ *and* $\mathbf{B}$ *be real* $n \times k$ *matrices. Then*

$$\sup_{\Gamma,\Delta} \operatorname{tr} \Gamma \mathbf{A} \Delta \mathbf{B}' = \Sigma_1^k \lambda_i(\mathbf{A})\, \lambda_i(\mathbf{B})$$

where the sup is over all $\Gamma \in \mathcal{O}_n$ and $\Delta \in \mathcal{O}_k$.

A discussion of and variations on this Theorem can be found in Marshall and Olkin (1979, p. 514).

Examples 2.4. In this example, $V = \mathcal{S}_p$—the vector space of all $p \times p$ real symmetric matrices and the inner product is

$$(\mathbf{x}_1, \mathbf{x}_2) = tr\, \mathbf{x}_1 \mathbf{x}_2$$

where tr denotes trace. The group G is the group of $p \times p$ orthogonal matrices, $\mathcal{O}_p$, and the group action is $\mathbf{x} \to \Gamma\ \mathbf{x}\ \Gamma'$ for $\Gamma \in \mathcal{O}_p$. Clearly G is a subgroup of $\mathcal{O}(V)$ for this example. Let

$$\mathcal{T} = \{\mathbf{x} | x_{11} \geqslant \ldots \geqslant x_{pp}, x_{ij} = 0 \quad \text{for } i \neq j\}$$

where x_{ij} is the i,j element of $\mathbf{x}$. A convenient choice for τ is to let $\tau(\mathbf{x})$ be the diagonal matrix in $\mathcal{S}_p$ with diagonal elements $(t(\mathbf{x}))_{ii} = \mu_i(\mathbf{x})$, $i = 1, \ldots, p$. A frame for $\mathcal{T}$ can be constructed as follows. Let $\mathbf{t}_i \in \mathcal{S}_p$ have the first i diagonal elements equal to one and all the remaining elements of $\mathbf{t}_i$ equal to zero, $i = 1, \ldots, p$. Also, let $\mathbf{t}_{p+1} = -\mathbf{t}_p$. That $T = \{\mathbf{t}_1, \ldots, \mathbf{t}_p, \mathbf{t}_{p+1}\}$ is a frame for $\mathcal{T}$ follows easily (see Example 2.2). We now claim that $m(\mathbf{u},\mathbf{y}) = (\mathbf{u},\mathbf{y})$ for $\mathbf{u},\mathbf{y} \in \mathcal{T}$. To see this, first choose δ large enough so that $\mathbf{u} + \delta\mathbf{I}$ and $\mathbf{y} + \delta\mathbf{I}$ have positive diagonal elements. Then

$$(*) \quad m(\mathbf{u},\mathbf{y}) = \sup_\Gamma tr\, \mathbf{u}\Gamma\mathbf{y}\Gamma' = \sup_\Gamma \{tr(\mathbf{u}+\delta\mathbf{I})\Gamma(\mathbf{y}+\delta\mathbf{I})\Gamma'\} - \delta tr(\mathbf{y}) - \delta tr(\mathbf{x}) + \delta^2\mathbf{p}$$
$$= m(\mathbf{u}+\delta\mathbf{I}, \mathbf{y}+\delta\mathbf{I}) - \delta tr(\mathbf{y}) - \delta tr(\mathbf{u}) + \delta^2\mathbf{p}.$$

Since $\mathbf{u}+\delta\mathbf{I}$ and $\mathbf{y}+\delta\mathbf{I}$ have positive diagonals and are in $\mathcal{T}$,

$$\lambda_i(\mathbf{u}+\delta\mathbf{I}) = \mu_i(\mathbf{u}+\delta\mathbf{I}),\ i = 1, \ldots, p$$

and the same holds for $\mathbf{y}$ in place of $\mathbf{u}$. With $n = k = p$, $\mathbf{A} = \mathbf{u}+\delta\mathbf{I}$ and $\mathbf{B} = \mathbf{y}+\delta\mathbf{I}$. Theorem 1 implies that

$$m(\mathbf{u}+\delta\mathbf{I}, \mathbf{y}+\delta\mathbf{I}) \leqslant \Sigma_1^p\, \mu_i(\mathbf{u}+\delta\mathbf{I})\, \mu_i(\mathbf{y}+\delta\mathbf{I}).$$

Since $\mu_i(\mathbf{u}+\delta\mathbf{I}) = u_{ii} + \delta$ and $\mu_i(\mathbf{y}+\delta\mathbf{I}) = y_{ii} + \delta$, there is obviously equality in the above inequality (just take $\Gamma = I$ in the definition of m). Substituting this into (*) and a bit of algebra show that $m(\mathbf{u},\mathbf{y}) = tr(\mathbf{uy}) = (\mathbf{u},\mathbf{y})$. Hence Corollary 1 is applicable and yields that $\mathbf{x} \leqslant \mathbf{y}$ iff

$$\Sigma_1^k \mu_i(\mathbf{x}) \leqslant \Sigma_1^k \mu_i(\mathbf{y}),\ k = 1, \ldots, p-1 \text{ and } \Sigma_1^p \mu_i(\mathbf{x}) = \Sigma_1^p \mu_i(\mathbf{y}).$$

In other words, $\mathbf{x} \leq \mathbf{y}$ iff the vector of eigenvalues of $\mathbf{x}$ is majorized by the vector of eigenvalues of $\mathbf{y}$. This result was established by Karlin and Rinott (1981) using a different argument.

Example 2.5. For this example, V is the vector space $\mathcal{L}_{p,n}$ of all $n \times p$ real matrices with inner product $(\mathbf{x}_1, \mathbf{x}_2) = tr(\mathbf{x}_1\mathbf{x}_2')$. For notational simplicity, it is assumed $n \geqslant p$; the contrary case is handled by a similar argument. The group G is $\mathcal{O}_n \times \mathcal{O}_p$ which acts on $\mathcal{L}_{p,n}$ by $\mathbf{x} \to \Gamma\mathbf{x}\Delta'$ for $\Gamma \in \mathcal{O}_n$ and $\Delta \in \mathcal{O}_p$. The convex cone $\mathcal{T}$ is $\mathcal{T} = \{\mathbf{x} | x_{11} \geqslant \ldots \geqslant x_{pp} \geqslant 0$, $x_{ij} = 0$ for all $i \neq j)$ where x_{ij} is the i,j element of $\mathbf{x}$. The maximal invariant τ is defined to be: $\mathbf{u} = \tau(\mathbf{y})$ is the element of $\mathcal{T}$ with $u_{ii} = \lambda_i(y)$, $i = 1, \ldots, p$. That τ is a maximal invariant is a consequence of the singular value decomposition theorem (Eckart and Young, 1939). To evaluate m on $\mathcal{T}$, consider $\mathbf{u},\mathbf{y} \in \mathcal{T}$. Then

$$m(\mathbf{u},\mathbf{y}) = \sup_{\Gamma,\Delta} tr(\mathbf{u}\Delta\mathbf{y}'\Gamma') = \Sigma_1^p \lambda_1(\mathbf{u})\, \lambda_i(\mathbf{y})$$

by Theorem 1. Since $\mathbf{u},\mathbf{y} \in \mathcal{T}$, it follows that $\lambda_i(\mathbf{u}) = u_{ii}$ and $\lambda_i(\mathbf{y}) = y_{ii}$. Thus, for

$\mathbf{u},\mathbf{y} \in \mathcal{T}$, $m(\mathbf{u},\mathbf{y}) = \Sigma_1^p \mu_{ii} y_{ii} = (\mathbf{u},\mathbf{y})$. To apply Corollary 1, we first need a frame of $\mathcal{T}$. Let $\mathbf{t}_i \in \mathcal{T}$ have its first i diagonal elements equal to one and all other elements equal to zero. Then it is easy to see that $T = \{\mathbf{t}_1, \ldots, \mathbf{t}_p\}$ is a frame for $\mathcal{T}$. A direct application of Corollary 1 shows that $\mathbf{x} \leq \mathbf{z}$ iff.

$$\Sigma_1^k \lambda_i(\mathbf{x}) \leq \Sigma_1^k \lambda_i(\mathbf{z}),\ k = 1, \ldots, p.$$

In other words, $\mathbf{x} \leq \mathbf{z}$ iff the vector of singular values of $\mathbf{x}$ is submajorized by the vector of singular values of z (see Example 2.3 for a discussion of submajorization).

Example 2.6. As in Example 2.5, take V to be $\mathcal{L}_{p,n}$ with the inner product $(\mathbf{x}_1, \mathbf{x}_2) = tr(\mathbf{x}_1\mathbf{x}_2')$ and again assume for convenience that $n \geq p$. Consider the group $G = \mathcal{O}_n$ which acts on $\mathcal{L}_{p,n}$ by $\mathbf{x} \to \Gamma\mathbf{x}$. Let $\mathcal{S}_p^+$ denote the convex cone of positive semi-definite $p \times p$ real matrices and for $\mathbf{s} \in S_p^+$ let $\mathbf{s}^{1/2}$ denote the unique element in S_p^+ which satisfies $\mathbf{s}^{1/2}\mathbf{s}^{1/2} = \mathbf{s}$. For this example, let $\mathcal{T}$ be $\mathcal{T} = \{\mathbf{x}|\mathbf{x} = \binom{\mathbf{s}}{0},\ \mathbf{s} \in \mathcal{S}_p^+\}$ and set $\tau(\mathbf{x}) = \binom{(\mathbf{x}'\mathbf{x})^{1/2}}{0} \in \mathcal{T}$. That $\tau(\mathbf{x})$ is a maximal invariant follows from Vinograde (1950). To characterize the group induced ordering, we first calculate m using Theorem 1. For $\mathbf{y} \in \mathcal{L}_{p,n}$,

$$m(\mathbf{u},\mathbf{y}) = \sup_\Gamma tr(\mathbf{u}\mathbf{y}'\Gamma') = \sup_\Gamma tr(\Gamma\mathbf{u}\mathbf{y}')$$

where the sup is over $\mathcal{O}_n$. Now, apply Theorem 1 with $n = k$, $\mathbf{A} = \mathbf{u}\mathbf{y}'$ and $\mathbf{B} = \mathbf{I}_n$ to see that $m(\mathbf{u},\mathbf{y}) = \Sigma_1^n \lambda_i(\mathbf{u}\mathbf{y}') = \Sigma_1^p \lambda_i(\mathbf{u}\mathbf{y}')$. The second equality holds since $\lambda_i(\mathbf{u}\mathbf{y}') = 0$ for $i > p$. In this example, the assumption that m restricted to $\mathcal{T}$ is the inner product, does not hold. However a description of the order can be given in terms of the Loewner ordering on $\mathcal{S}_p$. For $\mathbf{s}_1, \mathbf{s}_2 \in \mathcal{S}_p$, write $\mathbf{s}_1 \leq_L \mathbf{s}_2$ if $\mathbf{s}_2 - \mathbf{s}_1 \in \mathcal{S}_p^+$ (see Loewner (1934)).

LEMMA 1. $\mathbf{x} \leq \mathbf{y}$ *iff* $\mathbf{x}'\mathbf{x} \leq_L \mathbf{y}'\mathbf{y}$.

Proof. Assume $\mathbf{x} \leq \mathbf{y}$ so $m(\mathbf{u},\mathbf{x}) \leq m(\mathbf{u},\mathbf{y})$ for all $\mathbf{u} \in \mathcal{L}_{p,n}$.
Pick $\mathbf{u} = \alpha\beta'$ where $\alpha \in \mathcal{R}^n$, $\alpha'\alpha = 1$ and $\beta \in \mathcal{R}^p$. Then

$$m(\mathbf{u},\mathbf{y}) = \Sigma_1^p \lambda_i(\alpha\beta'\mathbf{y}') = \Sigma_1^p \mu_i^{1/2}(\alpha\beta'\mathbf{y}'\mathbf{y}'\beta\alpha') = (\beta'\mathbf{y}'\mathbf{y}'\beta)^{1/2}$$

since $\alpha\beta'\mathbf{y}'$ has rank one and $\alpha'\alpha = 1$. A similar expression holds for $\mathbf{x}$ so $(\beta'\mathbf{x}'\mathbf{x}\beta)^{1/2} \leq (\beta'\mathbf{y}'\mathbf{y}'\beta)^{1/2}$ for all $\beta \in \mathcal{R}^p$ which implies that $\mathbf{x}'\mathbf{x} \leq_L \mathbf{y}'\mathbf{y}$. Conversely, assume $\mathbf{x}'\mathbf{x} \leq_L \mathbf{y}'\mathbf{y}$ so for all $u \in \mathcal{L}_{p,n}$, $\mathbf{u}\mathbf{x}'\mathbf{x}\mathbf{u}' \leq_L \mathbf{u}\mathbf{y}'\mathbf{y}\mathbf{u}'$. This implies that (see Bellman (1960), p. 115) $\mu_i(\mathbf{u}\mathbf{x}'\mathbf{x}\mathbf{u}') \leq \mu_i(\mathbf{u}\mathbf{y}'\mathbf{y}\mathbf{u}')$, $i = 1, \ldots, p$ so $\lambda_i(\mathbf{u}\mathbf{x}') \leq \lambda_i(\mathbf{u}\mathbf{y}')$ for $i = 1, \ldots, p$. Hence $m(\mathbf{u},\mathbf{x}) \leq m(\mathbf{u},\mathbf{y})$ for all $\mathbf{u} \in \mathcal{L}_{p,n}$ so $\mathbf{x} \leq \mathbf{y}$ by Proposition 2. □

The result has a number of interesting consequences.

PROPOSITION 5. *The closed convex hull of* $\mathcal{O}_n$ *in* $\mathcal{L}_{n,n}$ *is* $\{\psi|\psi \in \mathcal{L}_{n,n}, \psi'\psi \leq_L I_n\}$.

Proof. In Lemma 1, take $n = p$ and $\mathbf{y} = \mathbf{I}_n$. Then $\mathbf{x} \leq \mathbf{I}_n$ means that $\mathbf{x}$ is in the convex hull of $\{\Gamma|\Gamma \in \mathcal{O}_n\}$ and by Lemma 1, this is equivalent to $\mathbf{x}'\mathbf{x} \leq_L I_n$. □

PROPOSITION 6. *For* $\mathbf{x},\mathbf{y} \in \mathcal{L}_{p,n}$, $\mathbf{x}'\mathbf{x} \leq_L \mathbf{y}'\mathbf{y}$ *iff.* $\mathbf{x} = \psi\mathbf{y}$ *where* $\psi \in \mathcal{L}_{n,n}$ *satisfies* $\psi'\psi \leq_L I_n$.

Proof. From Lemma 1, $\mathbf{x}'\mathbf{x} \leq_L \mathbf{y}'\mathbf{y}$ *iff* $\mathbf{x} \leq \mathbf{y}$. Also, $\mathbf{x} \leq \mathbf{y}$ *iff.* $\mathbf{x}$ is in the convex hull of $\{\Gamma\mathbf{y}|\Gamma \in \mathcal{O}_n\}$. By Proposition 5, this convex hull is just $\{\psi\mathbf{y}|\psi'\psi \leq_L I_n\}$. □

This completes our discussion of Example 2.6.

The reader should compare Examples 2.5 and 2.6 with the treatment in Jensen (1984). There is some overlap but the results do complement each other. The result in Proposition 6 is an extension of Vinograde's (1950) result and can be derived rather easily from Vinograde's result. The final example in this section is rather simple but shows that in some cases very little is gained from Proposition 2.

Example 2.7. In this example, let V be $\mathcal{R}^2$ and let G be the group generated by g_0 which is rotation through $\pi/2$ in the counterclockwise direction in $\mathcal{R}^2$. Thus, G has four elements which are $\{g_0^0, g_0, g_0^2, g_0^3\}$ and G is Abelian. Take $\mathcal{J}$ to be

$$\mathcal{J} = \{\mathbf{x}|x_1 > 0, x_2 \geq 0\} \cup \{0\}$$

and let $\tau(\mathbf{x})$ be the unique vector in $\mathcal{J}$ of the form $g_0^i\mathbf{x}$ for $i = 0, 1, 2, 3$. The calculation of m is easy and of very little help. About all one can say here is that $\mathbf{x} \leq \mathbf{y}$ iff $\mathbf{x}$ is in the convex hull of $\{g_0^i y|\ i = 0, 1, 2, 3\}$. Exactly the same remarks are in order when G is the group generated by the rotation through $2\pi/k$ ($k = 3, 4, \ldots$). Namely, m is easy to describe, but of no help in describing the ordering. Note that Proposition 4 (and hence Corollary 1) cannot be used in this example. However, if instead of a finite rotation group, we use a finite dihedral group (see Benson and Grove (1971), p. 7) acting on $\mathcal{R}^2$, then with the obvious choice for $\mathcal{J}$, Proposition 4 and Corollary 1 do apply directly.

3. The Decreasing Functions. In this section, we apply results of Marshall, Walkup and Wets (1967) to describe the decreasing functions of some of the group induced orderings discussed earlier. As in the last section, it is assumed that $(V, (\cdot,\cdot))$ is an inner product space acted on by a closed group $G \subseteq \mathcal{O}(V)$, τ is a maximal invariant function whose range is the convex cone $\mathcal{J}$ with $\tau(x) \in \{gx|g \in G\}$. The problem considered here is to give a useful analytic condition on $f{:}V \to R$ so that $x \leq y$ implies that $f(x) \geq f(y)$.

A solution to the problem just described will be given in the case that, when restricted to $\mathcal{J}$, the group induced ordering $\leq$ is a cone ordering. To be more precise, let D be a subset of an inner product space $(W, (\cdot,\cdot))$ and let $K \subseteq W$ be a fixed convex cone. A *cone ordering* induced on D by K is a relation $<$ defined by $x < y$ iff $y - x \in K$. A function $f{:}D \to R$ is decreasing on $(D, <)$ if $x < y$ implies $f(x) \geq f(y)$, for $x,y \in D$.

THEOREM 2 (**Marshall, Walkup and Wets (1967)**). *Suppose D is convex with a non-empty interior and $f{:}D \to R$ is continuous at the boundary of D. Let T be a frame for K. Then f is decreasing on $(D, <)$ iff.*

(3.1) $\quad f(x + \lambda t) \leq f(x)$ *for all $x \in D$ and all $t \in T$ and $\lambda > 0$ such that $x + \lambda t \in D$.*

Corollary 2 (Marshall, Walkup, Wets (1967). *In addition to the assumptions in Theorem 1, assume that f has a differential $df{:}D^\circ \to W$ on the interior of D. Then f is decreasing on $(D, <)$, iff*

$$(df(x), t) \leq 0 \tag{3.2}$$

for all $t \in T$ and $x \in D^\circ$.

Before applying these results to the problem at hand, a few preliminaries are needed. Given the convex cone $\mathcal{J}$ which is the range of τ, let M be the subspace of V which is generated by $\mathcal{J}$. Thus, $\mathcal{J}$ is a convex cone with a non-empty interior in the inner product space $(M, (\cdot,\cdot))$. Also, let

$$\mathcal{J}^* = \{x|x \in M, (x,y) \geq 0 \quad \text{for all } y \in \mathcal{J}\}$$

so $\mathcal{J}^*$ is the dual cone (in M) of $\mathcal{J}$. Of course, $\mathcal{J}^*$ is also a convex cone. The following result shows that the group induced ordering is in fact a cone ordering on $\mathcal{J}$ in the special case treated in Proposition 4.

PROPOSITION 7. *For β, $\gamma \in \mathcal{J}$, suppose that $m(\beta, \gamma) = (\beta, \gamma)$. Then, for $u,v \in \mathcal{J}$, $u \leq v$ iff $v - u \in \mathcal{J}^*$.*

Proof. Since $u, v \in \mathcal{J}$, $\tau(u) = u$ and $\tau(v) = v$. Thus, Proposition 4 implies that $u \leqslant v$ iff

$$(\beta, u) \leqslant (\beta, v) \quad \text{for all } \beta \in \mathcal{J} \text{ which holds iff.} \tag{3.3}$$

$$(\beta, v-u) \geqslant 0 \quad \text{for all } \beta \in \mathcal{J}, \tag{3.4}$$

and this is equivalent to the assertion that $v - u \in \mathcal{J}^*$. □

The above result gives a sufficient condition that $\leqslant$ be a cone ordering on $\mathcal{J}$. Further the cone ordering is determined by the convex cone $\mathcal{J}^*$. In examples 2.1 through 2.5, m on $\mathcal{J} \times \mathcal{J}$ is the inner product so Proposition 7 applies directly.

PROPOSITION 8. *Assume that the pre-ordering on $\mathcal{J}$ is the cone ordering defined by $\mathcal{J}^*$. Let T^* be a frame for $\mathcal{J}^*$. Then a function $f: V \to R^1$ which is continuous at the boundary of $\mathcal{J}$ is decreasing iff for each $x \in \mathcal{J}$ and $t \in T^*$,*

$$f(x + \lambda t) \leqslant f(x)$$

for all $\lambda > 0$ such that $x + \lambda t \in \mathcal{J}$.

Proof. Apply Theorem 2 with $D = \mathcal{J}$ and $K = \mathcal{J}^*$. □

COROLLARY 3. *Let the assumptions of Proposition 8 hold. Also assume that f has a differential $df: \mathcal{J}^0 \to M$. Then f is decreasing iff. for all $t \in T^*$,*

$$(df(x), t) \leqslant 0 \tag{3.4}$$

for $x \in \mathcal{J}^*$.

Proof. This is immediate from Corollary 2. □

The application of Corollary 3 to Examples 2.1, 2.2, and 2.3 is quite easy and the essential details can be found in Marshall, Walkup and Wets (1967). A discussion of Example 2.4 is much the same as that for Example 2.5 which we now give.

Example 2.5 continued. The notation and results given in Example 2.5 are assumed. First, the subspace M generated by $\mathcal{J}$ is the space of all $n \times p$ real matrices $\mathbf{u}$ with $u_{ij} = 0$ for $i \neq j$, so M is p dimensional. The dual cone $\mathcal{J}^* \subseteq M$ is

$$\mathcal{J}^* = \{\mathbf{u} | \mathbf{u} \in M, \Sigma_1^k u_{ii} \geqslant 0, k = 1, \ldots, p\}.$$

A frame T^* for $\mathcal{J}^*$ consists of $t_1, \ldots, t_p \in M$ where:

> t_i has its i,i diagonal 1, its $(i+1), (i+1)$ diagonal -1 and all other elements of t_i are 0, for $i = 1, \ldots, p-1$;
>
> t_p has its p,p diagonal 1 and all other elements 0.

This follows from Proposition 1 in Marshall, Walkup and Wets (1967). Let $f: \mathcal{L}_{p,n} \to R^1$ be a $\mathcal{O}_n \times \mathcal{O}_p$ invariant function and let $\bar{f}$ denote the restriction of f to $\mathcal{J}$. When $\bar{f}$ has a differential, then f is decreasing iff

$$(t, d\bar{f}(u)) \leqslant 0, t \in T^* \text{ and } u \in \mathcal{J}$$

which is equivalent to

$$\delta\bar{f}/\delta u_{11} \leqslant \delta\bar{f}/\delta u_{22} \leqslant \ldots \leqslant \delta\bar{f}/\delta u_{pp} \leqslant 0.$$

In Example 2.6, the group ordering is not an $\mathcal{J}^*$ cone ordering on $\mathcal{J}$, but is an $\mathcal{J}^*$ cone ordering in a different coordinate system. To be more precise, Lemma 1 shows that $x \leqslant y$ iff. $x'x \leqslant_L y'y$. The Loewner ordering $\leqslant_L$ is a cone ordering on $\mathcal{S}_p^+$. Thus, a decreasing function f on $\mathcal{L}_{p,n}$ in Example 2.6 can be characterized by first writing it as $f(x) = \bar{f}(x'x)$ and then using the Marshall, Walkup and Wests ((1967), Example 4) results.

4. Remarks on the Convolution Theorem. Again consider the general situation of an inner product space $(V(\cdot,\cdot))$ acted on by a compact group $G \subseteq \mathcal{O}(V)$. As usual, $\leq$ denotes the pre-order defined by G.

Definition 2. If for every two compact monotone sets A and B, the function

$$\psi(y) = \int_V I_A(x) I_B(y-x)dx$$

is decreasing (see Definition 1), then we say the convolution theorem (CT) holds for G.

It is a standard approximation argument to show that CT implies that for suitably smooth, integrable and decreasing f_1, f_2, the convolution

$$f(y) = (f_1 * f_2)(y) = \int_V f_1(x) f_2(y-x)dx$$

is again decreasing. Hence the term convolution theorem. This result has many applications in the area of probability inequalities—for example, see Marshall and Olkin (1974), Eaton and Perlman (1977), Marshall and Olkin (1979) and Eaton (1982).

CT was established for $V = R^n$ and $G = \mathcal{P}_n$ by Marshall and Olkin (1974). This result was extended to all reflection groups by Eaton and Perlman (1977). Examples of reflection groups are the groups considered in Examples 2.1, 2.2 and 2.3. When the group G acts transitively on $\{x|x \in V, \|x\| = 1\}$, then $x \leq y$ means that $\|x\| \leq \|y\|$ and all the decreasing functions have the form $x \to \eta(\|x\|)$ where η is decreasing on $[0, \infty)$. CT obviously holds for such cases. In summary, here is a listing of some groups for which CT is known to hold:

(i) All finite and infinite closed reflection groups (see Eaton and Perlman (1977)).

(ii) Any group G which acts transitively on $\{x|x \in V, \|x\| = 1\}$.

(iii) A product $G_1 \times G_2 \times \ldots \times G_k$ acting on the direct sum $V_1 \otimes V_2 \otimes \ldots \otimes V_k$. The action is coordinatewise, $(g_1, g_2, \ldots, g_k)(x_1, x_2, \ldots, x_k) = (g_1x_1, g_2x_2, \ldots, g_kx_k)$, where G_i acting on V_i is of the type (i) or (ii) above.

These are the only groups that I know for which CT holds. The remainder of this section is devoted to a discussion of a necessary condition on G in order that CT hold. Some examples are given where CT does not hold.

Recall that $x \leq y$ means $x \in C(y)$ where $C(y)$ is the convex hull of $\{gy|g \in G\}$. Also, a set B is monotone iff for all $x \in B$, $C(x) \subseteq B$.

Definition 3. Given any set A, let

$$\mathcal{S}(A) = \bigcup_{x \in A} C(x).$$

PROPOSITION 9. *The set $\mathcal{S}(A)$ is the smallest monotone set which contains A.*

Proof. To show $\mathcal{S}(A)$ is monotone, consider $u \in \mathcal{S}(A)$ so $u \in C(x)$ for some $x \in A$. Since $C(x)$ is monotone, $C(u) \subseteq C(x)$ so $C(u) \subseteq \mathcal{S}(A)$ which shows $\mathcal{S}(A)$ is monotone. Now, assume $B \supseteq A$ and B is monotone. If $x \in A$ then $x \in B$ so $C(x) \subseteq B$ as B is monotone. Hence $\bigcup_{x \in A} C(x) \subseteq B$. □

Here are some properties of $\mathcal{S}$ which are easily verified:

(i) $\mathcal{S}(\bigcup_\alpha A_\alpha) = \bigcup_\alpha \mathcal{S}(A_\alpha)$

(4.1) (ii) $\mathcal{S}(A_1 + A_2) \subseteq \mathcal{S}(A_1) + \mathcal{S}(A_2)$

(iii) A compact implies $\mathcal{S}(A)$ compact

In (4.1), the sign + denotes the usual Minkowski sum of two sets.

Next is a necessary condition for CT to hold.

PROPOSITION 10. *Assume that for each $x \neq 0$, $C(x)$ has a non-empty interior (For a discussion of this condition, see Eaton and Perlman (1977)). In order that CT hold, it is necessary that*

(4.2) $A + y \subseteq \mathcal{S}(A + x)$ for all $y \in C(x)$, for all $x \in V$, and for all compact monotone sets A.

Proof. Assume that (4.2) does not hold for some x, $y \in C(x)$ and A. Then, A must have a non-zero element and x must be non-zero. Let $z \in A$ with $z \neq 0$. Since $\phi \neq (C(z))^\circ \subseteq A$, the set A has a non-empty interior. Hence $A + y$ has a non-empty interior and the open set $N = (A + y)^\circ \cap (\mathcal{S}(A + x))^C$ is not empty. With l denoting Lebesgue measure, we have

$$l(A) = l(A + y) > l((A + y) \cap \mathcal{S}(A + x)),$$

since N is open and non-empty. For $u \in V$ let

$$\Psi(u) = \int I_{\mathcal{S}(A+x)}(w)\, I_A(w-u)\, dw.$$

Since $A + x \subseteq \mathcal{S}(A + x)$,

$$\Psi(x) = l((A + x) \cap \mathcal{S}(A + x)) = l(A + x) = l(A).$$

However,

$$\Psi(y) = l((A+y) \cap \mathcal{S}(A+x)) < l(A)$$

so $\Psi(y) < \Psi(x)$ and CT does not hold. □

PROPOSITION 11. *Each of the following conditions is equivalent to (4.2):*

(4.3) $A + C(x) \subseteq \mathcal{S}(A+x)$ for all $x \in V$ and for all compact monotone sets A,

(4.4) $C(z) + C(x) \subseteq \mathcal{S}(C(z) + x)$ for all $x, z \in V$,

(4.5) $\mathcal{S}(C(z) + x)$ is a convex set for all $x, z \in V$.

Proof. Clearly (4.3) implies (4.2). Conversely, if (4.2) holds, then

$$A + C(x) = \bigcup_{y \in C(x)} (A + y) \subseteq \mathcal{S}(A + x)$$

so (4.3) holds. Clearly (4.3) implies (4.4). To show (4.4) implies (4.3) first observe that when A is a monotone set,

$$A + C(x) = \bigcup_{z \in A} (C(z) + C(x)).$$

Since A is monotone, $A = \bigcup_{z \in A} C(x)$ so (4.1)(i) and (4.4) imply that

$$\mathcal{S}(A + x) = \mathcal{S}(\bigcup_{z \in A}(C(z) + C(x)) = \bigcup_{z \in A} \mathcal{S}(C(z) + x) \supseteq \bigcup_{x \in A}(C(z) + C(x)) = \mathbf{A} + C(x).$$

Hence (4.3) holds. To show (4.4) and (4.5) are equivalent, first assume (4.4) holds. Since $C(z) + C(x)$ is monotone, Proposition 9 implies that

(*) $$\mathcal{S}(C(z) + x) \subseteq C(z) + C(x).$$

Thus, when (4.4) holds there is equality in (*). But $C(z) + C(x)$ is convex as both $C(z)$ and $C(x)$ are convex so (4.5) holds. Conversely, assume that (4.5) holds and consider $u \in C(z)$ and $v \in C(x)$. It must be shown that $u + v \in \mathcal{S}(C(z) + x)$. Since $\mathcal{S}(C(z) + x) = \mathcal{S}(C(z) + gx))$ for all $g \in G$, it follows that $u + gx$ is in $\mathcal{S}(C(z) + x)$ for all $g \in G$ as $u \in C(z)$. But, if $\mathcal{S}(C(z) + x)$ is convex, this implies that all convex combinations (over $g \in G$) of $u + gx$ are also in $\mathcal{S}(C(z) + x)$. Since $v \in C(x)$, v can be represented as a convex combination of $\{gx | g \in G\}$ so $u + v$ is a convex combination of $\{u + gx | g \in G\}$. Hence (4.4) holds. □

The following example shows that CT does not hold for any finite rotation group acting on $\mathcal{R}^2$. It will be shown that condition (4.5) does not hold for these cases.

Example 4.1. Fix an integer $k \geqslant 3$ and let $\theta = 2\pi/k$. The case of $k = 2$ is trivial. Let G be the group generated by $g = g_\theta$ which is rotation (in the counter-clockwise direction) through the angle θ. Thus $G = \{I, g, \ldots, g^{k-1}\}$ has k elements. Let $z = \binom{1}{0}$ and let x be

$$x = 5(g_\eta z)$$

where $\eta = \theta/2$, so x has length 5. Then $u = z + x \in C(z) + x$ and has coordinates

$$\begin{pmatrix} u_1 \\ u_2 \end{pmatrix} = \begin{pmatrix} x_1 + 1 \\ x_2 \end{pmatrix}.$$

Applying g^{k-1} to the set $C(z) + x$ shows that the vector

$$\bar{u} = \begin{pmatrix} \bar{u}_1 \\ \bar{u}_2 \end{pmatrix} = \begin{pmatrix} x_1 + 1 \\ -x_2 \end{pmatrix} = g^{k-1}(x + gz)$$

is in the set $\mathcal{S}(C(z) + x)$. Hence, if $\mathcal{S}(C(z) + x)$ is to be convex, the vector

$$v = \frac{u + \bar{u}}{2} = \begin{pmatrix} x_1 + 1 \\ 0 \end{pmatrix}$$

must be in $\mathcal{S}(C(z) + x)$. However, a carefully drawn picture will convince the reader that v is not in $C(w)$ for any $w \in C(z) + x$. The case of $k = 4$ is a good starting point to see why $\mathcal{S}(C(z) + x)$ is not convex for the particular choices of z and x above (see Figure 1). Thus CT does not hold for any of the finite rotation groups acting on $\mathcal{R}^2$. However, CT does hold for the finite dihedral groups acting on $\mathcal{R}^2$ as these are reflection groups (see Benson and Grove (1971)).

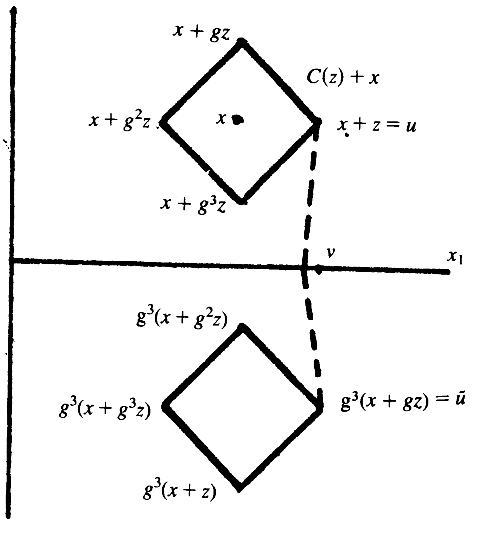


FIGURE 1. Case of $k = 4$. The dashed line gives the right most boundary of $\mathcal{S}(C(z) + x)$.

The final result of this section shows that the necessary condition (4.2) is satisfied for the situation considered in Proposition 4. More precisely, again assume that τ is a maximal invariant with a convex cone $\mathcal{T}$ as its range and $\tau(x) \in \{gx \mid g \in G\}$.

PROPOSITION 12. As in Proposition 4, assume that for $u, v \in \mathcal{T}$, $m(u,v) = (u,v)$. Then, for $u, v \in \mathcal{T}$

$$C(u) + C(v) = C(u + v) \tag{4.6}$$

and condition (4.2) holds.

Proof. Since $u + v \in C(u) + C(v)$ and since $C(u) + C(v)$ is monotone, it is clear that $C(u + v) \subseteq C(u) + C(v)$. Now, suppose $z \in C(u) + C(v)$ so $z = \gamma + \delta$ with $\gamma \in C(u)$ and $\delta \in C(v)$. Using the relations given by (2.5) and the results of Propositions 2, 3 and 4, we have

$$(4.7) \quad m(w,z) = m(w,\gamma + \delta) \leq m(w,\gamma) + m(w,\gamma) \leq m(w,u) + m(w,v) = m(\tau(w),u) + m(\tau(w),v) = (\tau(w), u + v) = m(\tau(w), u + v) = m(w,u + v)$$

for any $w \in V$. By Proposition 2, this implies that $z \leq u + v$ so $z \in C(u + v)$. Hence $C(u) + C(v) \subseteq C(u + v)$ so (4.6) holds. To show (4.2) holds, (4.4) will be verified. First observe that $C(z) = C(\tau(z))$ and

$$\mathcal{S}(C(z) + x) = \mathcal{S}(C(\tau(z)) + \tau(x)) \quad \text{for } z,x \in V.$$

Thus, by (4.6), we have

$$\begin{aligned} &C(z) + C(x) = C(\tau(z)) + C(\tau(x)) = \\ &C(\tau(z) + \tau(x)) \subseteq \mathcal{S}(C(\tau(z)) + \tau(x)) = \\ &\mathcal{S}(C(z) + x) \end{aligned}$$

so (4.4) and hence (4.2) holds. □

The above result shows that (4.2) holds for Examples 2.1–2.5 although CT is only known to hold for Examples 2.1–2.3. It is not known whether (4.2) holds for Example 2.6. Whether or not CT holds for Example 2.5 is an important unresolved problem.

The implications of (4.6) concerning the group G are not known, but are probably important in understanding when CT holds. Both these implications and useful conditions for CT to hold would be welcome contributions.

REFERENCES

BELLMAN, R. (1960). *Introduction to Matrix Analysis*. McGraw-Hill, New York.

BENSON, C. T. and GROVE, L. C. (1971). *Finite Reflection Groups*. Bogdon and Quigley. Terrytown-on-Hudson, New York.

EATON, M. L. (1982). A review of selected topics in multivariate probability inequalities. *Ann. Statist.* *10* 11–43.

EATON, M. L. and PERLMAN, M. (1977). Reflection groups, generalized Schur functions, and the geometry of majorization. *Ann. Prob.* *5* 829–860.

ECKART, C. and YOUNG, G. (1939). A principle axis transformation for non-Hermitian matrices. *Bull. Amer. Math. Soc.* *45* 118–121.

FAN, K. (1951). Maximum properties and inequalities for the eigenvalues of completely continuous operators. *Proc. Nat. Acad. Sci. U.S.A.* *37* 760–766.

HARDY, G. H., LITTLEWOOD, J. E., and PÓLYA, G. (1934, 1952). *Inequalities,* 1st ed., 2nd ed. Cambridge University Press, London and New York.

JENSEN, D. R. (1984). Invariant ordering and order preservation. In *Inequalities in Statistics and Probability,* Y. L. Tong, ed., Institute of Mathematical Statistics, Hayward, CA., 26–34.

KARLIN, S. and RINOTT, Y. (1981). Entropy inequalities for classes of probability distributions II. The multivariate case. *Adv. Appl. Prob.* *13* 325–351.

LEHMANN, E. L. (1959). *Testing Statistical Hypotheses*. Wiley, New York.

LOEWNER, C. (K. Löwner) (1934). Über monotone Matrixfunctionen. *Math. Z.* *38* 177–216.

MARSHALL, A. W. and OLKIN, I. (1974). Majorization in multivariate distributions. *Ann. Statist.* *2* 1189–1200.

MARSHALL, A. W. and OLKIN, I. (1979). *Inequalities: Theory of Majorization and Its Applications*. Academic Press, New York.

MARSHALL, A. W., WALKUP, D. W., and WETS, R. J. B. (1967). Order preserving functions: applications to majorization and order statistics. *Pacific J. Math.* *23* 569–584.

von Neumann, J. (1937). Some matrix-inequalities and metrization of a matrix space. *Tomsk. Univ. Rev. 1* 286–300.

Ostrowski, A. M. (1952). Sur quelques applications des fonctiones convexes et concaves au sens de I. Schur. *J. Math. Pures. Appl.* [9] *31* 253–292.

Rado, R. (1952). An Inequality. *J. London Math. Soc. 27* 1–6.

Schur, I. (1923). Uber eine Klasse von Mittelbildungen mit Anwendungen die Determinanten. *Theorie Sitzungsber. Berlin. Math. Gesellschaft 22* 9–20.

Vinograde, B. (1950). Canonical positive definite matrices under internal linear transformations. *Proc. Amer. Math. Soc. 1* 159–161.

Inequalities in Statistics and Probability
IMS Lecture Notes–Monograph Series Vol. 5 (1984), 26–34

INVARIANT ORDERING AND ORDER PRESERVATION[1]

BY D. R. JENSEN
Virginia Polytechnic Institute and State University

Suppose $\mathcal{G}$ is a group of one-to-one transformations of a set $\mathcal{X}$ onto $\mathcal{X}$, M is maximal invariant taking values in an ordered set $(\mathcal{M}, \succeq_M)$, and $\succeq$ is an ordering on $\mathcal{X}$ induced from $\succeq_M$. Properties of $(\mathcal{X}, \succeq)$ are studied in Part I including lattice properties and order preservation. Examples include an ordering on $\mathcal{F}_{n\times m}$ having properties of Loewner's (1934) ordering for Hermitian varieties, a unitary ordering on $\mathcal{F}_{n\times m}$ giving a lattice, and orderings on $\mathcal{R}^n$ invariant under various groups. Applications to a variety of problems in statistics and applied probability are given in Part II..

PART I. THEORY

1. Introduction. Many problems exhibit symmetries as invariance under a group $\mathcal{G}$ acting on a set $\mathcal{X}$. Invariance principles require that solutions be invariant, and reduction by invariance preserves essentials while discarding irrelevant details. Because order relations often assume a prominent role in the analysis of such problems, it is instructive to consider orderings symmetric under $\mathcal{G}$.

Often $\mathcal{X}$ is finite-dimensional; examples are the Euclidean space $\mathcal{R}^n$, the linear space $\mathcal{F}_{n\times m}$ of $(n\times m)$ matrices over the complex field $\mathcal{C}$, the Hermitian $(n\times n)$ matrices $\mathcal{H}_n$, and the cone $\mathcal{H}_n^+$ of positive semidefinite Hermitian varieties. Typical groups of transformations are the classical groups. An ordering on $\mathcal{H}_n$ in wide usage was studied by Loewner (1934) in which $\mathbf{A} \succeq_L \mathbf{B}$ on $\mathcal{H}_n$ if and only if $\mathbf{A} - \mathbf{B} \in \mathcal{H}^+_n$. This ordering is invariant under the general linear group $Gl(n)$ acting on $\mathcal{H}_n$ by congruence, for $\mathbf{A} \succeq_L \mathbf{B}$ on $\mathcal{H}_n$ if and only if $\mathbf{CAC^*} \succeq_L \mathbf{CBC^*}$ on $\mathcal{H}_n$ for every $\mathbf{C} \in Gl(n)$, with $\mathbf{C^*}$ the conjugate transpose of $\mathbf{C}$. The relation $\succeq_L$ as an ordering on $\mathcal{F}_{n\times n}$ was considered by Hartwig (1976).

Here we study symmetric orderings induced through maximal invariants, the preservation of such orderings, and the possible transitivity of lattice properties. Our principal motivation stems from needed orderings on all of $\mathcal{F}_{n\times m}$ and not just $\mathcal{F}_{n\times n}$ or its Hermitian varieties.

2. The Basic Results. A set $\mathcal{X}$ together with a binary relation $\succeq$ is said to be *linearly ordered* if the relation is reflexive, transitive, antisymmetric, and complete. The relation is a *partial ordering* if it is reflexive, transitive, and antisymmetric, and a *preordering* if it is reflexive and transitive. A partially ordered set $(\mathcal{X}, \succeq)$ is a *lower semi-lattice* if for any two elements x,y there is an element $v = x \wedge y \in \mathcal{X}$ that is a greatest lower bound for x,y; an *upper semi-lattice* if there is a least upper bound $u = x \vee y$ for x,y in $\mathcal{X}$; and a *lattice* if it is both a lower and upper semi-lattice.

Let $\mathcal{G}$ be a group of one-to-one transformations from $\mathcal{X}$ onto $\mathcal{X}$. A function f on $\mathcal{X}$ is said to be *invariant* under $\mathcal{G}$ if, for any $(x,g) \in \mathcal{X} \times \mathcal{G}$, $f(gx) = f(x)$, and to be *maximal invariant* if it is invariant and if $f(x) = f(y)$ implies $y = gx$ for some $g \in \mathcal{G}$. The $\mathcal{G}$-*orbit* of $x_0 \in \mathcal{X}$

[1] This work was supported in part by the U.S. Army Research Office through Grant No. DAAG–29–78–G–0172.

AMS 1980 subject classifications. Primary, 06A10; Secondary, 15A45.

Key words and phrases: Invariant orderings, maximal invariants, monotone functions, orderings on $\mathcal{F}_{n\times m}$, Moore-Penrose inverses, applications.

is the set $0(x_0) = \{x \in \mathcal{X} | x = gx_0, g \in \mathcal{G}\}$ in $\mathcal{X}$. An invariant function is constant on each orbit; a maximal invariant function takes distinct values on different orbits. The orbits are equivalence classes under the relation $y \equiv x \bmod \mathcal{G}$, in one-to-one correspondence with the values of a maximal invariant. If the image $\mathcal{V}$ of $\mathcal{X}$ under f has a partial ordering $\succeq_V$, the mapping f is said to be *monotone* if f is order-preserving, i.e. if $x \succeq y$ on $\mathcal{X}$ implies $f(x) \succeq_V f(y)$ on $\mathcal{V}$. In particular, a real function ϕ is monotone on $(\mathcal{X}, \succeq)$ if $x \succeq y$ implies $\phi(x) \geq \phi(y)$; if ϕ is a norm it is a *monotone norm*.

Henceforth $M(x)$ on $\mathcal{X}$ is a maximal invariant function with range $\mathcal{M}$. If in addition $(\mathcal{M}, \succeq_M)$ is ordered, then a binary relation on $\mathcal{X}$ may be induced as follows.

Definition 1. Let $M: \mathcal{X} \to \mathcal{M}$ be maximal invariant under $\mathcal{G}$. If $(\mathcal{M}, \succeq_M)$ is ordered, then x,y are said to be related as $x \succeq y$ on $\mathcal{X}$ if and only if $M(x) \succeq_M M(y)$ on $\mathcal{M}$.

The following theorem is basic. It asserts that the induced relation on $\mathcal{X}$ is an ordering, that this ordering is invariant, that $(\mathcal{X}, \succeq)$ may inherit lattice properties from $(\mathcal{M}, \succeq_M)$, and that the functions monotone on $(\mathcal{X}, \succeq)$ may be characterized. That a property holds up to equivalence means that the conventional definition applies when elements on the same $\mathcal{G}$-orbit are identified.

THEOREM 1. *Let $\succeq$ be a binary relation on $\mathcal{X}$ induced as in Definition 1*. (i) *The relation is invariant in the sense that $x \succeq y$ on $\mathcal{X}$ if an only if $gx \succeq g'y$ for any $g,g' \in \mathcal{G}$.* (ii) *If $(\mathcal{M}, \succeq_M)$ is partially ordered, then $(\mathcal{X}, \succeq)$ is preordered and is antisymmetric up to equivalence*. (iii) *If $(\mathcal{M}, \succeq_M)$ is completely ordered, then the ordering $\succeq$ on $\mathcal{X}$ is complete up to equivalence*, (iv) *If $(\mathcal{M}, \succeq_M)$ is a lower or upper semi-lattice, then $(\mathcal{X}, \succeq)$ is a lower or upper semi-lattice, respectively, up to equivalence*. (v) *A real function f is monotone on $(\mathcal{X}, \succeq)$ if and only if f is a composition of the type*

$$f(x) = \psi(M(x)) = [\psi \circ M](x) \tag{2.1}$$

with ψ a function from the class Ψ of all functions monotone on $(\mathcal{M}, \succeq_M)$.

Proof. For conclusions (i)–(iv) and sufficiency in (v) argue orbit by orbit. The necessity of (v) follows from a result on page 216 of Lehmann (1959), i.e. $f(x) = [\xi \circ M](x)$ for some function ξ on $\mathcal{M}$, together with the monotonicity of f. □

Often there is wide latitude in the choice of an invariant ordering. If $\mathcal{M}$ is a vector space, $\mathcal{K} \subset \mathcal{M}$ a cone, and if $M_1 \succeq_M M_2$ is equivalent to $M_1 - M_2 \in \mathcal{K}$, then $\succeq_M$ is a preordering; $(\mathcal{M}, \succeq_M)$ is partially ordered if and only if $\mathcal{K} \cap (-\mathcal{K}) = \{0\}$ (cf. Wong and Ng (1973)). For cone orderings on $\mathcal{M}$ the functions Ψ monotone on $(\mathcal{M}, \succeq_M)$ are characterized in Marshall, Walkup, and Wets (1967). Elsewhere in this volume Eaton (1984) requires that $x \succeq_E y$ on $\mathcal{X}$ if and only if y lies in the convex hull of the $\mathcal{G}$-orbit of x; given an inner product on $\mathcal{X}$, he characterizes this ordering by quasi-linearization in terms of a maximal invariant function. From the correspondence of orbits to points in $\mathcal{M}$, it is clear that both $\mathcal{X}$ and $\mathcal{M}$ may be ordered using $\succeq_E$, the latter depending only on $\mathcal{G}$ and leaving no latitude for choice. Although that approach and the approach taken here differ, there is common ground as may be seen on comparing our examples with those of Eaton (1984).

We now specialize $\mathcal{X}$ and $\mathcal{G}$ to familiar finite-dimensional linear spaces and the classical groups, respectively. On occasion properties of these spaces support results beyond those of Theorem 1.

3. Orderings on $\mathcal{R}^n$ and $\mathcal{F}_{n \times m}$. Let $\mathcal{R}_+^n$ be the positive orthant of $\mathcal{R}^n$; write $\mathcal{D}_n = \{\mathbf{x} \in \mathcal{R}^n | x_1 \geq x_2 \geq \ldots \geq x_n\}$ and $\mathcal{D}_n^+ = \mathcal{D}_n \cap \mathcal{R}_n^+$; and henceforth consider $\mathcal{F}_{n \times m}$ with $n \geq m$. Denote by $\mathcal{U}(n)$ the unitary $(n \times n)$ matrices and by $\mathcal{S}(n,m)$ the Stiefel manifold in

$\mathcal{F}_{n\times m}$ whose elements satisfy $\mathbf{A}^*\mathbf{A} = \mathbf{I}_m$. The *polar factorization* of $\mathbf{A} \in \mathcal{F}_{n\times m}$ is $\mathbf{A} = \mathbf{LS}$ with $\mathbf{L} \in \mathcal{S}(n,m)$ and $\mathbf{S} = (\mathbf{A}^*\mathbf{A})^{1/2}$ as the Hermitian square root. Its *singular decomposition* is $\mathbf{A} = \mathbf{PD}_\alpha\mathbf{Q}^*$ with $\mathbf{P} \in \mathcal{S}(n,m)$, $\mathbf{Q} \in \mathcal{U}(m)$, and $\mathbf{D}_\alpha = \text{Diag}(\alpha_1, \ldots, \alpha_m)$, a real diagonal matrix of the ordered *singular values* of $\mathbf{A}$, i.e. the non-negative square roots of the characteristic values of $\mathbf{A}^*\mathbf{A}$. Let σ: $\mathcal{F}_{n\times m} \to \mathcal{D}^+_m$ map $\mathbf{A}$ into its ordered singular values.

The following construction is essentially due to von Neumann (1937). For functions γ: $\mathcal{R}^m \to \mathcal{R}^1$ consider the properties

P1. $\gamma(\epsilon_1 x_{i_1}, \ldots, \epsilon_m x_{i_m}) = \gamma(x_1, \ldots, x_m)$, where $\{\epsilon_i = \pm 1; 1 \leq i \leq m)$ and $(i_1, i_2, \ldots, i_m)$ is any permutation of $(1, 2, \ldots, m)$;

P2. If $\{|x_i| \leq |y_i|, 1 \leq i \leq m\}$, then $\gamma(x_1, \ldots, x_m) \leq \gamma(y_1, \ldots, y_m)$ with strict inequality if $|x_i| < |y_i|$ for some i.

P3. $\gamma(x_1, \ldots, x_m) \geq 0$;

P4. $\gamma(cx_1, \ldots, cx_m) = |c|\gamma(x_1, \ldots, x_m)$.

P5. $\gamma(x_1 + y_1, \ldots, x_m + y_m) \leq \gamma(x_1, \ldots, x_m) + \gamma(y_1, \ldots, y_m)$.

Definition 2. Let Γ be the class of functions γ: $\mathcal{R}^m \to \mathcal{R}^1$ having properties P1 and P2, and let $\Gamma_0 \subset \Gamma$ have the additional properties P3, P4 and P5.

Definition 3. Let Φ be the class of functions ϕ: $\mathcal{F}_{n\times m} \to \mathcal{R}^1$ generated by compositions as $\Phi = \{\phi | \phi = \gamma \circ \sigma, \gamma \in \Gamma\}$. Let Φ_0 be the subclass of functions in Φ such that $\phi_0 = \{\phi | \phi = \gamma \circ \sigma, \gamma \in \Gamma_0\}$.

Functions in Γ_0 are the symmetric gauge functions on $\mathcal{R}^m$, and those in Φ_0 are the unitarily invariant norms on $\mathcal{F}_{n\times m}$; von Neumann (1937) showed that these classes generate each other (cf. also Schatten (1970)).

3.1 Orderings on $\mathcal{R}^n$. If $\mathcal{G}$ is the group of 2^n reflections about the coordinate planes in $\mathcal{R}^n$, then the orbit of $\mathbf{x} \in \mathbb{R}^n$ has the vertices $\{(\epsilon_1 x_1, \ldots, \epsilon_n x_n); \epsilon_i = \pm 1, 1 \leq i \leq n\}$ of a parallelotope, and a maximal invariant is $M(\mathbf{x}) = (|x_1|, |x_2|, \ldots, |x_n|)$ with range $\mathcal{R}^n_+$. If ordered by coordinates such that $\mathbf{u} \succeq_M \mathbf{v}$ on $\mathcal{R}^n_+$ if and only if $\{u_i \geq v_i; i = 1, \ldots, n\}$, the range $(\mathcal{R}^n_+, \succeq_M)$ is a lattice (cf. Vulikh (1967), for example). The order induced by Definition 1 is that $\mathbf{x} \succeq \mathbf{y}$ on $\mathcal{R}^n$ if and only if $\{|x_i| \geq |y_i|; i = 1, \ldots, n\}$. Theorem 1 now assures that $(\mathcal{R}^n, \succeq)$ is partially ordered symmetrically up to equivalence, and that $(\mathcal{R}^n, \succeq)$ is a lattice up to equivalence under $\mathcal{G}$. For $\mathbf{x},\mathbf{y} \in \mathcal{R}^n$, $\mathbf{x} \wedge \mathbf{y}$ is the orbit identified with $\mathbf{w} \in \mathcal{R}^n$ having $\{w_i = \min(|x_i|, |y_i|); i = 1, \ldots, n\}$, and $\mathbf{x} \vee \mathbf{y}$ is the orbit identified with $\mathbf{z} \in \mathcal{R}^n$ having $\{z_i = \max(|x_i|, |y_i|); i = 1, \ldots, n\}$. By conclusion (v) of Theorem 1 the monotone functions on $(\mathcal{R}^n, \succeq)$ are generated by the class Ψ of functions on $\mathcal{R}^n_+$ increasing (i.e. nondecreasing) in each argument.

Other orderings of interest on $\mathcal{R}^n$ are symmetric under the permutation group $\mathcal{G}$. Then, with $\{x_{(1)} \geq x_{(2)} \geq \ldots \geq x_{(n)}\}$ as the order statistics, $M(\mathbf{x}) = (x_{(1)}, x_{(2)}, \ldots, x_{(n)})$ is a maximal invariant on $\mathcal{R}^n$ with range $\mathcal{D}_n$. A fundamental ordering on $\mathcal{R}^n$ is obtained via Theorem 1 through *majorization* on $\mathcal{D}_n$ such that $\mathbf{u} \succeq_M \mathbf{v}$ on $\mathcal{D}_n$ if and only if

$$\Sigma_1^k u_i \geq \Sigma_1^k v_i; k = 1, 2, \ldots . n-1 \tag{3.1}$$

$$\Sigma_1^n u_i = \Sigma_1^n v_i. \tag{3.2}$$

The result is a partial ordering on $\mathcal{R}^n$ up to equivalence. Ordering the elements of $\mathbf{x},\mathbf{y} \in \mathcal{R}^n$ before applying (3.1) and (3.2) is justified formally by (i) of Theorem 1. The functions monotone on $(\mathcal{R}^n, \succeq)$, i.e. the $\mathcal{S}$-convex functions of Schur (1923), may be generated from functions on $\mathcal{D}_n$ and conversely. Part (v) of Theorem 1 thus yields Proposition H.1

of Marshall and Olkin (1979, p. 92) as well as its converse.

Similar remarks apply to a weak majorization on $\mathcal{D}_n$ in which $k = 1, \ldots, n$ in (3.1) and (3.2) is deleted. Examples of functions on $\mathcal{R}^n$ montone under the induced ordering are the symmetric gauge functions (cf. Fan (1951)). Functions monotone under majorization and various weak majorizations are treated in Marshall and Olkin (1979).

3.2 Left-Unitary Ordering on $\mathcal{F}_{n\times m}$. Here $\mathcal{G}$ is the unitary group acting from the left, i.e. $\mathbf{A} \to \mathbf{UA}$ on $\mathcal{F}_{n\times m}$ with $\mathbf{U} \in \mathcal{U}(n)$. A maximal invariant is $M(\mathbf{A}) = \mathbf{A^*A}$ with range $\mathcal{H}_m^+$ (cf. Vinograde (1950)). The ordering considered here is as follows.

Definition 4. Two matrices $\mathbf{A},\mathbf{B} \in \mathcal{F}_{n\times m}$ are said to be ordered as $\mathbf{A} \succeq \mathbf{B}$ if and only if $\mathbf{A^*A} \succeq_L \mathbf{B^*B}$ on $\mathcal{H}_m^+$, i.e. $\mathbf{A^*A} - \mathbf{B^*B} \in \mathcal{H}_m^+$. The ordering $\mathbf{A} \succ \mathbf{B}$ is strict whenever $\mathbf{A^*A} - \mathbf{B^*B}$ is positive definite.

Basic properties of $(\mathcal{F}_{n\times m}, \succeq)$, ordered as in Definition 4, are given next, where $(\mathcal{H}_m^+, \succeq_L)$ is used in lieu of $(\mathcal{M}, \succeq_M)$.

THEOREM 2. *Let $\succeq$ be a relation on $\mathcal{F}_{n\times m}$ induced from $(\mathcal{H}_m^+, \succeq_L)$ as in Definition 4.* (i) *The relation $\succeq$ is invariant in the sense that $\mathbf{A} \succeq \mathbf{B}$ on $\mathcal{F}_{n\times m}$ if and only if $\mathbf{PAC} \succeq \mathbf{QBC}$ for any $\mathbf{P},\mathbf{Q} \in \mathcal{U}(n)$ and $\mathbf{C} \in Gl(m)$.* (ii) *$(\mathcal{F}_{n\times m}, \succeq)$ is partially ordered up to equivalence under $\mathcal{G}$.* (iii) *If $m = 1$ the ordering $\succeq$ is complete up to equivalence.* (iv) *If $\mathbf{A} \succeq \mathbf{B}$ on $\mathcal{F}_{n\times m}$, then $\mathbf{AG} \succeq \mathbf{BG}$ on $\mathcal{F}_{n\times s}$ for any $\mathbf{G} \in \mathcal{F}_{m\times s}$.* (v) *For $\mathbf{A},\mathbf{B} \in \mathcal{F}_{n\times m}$ a necessary and sufficient condition that $\mathbf{A} + \mathbf{B} \succeq \mathbf{A}$ is that $\mathbf{B^*B} + \mathbf{A^*B} + \mathbf{B^*A} \in \mathcal{H}_m^+$. In particular, if $\mathbf{A^*B} = \mathbf{0}$, then $\mathbf{A} + \mathbf{B} \succeq \mathbf{A}$.*

Proof. Conclusions (i)–(iii) follow from their antecedents in Theorem 1 together with properties of $(\mathcal{H}_m^+, \succeq_L)$; conclusion (iv) is immediate; and (v) follows on expanding $(\mathbf{A} + \mathbf{B})^*(\mathbf{A} + \mathbf{B})$. □

Claims for lattice properties of $(\mathcal{F}_{n\times m}, \succeq)$ are not available through Theorem 1. Halmos (1958, p. 142) showed that $(\mathcal{H}_m, \succeq_L)$ is not a lattice.

Well known properties of $(\mathcal{H}_n, \succeq_L)$ are 1) if $\mathbf{A} \succeq_L \mathbf{B} \succ_L \mathbf{0}$, then $\mathbf{B}^{-1} \succeq_L \mathbf{A}^{-1} \succ_L \mathbf{0}$ (cf. Loewner (1934)) and 2) if $\mathbf{A} \succeq_L \mathbf{B}$, then their ordered characteristic values, $Ch(\mathbf{A}) = \{\alpha_1 \geq \ldots \geq \alpha_n\}$ and $Ch(\mathbf{B}) = \{\beta_1 \geq \ldots \geq \beta_n\}$, satisfy $\{\alpha_i \geq \beta_i;\ 1 \leq i \leq n\}$ (cf. Bellman (1960), p. 115). The first was extended to singular matrices in $\mathcal{H}_n^+$ by Milliken and Akdeniz (1977) using the pseudo-inverse of Moore (1920) and Penrose (1955). Thus these inverse operators are order-reversing on the boundary and interior of $\mathcal{H}_n^+$ under $\succeq_L$. Corresponding properties are shown next for $(\mathcal{F}_{n\times m}, \succeq)$ in terms of 1′) the Moore-Penrose inverse operator $\mathbf{A} \to \mathbf{A}^\dagger$ on $\mathcal{F}_{n\times m}$ and 2′) the singular-value mapping $\mathbf{A} \to \sigma(\mathbf{A})$.

THEOREM 3. *Let $\sigma(\mathbf{A}) = \{\alpha_1 \geq \ldots \geq \alpha_m\}$ and $\sigma(\mathbf{B}) = \{\beta_1 \geq \ldots \geq \beta_m\}$ be the ordered singular values of $\mathbf{A},\mathbf{B} \in \mathcal{F}_{n\times m}$, and let $\mathbf{A}^\dagger$ and $\mathbf{B}^\dagger$ be their respective Moore-Penrose inverses.* (i) *If $\mathbf{A} \succeq \mathbf{B}$, then $\alpha_i \geq \beta_i$ for $i = 1, \ldots, m$.* (ii) *If $\mathbf{A} \succeq \mathbf{B}$ and if $\mathbf{A}$ and $\mathbf{B}$ have rank $s \leq m$, then $(\mathbf{B}^\dagger)^* \succeq (\mathbf{A}^\dagger)^*$ on $\mathcal{F}_{n\times m}$.*

Proof. Conclusion (i) is a restatement of property 2) of $(\mathrm{H}_m^+, \succeq_L)$. Conclusion (ii) follows because $\mathbf{A} \succeq \mathbf{B}$ implies $\mathbf{A^*A} \succeq_L \mathbf{B^*B}$, which implies (cf. Milliken and Akdeniz (1977)) that $(\mathbf{B^*B})^\dagger \succeq_L (\mathbf{A^*A})^\dagger$, which in turn implies $\mathbf{B}^\dagger(\mathbf{B}^\dagger)^* \succeq_L \mathbf{A}^\dagger(\mathbf{A}^\dagger)^*$ and thus (ii). The last step uses the singular decomposition $\mathbf{F} = \mathbf{PDQ^*}$ with $\mathbf{P} \in \mathcal{U}(n)$, $\mathbf{Q} \in \mathcal{U}(m)$, and $\mathbf{D}(n\times m)$, and the fact that $\mathbf{F}^\dagger = \mathbf{QD}^\dagger\mathbf{P}^*$. □

The functions monotone on $(\mathcal{F}_{n\times m}, \succeq)$ may be generated by composition as $\{f(\mathbf{X}) = \psi(\mathbf{X}^*\mathbf{X});\ \psi \in \Psi\}$, with Ψ the real functions monotone on $(\mathcal{H}^+_m, \succeq_L)$ as characterized by Marshall, Walkup, and Wets (1967). Other results follow.

THEOREM 4. *Let Φ be the class of functions on $(\mathcal{F}_{n\times m}, \succeq)$ as in Definition 3. If $\phi \in \Phi$, then ϕ is monotone. In particular, any $\phi \in \Phi_0$ is a monotone norm.*

Proof. That ϕ is order-preserving follows on combining conclusion (i) of Theorem 3 with property (ii) of Definition 2, for $\mathbf{A} \succeq \mathbf{B}$ implies $\phi(\mathbf{A}) = \gamma(\sigma_1(\mathbf{A}), \dots, \sigma_m(\mathbf{A})) \geqslant \gamma(\sigma_1(\mathbf{B}), \dots, \sigma_m(\mathbf{B})) = \phi(\mathbf{B})$. In particular, any $\phi \in \Phi_0$, a norm, is a monotone norm. □

In conclusion, note that the ordering $\succeq$ of Definition 4 is an alternative to Loewner's ordering $\succeq_L$ on H_n; clearly the two orderings coincide on $\mathcal{H}^+_n$. Examples of elements in $\mathcal{H}_2$ ordered by $\succeq$ but not $\succeq_L$ are the diagonal matrices $\mathbf{A} = \mathrm{Diag}(2,-2)$ and $\mathbf{B} = \mathrm{Diag}(1, -1)$.

3.3 Unitary Orderings on $\mathcal{F}_{n\times m}$. Let $\mathcal{G}$ be the unitary group on $\mathcal{F}_{n\times m}$ taking $\mathbf{A}$ into $\mathbf{UAV}$ with $\mathbf{U} \in \mathcal{U}(n)$ and $\mathbf{V} \in \mathcal{U}(m)$. A maximal invariant is $\sigma(\mathbf{A})$ with range $\mathcal{D}^+_m$, each orbit in $\mathcal{F}_{n\times m}$ having matrices with the same ordered singular values.

Orderings on $\mathcal{D}^+_m$ of interest here are those of Section 3.1: (i) ordering by coordinates, (ii) majorization, and (iii) weak majorization. Properties of $(\mathcal{F}_{n\times m}, \succeq)$ under the induced orderings follow from Theorem 1 as before. Functions monotone on $(\mathcal{F}_{n\times m}, \succeq)$ are equivalent to compositions $\{f(\mathbf{A}) = \psi(\sigma(\mathbf{A}));\ \psi \in \Psi\}$ where, for the three cases, (i) Ψ consists of functions on $\mathcal{D}^+_m$ increasing in each argument, (ii) Ψ consists of $\mathcal{S}$-convex functions, and (iii) Ψ is the class Γ_0 of Definition 2. The latter combines Theorem 1(v) with a result of Fan (1951). For this case the class of monotone functions is Φ_0, the unitarily invariant norms being monotone on $(\mathcal{F}_{n\times m}, \succeq)$ when ordering is induced through weak majorization of the singular values. For case (i) $(\mathcal{F}_{n\times m}, \succeq)$ is a lattice up to equivalence under $\mathcal{G}$.

PART II. APPLICATIONS

1. Introduction. The foregoing concepts apply in a variety of settings, where different orderings serve different purposes and a careful choice may yield results not otherwise attainable. Unless stated otherwise, we take $(\mathcal{F}_{n\times m}, \succeq)$ to be ordered as in Definition 4.

Subsequently define $\mathcal{F}_{n\times m}$ over $\mathcal{R}^1$ and let $\mathcal{O}(n)$ be the group of real orthogonal $(n\times n)$ matrices. Let $S_{n,m}(\Theta,\Gamma\times\Xi)$ be the class of ellipsoidal matrix distributions on $\mathcal{F}_{n\times m}$ with typical density

$$f(\mathbf{Y}) = g(tr(\mathbf{Y}-\Theta)\Xi^{-1}(\mathbf{Y}-\Theta)'\Gamma^{-1}); \tag{1.1}$$

let $U_{n,m}(\Theta,\Gamma\times\Xi)$ be the subclass of distributions unimodal in the sense of Anderson (1955); let $G_{n,m}(\Theta,\Gamma\times\Xi)$ be the subclass of these consisting of scale mixtures of matrix Gaussian laws; and let $L_{n,m}(\Theta,\Lambda)$ be the distribution on $\mathcal{F}_{n\times m}$ with typical density

$$f(\mathbf{Y}) = h(\Lambda'(\mathbf{Y}-\Theta)'(\mathbf{Y}-\Theta)\Lambda). \tag{1.2}$$

Here $\Theta \in \mathcal{F}_{n\times m}$ consists of location parameters, while $\Xi \in \mathcal{H}^+_m$, $\Gamma \in \mathcal{H}^+_n$, and $\Lambda \in \mathcal{F}_{m\times m}$ are scale parameters, all nonsingular. These distributions are considered in Jensen and Good (1981). When $m = 1$ these specialize to the classes $S_n(\theta,\Gamma)$, $U_n(\theta,\Gamma)$, and $G_n(\theta,\Gamma)$ on $\mathcal{R}^n$. The distribution of $\mathbf{W}$ is denoted by $\mathcal{L}(\mathbf{W})$.

A useful ordering for probability measures is the following (compare Sherman (1955)), where $\mathcal{X}$ is a linear space having the zero element $0 \in \mathcal{X}$.

Definition 5. Let $(\mathcal{X}, \mathcal{B}, \cdot)$ be a measurable space. The probability measure μ is *more peaked* about $0 \in \mathcal{X}$ than ν if $\mu(A) \geq \nu(A)$ for every set in the class τ of compact convex measurable sets A symmetric under reflection, i.e. $x \in A$ implies $-x \in A$. Denote this ordering by $\mu \succeq_p \nu$.

2. Peakedness of Measures. For suitable measures μ and ν on $\mathcal{R}^n$ symmetric about $\mathbf{0}$ and having the dispersion matrices Σ_μ and Σ_ν such that $\Sigma_\nu \succeq_L \Sigma_\mu$, μ is more peaked about $\mathbf{0}$ than ν. This was shown for Gaussian measures on $\mathcal{R}^n$ by Anderson (1955) and for $S_n(\mathbf{0},\Gamma)$ by Fefferman, Jodeit, and Perlman (1972). We extend these results to linear transformations from $\mathcal{R}^n$ to $\mathcal{R}^m$ with $m \leq n$, and we also supply a converse.

THEOREM 5. *Let μ_A and μ_B be probability measures on $\mathcal{R}^m$ induced by $\mathbf{y} \to \mathbf{A}'\mathbf{y}$ and $\mathbf{y} \to \mathbf{B}'\mathbf{y}$ from $\mathcal{L}(\mathbf{y}) \in S_n(\mathbf{0},\mathbf{I}_n)$ with $\mathbf{A},\mathbf{B} \in \mathcal{F}_{n\times m}$. Then $\mu_A \succeq_p \mu_B$ on $\mathcal{R}^m$ if and only if $\mathbf{B} \succeq \mathbf{A}$ on $(\mathcal{F}_{n\times m}, \succeq)$.*

Proof. (i) Clearly $\mathbf{A}'\mathbf{A} = \Sigma_A$ and $\mathbf{B}'\mathbf{B} = \Sigma_B$ are the scale parameters of μ_A and μ_B. If $\mathbf{B} \succeq \mathbf{A}$, then $\Sigma_B \succeq_L \Sigma_A$ and the ordering $\mu_A \succeq_p \mu_B$ follows from that of Fefferman, Jodeit, and Perlman (1972). (ii) Conversely, suppose that $\mu_A \succeq_p \mu_B$ but neither $\mathbf{A} \succeq \mathbf{B}$ nor $\mathbf{B} \succeq \mathbf{A}$. As $\succeq_p$ is preserved under nonsingular linear transformations, simultaneously reduce Σ_A to $\mathbf{I}_m$ and Σ_B to $\mathbf{D} = \mathrm{Diag}(\delta_1, \ldots, \delta_m)$ where, for some $k \in (1,m)$, $\delta_1 \leq \ldots \leq \delta_k < 1 \leq \delta_{k+1} \leq \ldots \leq \delta_m\}$. Then the marginal measures ν_A and ν_B on $\mathcal{R}^k$ are ordered as $\nu_B \succeq_p \nu_A$, and necessity follows by contradiction using the fact that probability measures on $\mathcal{R}^n$ are tight. □

3. Linear Estimation. The problem is to estimate Θ in the matrix model $\mathbf{Y} = \mathbf{X}\Theta + \mathbf{E}$ with $\mathbf{Y} \in \mathcal{F}_{n\times m}$ observable, $\mathbf{X} \in \mathcal{F}_{n\times r}$ known of rank $r \leq n$, and $\mathbf{E} \in \mathcal{F}_{n\times m}$ a matrix of random errors. Minimizing $Q(\Theta) = tr(\mathbf{Y}-\mathbf{X}\Theta)'(\mathbf{Y}-\mathbf{X}\Theta)$ as Θ varies yields the least-squares solution $\hat{\Theta} = (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{Y}$ for Θ and $\mathbf{X}\hat{\Theta}$ for approximating $\mathbf{Y}$. We show much stronger minimizing properties.

THEOREM 6. *Suppose $\mathbf{Y} = \mathbf{X}\Theta + \mathbf{E}$ and order $(\mathcal{F}_{n\times m}, \succeq)$ as in Definition 4.* (i) $\hat{\Theta} = (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{Y}$ *is minimizing on $(\mathcal{F}_{n\times m}, \succeq)$ in the sense that $(\mathbf{Y}-\mathbf{X}\Theta) \succeq (\mathbf{Y}-\mathbf{X}\hat{\Theta})$ for every $\Theta \in \mathcal{F}_{r\times m}$.* (ii) $\psi(\mathbf{Y}-\mathbf{X}\hat{\Theta}) \leq \psi(\mathbf{Y}-\mathbf{X}\Theta)$ *for every ψ in the class Ψ of functions monotone on $(\mathcal{F}_{n\times m}, \succeq)$.* (iii) $\hat{\Theta}$ *is the minimum-norm solution to* $\min_{\Theta \in \mathcal{F}_{r\times m}} \|\mathbf{Y}-\mathbf{X}\Theta\|_\Phi$ *for every unitarily invariant norm $\|\cdot\|_\Phi$ on $\mathcal{F}_{n\times m}$.*

Proof. Because $\mathbf{X}'\mathbf{X}\hat{\Theta} - \mathbf{X}'\mathbf{Y} = \mathbf{0}$, the expansion

$$(\mathbf{Y}-\mathbf{X}\Theta)'(\mathbf{Y}-\mathbf{X}\Theta) = (\mathbf{Y}-\mathbf{X}\hat{\Theta})'(\mathbf{Y}-\mathbf{X}\hat{\Theta}) + (\hat{\Theta}-\Theta)'\mathbf{X}'\mathbf{X}(\hat{\Theta}-\Theta)$$

yields conclusion (i) directly. Conclusion (ii) follows by monotonicity, and conclusion (iii) from Theorem 4, where the unitarily invariant norms are shown to be monotone. □

Conclusion (iii) was obtained by Rao (1980) as a consequence of ordering the singular values $\sigma(\mathbf{Y}-\mathbf{X}\Theta)$ and $\sigma(\mathbf{Y}-\mathbf{X}\hat{\Theta})$. Our conclusion (i) is stronger, as it implies Rao's ordering using Theorem 3.

4. Ordered Designs. Specialize the foregoing model with $m = 1$ to $\mathbf{y} = \mathbf{X}\beta + \mathbf{e}$ and consider the choice of design as it pertains to testing $\mathbf{H}$: $\beta = \mathbf{0}$ against A: $\beta \neq \mathbf{0}$. If $\mathcal{L}(\mathbf{e}) \in U_n(\mathbf{0},\sigma^2\mathbf{I}_n)$, then the power function of the normal-theory test for H against A depends on the parameters (β,σ^2) and the design $\mathbf{X}$ only through $\lambda = \beta'\mathbf{X}'\mathbf{X}\beta/\sigma^2$, and it increases monotonically with λ; cf. Theorem 3.2 of Jensen (1979). Suppose one of the designs $\mathbf{A},\mathbf{B} \in \mathcal{F}_{n\times r}$ is to be chosen. A connection between the ordering $(\mathcal{F}_{n\times r}, \succeq)$ of Definition 4 and the power of this test is given in the following theorem for any unimodal spherical law of errors.

THEOREM 7. *Suppose* $\mathcal{L}(\mathbf{y}) \in U_n(\mathbf{X}\beta, \sigma^2\mathbf{I}_n)$ *with* $\mathbf{X} \in \mathcal{F}_{n\times r}$ *a design matrix of rank* $r \le n$. *Of two designs* $\mathbf{A},\mathbf{B} \in \mathcal{F}_{n\times r}$, *the F-test at level* α *for testing* $\mathbf{H}$: $\beta = \mathbf{0}$ *against* A:$\beta \neq \mathbf{0}$ *is uniformly more powerful using Design A than using Design B if and only if* $\mathbf{A} \succeq \mathbf{B}$ *on* $\mathcal{F}_{n\times r}$.

Proof. Sufficiency follows from the preceding paragraph because the power function depends only on λ. A proof for necessity parallels that of Theorem 5 on identifying subspaces where each is more powerful when $\mathbf{A} \not\succeq \mathbf{B}$. □

5. The $\mathbf{T^2}$ Chart under Scale Ordering. Hotelling's (1947) T^2 chart uses samples from a vector-valued process to monitor the stationarity of means over time. Its *run length* is the number of successive samples taken before signaling that the process is not in control. Suppose successive observations are independent Gaussian vectors on $\mathcal{R}^m$ having parameters (μ,Σ), with Σ characteristic of the process. In practice efforts are made to tighten the process to reduce its variability. The effects of such reduction on run lengths are as follows.

THEOREM 8. *Let* N_1 *be the run length of Hotelling's (1947)* T^2 *chart for monitoring a stationary Gaussian process with parameters* (μ, Σ_1), *and let* N_2 *be the run length of* T^2 *for a tightened process with parameters* (μ, Σ_2). *Then* N_2 *is stochastically smaller than* N_1 *for all* $\mu \in \mathcal{R}^m$ *if and only if* $\Sigma_1 \succeq_L \Sigma_2$ *on* $(\mathcal{H}_m^+, \succeq_L)$.

Proof. The proof parallels that of Theorem 7. □

6. Canonical Analysis. Let $\mathbf{z} = (\mathbf{z}_1', \mathbf{z}_2')'$ be Gaussian on $\mathcal{R}^n$ having zero means and the dispersion matrix

$$\Omega = \begin{pmatrix} \mathbf{I}_r & \mathbf{R} \\ \mathbf{R}' & \mathbf{I}_s \end{pmatrix}, \mathbf{R} \in \mathcal{F}_{r\times s}, \; r \le s, \tag{6.1}$$

and consider the quadratic forms $U_1 = \mathbf{z}_1'\mathbf{z}_1$ and $U_2 = \mathbf{z}_2'\mathbf{z}_2$. Their joint distribution depends only on $\sigma(\mathbf{R})$, the canonical correlations of Hotelling (1936). If an ordering on $(\mathcal{F}_{r\times s}, \succeq)$ is induced by coordinate-wise ordering of the singular values on $\mathcal{D}_r^+$, then a monotonicity property of certain probabilities is given in the following.

THEOREM 9. *Let* $\mu_R(\cdot)$ *be the joint measure of* $U_1 = \mathbf{z}_1'\mathbf{z}_1$ *and* $U_2 = \mathbf{z}_2'\mathbf{z}_2$ *on* $\mathcal{R}_+^2$ *having the cross correlation matrix* $\mathbf{R}$ *in (6.1) with* $r = s$. *Then for every measurable set* $A \subset \mathcal{R}_+^1$, *the measure* $\mu_R(A\times A)$ *is order-preserving in the sense that* $\Xi \succeq \Gamma$ *on* $(\mathcal{F}_{r\times r}, \succeq)$ *implies that* $\mu_\Xi(A\times A) \geq \mu_\Gamma(A\times A)$.

Proof. An expansion in the Lancaster (1958) canonical form was given by Jensen (1970) for the joint distribution, the coefficients $G_k(\rho)$ depending on $\rho = \sigma(\mathbf{R})$. The proof consists of integrating the expansion over $A\times A$ and showing that the resulting expression is an increasing function of ρ. □

We next show that a bivariate Chebyshev bound is monotone when considered as a function of $\mathbf{R}$ on $(\mathcal{F}_{r\times s}, \succeq)$ for any $r \le s$. This result is essentially distribution-free. Define

$$B(\delta_1, \delta_2; \mathbf{R}) = ((s-r)/\delta_2) + \Sigma_{i=1}^r\{(\delta_1+\delta_2) + [(\delta_1+\delta_2)^2 - 4\rho_i^2\delta_1\delta_2]^{1/2}\}/2\delta_1\delta_2 \tag{6.2}$$

in terms of the canonical correlations $\{\rho_1, \ldots, \rho_r\}$ of $\mathbf{y}_1$ and $\mathbf{y}_2$ for fixed δ_1 and δ_2.

THEOREM 10. *Let* $\mathbf{y} = (\mathbf{y}_1', \mathbf{y}_2')'$ *be a random vector of order* $r + s = n$, *with* $r \leq s$, *having the mean* $\mu = (\mu_1', \mu_2')'$, *the dispersion matrix* $\Sigma = [\Sigma_{ij}]$, *and* $\mathbf{R} = \Sigma_{11}^{-1/2}\Sigma_{12}\Sigma_{22}^{-1/2}$. *Then* (i) $P((\mathbf{y}_1-\mu_1)'\Sigma_{11}^{-1}(\mathbf{y}_1-\mu_1) \leq \delta_1, (\mathbf{y}_2-\mu_2)'\Sigma_{22}^{-1}(\mathbf{y}_2-\mu_2) \leq \delta_2) \geq 1-B(\delta_1,\delta_2;\mathbf{R})$; (ii) $B(\delta_1,\delta_2;\mathbf{R})$ *is monotone decreasing on* $(\mathcal{F}_{r\times s}, \succeq)$.

Proof. Conclusion (i) is given in Jensen (1982). Conclusion (ii) follows from expression (6.2), from the mapping $\sigma(\mathbf{R}) = (\rho_1, \dots, \rho_r)$, and from the fact that $\Sigma_{11}^{-1/2}\Sigma_{12}\Sigma_{22}^{-1/2} \succeq \Omega_{11}^{-1/2}\Omega_{12}\Omega_{22}^{-1/2}$ on $(\mathcal{F}_{r\times s}, \succeq)$, for example, if and only if their singular values are pairwise ordered on $\mathcal{D}_r^+$. □

Observe that Theorems 9 and 10 remain valid when $(\mathcal{F}_{r\times s}, \succeq)$ is ordered as in Definition 4. This follows from conclusion (i) of Theorem 3.

7. Signal Detection under Symmetry. Each channel of a k-channel receiver accepts an input vector $\mathbf{y} \in \mathcal{R}^n$ that is either processed as a signal or suppressed as noise depending on whether the input amplitude $||\mathbf{y}|| = (\mathbf{y}'\mathbf{y})^{1/2}$ does or does not exceed a threshold value c. Thus for k channels $\mathbf{Y} = [\mathbf{y}_1, \dots, \mathbf{y}_k]$ represents the input, typically correlated between channels, and a probabilistic assessment of the system performance focuses on expressions of the type

$$F(c_1, \dots, c_k; \Theta) = P_\Theta(||\mathbf{y}_1|| \leq c_1, \dots, ||\mathbf{y}_k|| \leq c_k), \tag{7.1}$$

when signals of varying strengths actually enter the system. A useful ordering is the following.

THEOREM 11. *Suppose* $\mathcal{L}(\mathbf{Y}) \in L_{n,k}(\Theta,\Lambda)$ *and its pdf is unimodal. Then the probability*

$$F(c_1, \dots, c_k;\Theta) = P_\Theta(||\mathbf{y}_1|| \leq c_1, \dots, ||\mathbf{y}_k|| \leq c_k),$$

when considered as a function of Θ *with* $\{c_1, \dots, c_k\}$ *fixed, is monotone decreasing on* $(\mathcal{F}_{n\times k}, \succeq)$ *under the ordering of Definition 4.*

Proof. Let $A \subset \mathcal{F}_{n\times k}$ be the set

$$A = \{\mathbf{Y} \in \mathcal{F}_{n\times k} | \mathbf{y}_1'\mathbf{y}_1 \leq c_1^2, \dots, \mathbf{y}_k'\mathbf{y}_k \leq c_k^2\};$$

Let $\mathcal{G}$ be the group $\mathcal{O}(n)$ acting on $\mathbf{Y}$ from the left; and observe that (i) A is a convex $\mathcal{G}$-invariant subset of $\mathcal{F}_{n\times k}$ and (ii) the pdf of $(\mathbf{Y}-\Theta)$ is a nonnegative real-valued, $\mathcal{G}$-invariant and unimodal function on $\mathcal{F}_{n\times k}$. Conditions (i) and (ii) satisfy the requirements of Theorem 5 of Mudholkar (1966) which assures that

$$P_\Theta(A) \geq P_{\Theta_0}(A) \tag{7.2}$$

for any Θ in the convex hull of the $\mathcal{G}$-orbit of Θ_0. This orbit is characterized by constant values of the maximal invariant function $\Theta_0'\Theta_0$. On taking sections, we infer that Θ is in the convex hull of the $\mathcal{G}$-orbit of Θ_0 if for all $\mathbf{a} \in \mathcal{R}^k$, $a'\Theta'\Theta a \leq a'\Theta_0'\Theta_0 a$, i.e. if $\Theta_0'\Theta_0 \succeq_L \Theta'\Theta$. But from Definition 4, this fact together with (7.2) are equivalent to the assertion of the theorem. □

In practice this assures that the larger the shift in the sense of the ordering $\succeq$, the greater the probability that signals are correctly identified and processed as signals in one or more channels. In particular, in back-up systems designed with redundancies, the detection probability will increase with the magnitude of the signal.

Acknowledgment. Thanks are due a referee for suggestions toward revision.

REFERENCES

ANDERSON, T. W. (1955). The integral of a symmetric unimodal function over a symmetric convex set and some probability inequalities. *Proc. Amer. Math. Soc. 6* 170–176.

BELLMAN, R. (1960). *Introduction to Matrix Analysis.* McGraw-Hill, New York.

EATON, M. L. (1984). On group induced orderings, monotone functions, and convolution theorems. In *Inequalities in Statistics and Probability*, Y. L. Tong, ed., Institute of Mathematical Statistics, Hayward, Ca., 13–25.

FAN, K. (1951) Maximum properties and inequalities for the eigenvalues of completely continuous operators. *Proc. Nat. Acad. Sci. U. S. A. 37* 760–766.

FEFFERMAN, C., JODEIT, M., and PERLMAN, M. D. (1972). A spherical surface measure inequality for convex sets. *Proc. Amer. Math. Soc. 33* 114–119.

HALMOS, P. (1958) *Finite-Dimensional Vector Spaces,* Second. ed., Van Nostrand, New York.

HARTWIG, R. E. (1976). Block generalized inverses. *Arch. Rat. Mech. Anal. 61* 197–251.

HOTELLING, H. (1947). Multivariate quality control. In *Techniques of Statistical Analysis*, C. Eisenhart, M. W. Hastay, and W. A. Wallis, eds., McGraw-Hill, New York, 111–184.

JENSEN, D. R. (1970). The joint distribution of quadratic forms and related distributions. *Australian J. Statist. 12* 13–22.

JENSEN, D. R. (1979). Linear models without moments. *Biometrika 66* 611–617.

JENSEN, D. R. (1982). Inequalities for joint distributions of quadratic forms. *SIAM J. Appl. Math. 42* 297–301.

JENSEN, D. R. and GOOD, I. J. (1981). Invariant distributions associated with matrix laws under structural symmetry. *J. Roy. Statist. Soc. B 43* 327–332.

LANCASTER, H. O. (1958). The structure of bivariate distributions. *Ann. Math. Statist. 29* 719–736.

LEHMANN, E. L. (1959). *Testing Statistical Hypotheses.* Wiley, New York.

LOEWNER, C. (1934). Über monotone Matrixfunktionen. *Math. Z. 38* 177–216.

MARSHALL, A. W. and OLKIN, I. (1979). *Inequalities: Theory of Majorization and Its Applications.* Academic Press, New York.

MARSHALL, A. W., WALKUP, D. W., and WETS, R. J. B. (1967). Order-preserving functions; applications to majorization and order statistics. *Pacific J. Math. 23* 569–584.

MILLIKEN, G. A. and AKDENIZ, F. (1977). A theorem on the difference of generalized inverses of two nonnegative matrices. *Comm. Statist. A6* 73–79.

MOORE, E. H. (1920). On the reciprocal of the general algebraic matrix (abstract). *Bull. Amer. Math. Soc. 26* 394–395.

MUDHOLKAR, G. S. (1966). The integral of an invariant unimodal function over an invariant convex set—An inequality and applications. *Proc. Amer. Math. Soc. 17* 1327–1333.

NEUMANN, J. von (1937). Some matrix inequalities and metrization of matric space. *Tomsk Univ. Rev. 1* 286–300.

PENROSE, R. (1955). A generalized inverse for matrices. *Proc. Cambridge Philos. Soc. 51* 406–413.

RAO, C. R. (1980). Matrix approximations and reduction of dimensionality in multivariate statistical analysis. In *Multivariate Analysis— V*, P. R. Krishnaiah, ed., North Holland, 3–22.

SCHATTEN, R. (1970). *Norm Ideals of Completely Continuous Operators.* Springer-Verlag, Berlin and New York.

SCHUR, I. (1923). Über eine Klasse von Mittelbildungen mit Anwendungen die Determinantentheorie Sitzungsber. *Berlin Math. Gesellschaft 22* 9–20.

SHERMAN, S. (1955). A theorem on convex sets with applications. *Ann. Math. Statist. 26* 763–766.

VINOGRADE, B. (1950). Canonical positive definite matrices under internal linear transformations. *Proc. Amer. Math. Soc. 1* 159–161.

VULIKH, B. Z. (1967). *Introduction to the Theory of Partially Ordered Spaces.* Wolters-Noordhoff, Groningen.

WONG, Y.-C. and NG, K.-F. (1973). *Partially Ordered Topological Vector Spaces.* Clarendon, Oxford.

Inequalities in Statistics and Probability
IMS Lecture Notes–Monograph Series Vol. 5 (1984), 35–40

RANDOM REPLACEMENT SCHEMES AND MULTIVARIATE MAJORIZATION[1]

BY SAMUEL KARLIN and YOSEF RINOTT
Stanford University and the Hebrew University of Jerusalem

In this note we obtain certain inequalities comparing random replacement schemes to sampling with replacement. Some of the results are related to multivariate majorization and Schur functions.

1. Various Stochastic Comparisons and Random Replacement Schemes. Let $\mathscr{A} = \{a_1, \ldots, a_N\}$, $a_i \in \mathscr{R}$ = the real line. We shall consider a sample of size n ($n \leq N$) drawn from A, and denote the observations by $X_1, \ldots, X_n$. In a symmetric random replacement scheme the observation X_1 is drawn with equal probabilities from A, i.e., $P(X_1 = a_i) = 1/N$, $i = 1, \ldots, N$. The element drawn for X_1 is replaced in A with probability π_1, and removed from A with probability $1-\pi_1$. Then X_2 is sampled, and the element which is drawn is replaced with probability π_2. Continuing to X_{n-1}, the vector $\pi = (\pi_1, \ldots, \pi_{n-1})$ defines the random replacement scheme $R(\pi)$. Note that for $\pi = \mathbf{0} = (0, \ldots, 0)$, $R(\pi)$ is equivalent to sampling without replacement while for $\pi = \mathbf{1} = (1, \ldots, 1)$, $R(\pi)$ corresponds to sampling with replacement and $X_1, \ldots, X_n$ are i.i.d.

It follows from Joag-Dev and Proschan (1983) that under $R(\mathbf{0})$, $X_1, \ldots, X_n$ are *negatively associated, i.e.,*

$$E\{\phi(X_i, i \in A)\psi(X_j, j \in B)\} \leq E\phi(X_i, i \in A)E\psi(X_j, j \in B) \tag{1.1}$$

for any partition A,B of $1, \ldots, n$, where ϕ and ψ are increasing functions.

In particular, (1.1) implies

$$E\{\Pi_{i=1}^n \varphi_i(X_i)\} \leq \Pi_{i=1}^n E\varphi_i(X_i) \tag{1.2a}$$

for any functions φ_i, all increasing (or all decreasing) and nonnegative. Note that (1.2a) can be written as

$$E_{R(\mathbf{0})}\{\Pi_{i=1}^n \varphi_i(X_i)\} \leq E_{R(\mathbf{1})}\{\Pi_{i=1}^n \varphi_i(X_i)\}. \tag{1.2b}$$

Inequalities for sampling schemes were obtained by various authors including Sen (1970), Rosén (1972), Serfling (1973), Kemperman (1973), Karlin (1974), Van Zwet (1983), and Krafft and Schaefer (preprint). The question of characterizing the class of functions for which

$$E_{R(\pi)}\psi(X_1, \ldots, X_n) \leq E_{R(\mathbf{1})}\psi(X_1, \ldots . X_n) \tag{1.3}$$

remains unresolved. The next result provides a class of functions for which (1.3) holds.

THEOREM 1. $\mathbf{E}_{R(\pi)}\{\Pi_{i=1}^n \varphi(X_i)\} \leq \mathbf{E}_{\mathbf{R}(\mathbf{1})}\{\Pi_{i=1}^n \varphi(X_i)\}$ *for any* $\varphi \geq 0$.

Proof. We write π instead of $R(\pi)$ as an index for the expectation. For $n = 2$

[1] *Supported in part by NIH Grant 5R01 GM10452-20 and NSF Grant MSC79-24310.*
AMS 1980 subject classifications. 62D05, 62H05.
Key words and phrases: Schur functions, negative association, birthday problem.

$$E_{\pi_1}\{\varphi(X_1)\varphi(X_2)\} = \pi_1/N^2 (\Sigma_{k=1}^N \varphi(a_k))^2 + \{(1-\pi_1)/(N(N-1))\} \Sigma_{k=1}^N \varphi(a_k)(\Sigma_{j\neq k}\varphi(a_j))$$

$$(1.4) \qquad = \pi_1/N^2 (\Sigma_{k=1}^N \varphi(a_k))^2 + \{(1-\pi_1)/(N(N-1))\} \Sigma_{k=1}^N \varphi(a_k)(\Sigma_{j=1}^N \varphi(a_j) - \varphi(a_k))$$

$$= \pi_1/N^2 (\Sigma_{k=1}^N \varphi(a_k))^2 + \{(1-\pi_1)/(N(N-1))\}\{(\Sigma_{k=1}^N \varphi(a_k))^2 - \Sigma_{k=1}^N \phi^2(a_k)\}.$$

Now $\Sigma\varphi^2(a_k) \geqslant (\Sigma\varphi(a_k))^2/N$, and therefore with $\Sigma\varphi(a_k) = m$ the last expression in (1.4) is bounded above by

$$\pi_1 m^2/N^2 + \{(1-\pi_1)/(N(N-1))\} \{m^2(1 - 1/N)\} = m^2/N^2 = N^{-2}(\Sigma_{k=1}^N \varphi(a_k))^2 = E_1(\varphi(X_1)\varphi(X_2)\}$$

and the case $n = 2$ is established. We now proceed by induction.

Let $\psi(X_1, \ldots, X_n) = \Pi_{i=1}^n \phi(X_i)$. Then (see Karlin (1974), Lemma 3.1)

$$E_{\{\pi_1, \ldots, \pi_{n-1}\}}\psi(X_1, \ldots, X_n) = (1/N) \Sigma_{k=1}^N \pi_1 E_{\{\pi_2, \ldots, \pi_{n-1}\}}\psi(a_k, X_2, \ldots, X_n)$$
$$+ (1/N)\Sigma_{k=1}^N (1-\pi_1) E'_{\{\pi_2, \ldots, \pi_{n-1}\}}\psi(a_k, X_2, \ldots, X_n)$$

where E' computes the expectation when a_k is removed from the sample space of $X_2, \ldots, X_n$. Invoking the induction hypothesis, i.e., Theorem 1 holding for $n-1$ variables this leads to

$$E_{\{\pi_1, \ldots, \pi_{n-1}\}}\psi(X_1, \ldots, X_n) \leqslant \pi_1 E_{\{1, \ldots, 1\}}\psi(X_1, \ldots, X_n).$$
$$+ (1-\pi_1)E_{\{0,1, \ldots, 1\}}\psi(X_1, \ldots, X_n).$$

Hence in order to complete the induction argument it suffices to prove

$$(1.5) \qquad E_{\{0,1, \ldots, 1\}}\psi(X_1, \ldots, X_n) \leqslant E_{\{1, \ldots, 1\}}\psi(X_1, \ldots, X_n).$$

Since $\varphi(a_i)$ is only a relabeling of a_i we assume $\varphi(a_i) = a_i \geqslant 0$ and also without loss of generality $a_1 \leqslant a_2 \leqslant \ldots \leqslant a_N$. With this (1.5) becomes

$$(1.6) \qquad \{N(N-1)^{n-1}\}^{-1}\Sigma_{k=1}^N a_k(\Sigma_{j=1}^N a_j - a_k)^{n-1} \leqslant N^{-n}(\Sigma_{j=1}^N a_j)^n$$

and with $b_k = a_k/(\Sigma_{j=1}^N a_j)$ so that $\Sigma_{k=1}^N b_k = 1$ we obtain that (1.6) is equivalent to

$$(1.7) \qquad \Sigma_{k=1}^N b_k(1-b_k)^{n-1} \leq \{1-(1/N)\}^{n-1}.$$

We now prove (1.7) by induction on N. For $N = 1$, (1.7) is trivial. If the maximum of $\Sigma_{k=1}^N b_k(1-b_k)^{n-1}$ is obtained at a boundary point of the simplex $\{b_i \geqslant 0, \Sigma_{i=1}^N b_i = 1\}$, then some $b_i = 0$ and by the induction hypothesis at the maximum point

$$\Sigma_{k=1}^N b_k(1-b_k)^{n-1} \leqslant \{1-(1/(N-1))\}^{n-1} \leqslant \{1-(1/N)\}^{n-1}.$$

If the maximum is at an interior point, then by differentiating $\Sigma b_k(1-b_k)^{n-1} - \lambda(\Sigma b_k - 1)$ we obtain the equation $(1-b_k)^{n-1} - (n-1)b_k(1-b_k)^{n-2} - \lambda = 0$, and equivalently

$$(1.8) \qquad n(1-b_k)^{n-1} - (n-1)(1-b_k)^{n-2} - \lambda = 0.$$

Summing in (1.8) over k we have

$$(1.9) \qquad n\Sigma_{k=1}^N (1-b_k)^{n-1} - (n-1)\Sigma_{k=1}^N (1-b_k)^{n-2} = N\lambda.$$

Now, using the Tchebycheff rearrangement inequality,

$$(1.10) \qquad \Sigma_{k=1}^N (1-b_k)^{n-1} \geqslant (1/N)\Sigma_{k=1}^N (1-b_k) S_{k=1}^N (1-b_k)^{n-2} = ((N-1)/N)\Sigma_{k=1}^N (1-b_k)^{n-2}$$

and therefore from (1.9)

$$(N/n)\lambda \geqslant ((N-1)/N)\Sigma_{k=1}^N (1-b_k)^{n-2} - ((n-1)/n)\Sigma_{k=1}^N (1-b_k)^{n-2} \geqslant 0.$$

Returning to the expressing in (1.8), $\lambda \geq 0$ implies that the polynomial $nx^{n-1}-(n-1)x^{n-2}-\lambda$ has only one positive root by Descartes' rule of signs. Therefore, an interior maximum of

$\Sigma_{k=1}^{N} b_k(1 - b_k)^{n-1}$ can occur only when all values of $1-b_k$ are equal to this root and hence if an interior maximum exists it must occur at $b_1 = \dots = b_n = 1/N$. At this point (1.7) holds with equality and thus (1.7) is now established. □

Remark. The inequality

$$E_{\{0,1,\dots,1\}}\{\Pi_{i=1}^{n}\varphi_i(X_i)\} \leq E_{\{1,\dots,1\}}\{\Pi_{i=1}^{n}\varphi_i(X_i)\}$$

for *different* φ_i increasing does *not* hold in general. To see this note that for $\varphi_1 \equiv 1/N$ and $\varphi_i(a_i) = a_i$, $i = 2, \dots, n$ we would have to prove instead of (1.7)

$$\Sigma_{k=1}^{N} N^{-1}(1 - b_k)^{n-1} \leq (1 - (1/N))^{n-1}. \tag{1.11}$$

However (1.11) holds with equality when all $b_i = 1/N$ and the inequality is reversed for any other choice of b_i.

As a special case of Theorem 1 we obtain

$$P_{R(\pi)}[X_1 \leq c, \dots, X_n \leq c] \leq \{P(X_1 \leq c)\}^n$$

where on the right-hand side X_1 takes the values $\{a_1, \dots, a_n\}$ with equal probabilities, i.e., $P(X_1 = a_i) = 1/N$. Also,

$$P_{R(\pi)}[X_1 \geq c, \dots, X_n \geq c] \leq \{P(X_1 \geq c)\}^n.$$

The next result should be compared with Theorem 3.1 of Karlin (1974).

THEOREM 2. *Let*

$$\psi(x_1, \dots, n_n) = \Sigma_{1 \leq i_j < \dots < i_r \leq n}\Pi_{j=1}^{r}\varphi(x_{i_j}), \quad r \leq n$$

where $\varphi \geq 0$. *Then*

$$E_{R(\pi)}\psi(X_1, \dots, X_n) \leq E_{R(1)}\psi(X_1, \dots, X_n). \tag{1.12}$$

The case $r = n$ coincides with Theorem 1.

Proof. The case $r = 1$ is trivial so we take $r \geq 2$. Note that it suffices to assume $0 \leq a_1 \leq \dots \leq a_N$ and take $\psi(x_1, \dots, x_n) = \Sigma_{1 \leq i_j < \dots < i_r \leq n}\Pi_{j=1}^{r} x_{i_j}$.

Again, it suffices to prove (1.5) which for the present ψ becomes

$$\binom{n-1}{r}(N(N-1)^r)^{-1}\Sigma_{k=1}^{N}(1 - b_k)^r + \binom{n-1}{r-1}(N(N-1)^{r-1})^{-1}\Sigma_{k=1}^{N} b_k(1 - b_k)^{r-1} \leq N^{-r}\binom{n}{r} \tag{1.13}$$

where $\Sigma_{k=1}^{N} b_k = 1, b_i \geq 0$, which reduces to

$$((n-r)/(N-1))\,\Sigma_{k=1}^{N}(1 - b_k)^r + r\,\Sigma_{k=1}^{N} b_k(1 - b_k)^{r-1} \leq (1 - (1/N))^{r-1}n. \tag{1.14}$$

We prove (1.14) by induction on N. As before, on the boundary where some $b_i = 0$, (1.14) follows readily from the induction hypothesis. Differentiation with respect to b_k of the left hand side of (1.14) with the constraint $\Sigma_{k=1}^{N} b_k = 1$ yields

$$-((n-r)/(N-1))r(1-b_k)^{r-1} + r(1-b_k)^{r-1} - r(r-1)b_k(1-b_k)^{r-2} - \lambda = 0 \tag{1.15}$$

or

$$r(r-(n-r)/(N-1))(1-b_k)^{r-1} - r(r-1)(1-b_k)^{r-2} - \lambda = 0 \tag{1.16}$$

Summation over k produces

$$r(r-(n-r)/(N-1))\Sigma_{k=1}^{N}(1-b_k)^{r-1} - r(r-1)\Sigma_{k=1}^{N}(1-b_k)^{r-2} = \lambda N$$

and invoking the inequality

$$\Sigma_{k=1}^{N}(1 - b_k)^{r-1} \geq ((N-1)/N)\Sigma_{k=1}^{N}(1 - b_k)^{r-2}; \quad \text{see (1.10)}$$

we have

$$(N/r)\lambda \geq (r-(n-r)/(N-1))((N-1)/N) - (r-1) \ \Sigma_{k=1}^{N}(1-b_k)^{r-2}$$
$$= ((N-n)/N)\,\Sigma_{k=1}^{N}(1-b_k)^{r-2} \geq 0.$$

Thus, again (1.16) has a unique critical point with $b_k \equiv 1/N$. Since for $b_k \equiv 1/N$ (1.13) hold with equality, the result follows. □

Examples. Theorem 2 implies inequality (1.12) for $\psi(x_1, \ldots, x_n) = (x_1 + \ldots + x_n)^\alpha$ for any integer $\alpha > 0$. For $\Sigma_{k=1}^{N} a_i < 1$ we obtain by expansion that (1.12) also holds with $\psi(x_1, \ldots, x_n) = [1-(x_1 + \ldots + x_n)^\alpha]^{-1}$ or any positive combination $\Sigma c_k(x_1 + \ldots + x_n)^{\alpha_k}$, $\alpha_k \geq 0$ integers.

The preceding inequalities are related to multivariate majorization and Schur function as explained next.

2. Multivariate Majorization and Negative Association. A function $\varphi(\mathbf{x})$ defined on $\mathcal{R}^N$ *is said to be Schur concave* if $\varphi(\mathbf{x}) \geq \varphi(\mathbf{y})$ for any $\mathbf{x},\mathbf{y} \in \mathcal{R}^N$ such that $\mathbf{x} = \mathbf{yM}$ for some matrix $\mathbf{M} \in \mathcal{D}$ = the class of $N\times N$ doubly stochastic matrices. See Marshall and Olkin (1979) for details, references and historical remarks. Let $\mathbf{X}$ and $\mathbf{Y}$ be $n\times N$ matrices whose columns are $\mathbf{x}_1, \ldots, \mathbf{x}_N$, and $\mathbf{y}_1, \ldots, \mathbf{y}_N$, respectively. The inequality $\Sigma_{i=1}^{N} g(\mathbf{x}_i) \geq \Sigma_{i=1}^{N} g(\mathbf{y}_i)$ holds for every concave function g defined on $\mathcal{R}^n$ if and only if there exists a matrix $\mathbf{M} \in \mathcal{D}$ such that $\mathbf{X} = \mathbf{YM}$. (This result is due to Hardy, Littlewood and Pólya (1934) for $n = 1$, and to Sherman (1951), Stein and Blackwell (1953)).

In particular, the matrix function $\psi(\mathbf{X}) = \Sigma_{i=1}^{N} g(\mathbf{x}_i)$ satisfies $\psi(\mathbf{X}) \geq \psi(\mathbf{Y})$ whenever $\mathbf{X} = \mathbf{YM}$ provided g is concave. Related notions of multivariate Schur concavity and probabilistic applications were studied by Rinott (1973), Marshall and Olkin (1979), Karlin and Rinott (1981), Tong (1982) and Karlin and Rinott (1983). In some of the applications one obtains the inequality $\psi(\mathbf{X}) \geqslant \psi(\mathbf{Y})$ whenever $\mathbf{X} = \mathbf{YM}$ where $\mathbf{M}$ belongs to a subclass of $\mathcal{D}$. Of particular interest is the class $\mathcal{T}$ of matrices which can be represented as products of matrices of the form $t\mathbf{I} + (1-t)\mathbf{P}$ where $\mathbf{I}$ is the $N\times N$ identity matrix, $\mathbf{P}$ is a permutation matrix which interchanges only two coordinates, and $0 \leq t \leq 1$.

Our next theorem describes an example of Schur concavity with respect to the class $\mathcal{T}$. A probabilistic interpretation of the result in terms of a birthday-problem of coincidence probabilities will be given. We first need a lemma which extends Ostrowski's (1952) well-known criterion for Schur concavity. The proof can be found in Rinott (1973), Marshall and Olkin (1979).

LEMMA 1. *A differentiable function* $\psi\colon \mathcal{R}^{nN} \to R$ *is multivariate Schur concave with respect to* $\mathcal{T}$, *i.e.,* $\psi(\mathbf{X}) \leq \psi(\mathbf{XT})$ *for every* $\mathbf{T} \in \mathcal{T}$ *and* $n\times N$ *matrix* $\mathbf{X} = \|x_{ij}\|$ *if and only if*

(i) $\psi(\mathbf{X}) = \psi(\mathbf{XP})$ *for every* $N\times N$ *permutation matrix* $\mathbf{P}$; *and*

(ii) $\Sigma_{i=1}^{n}(x_{ij}-x_{ik})[\partial\psi(\mathbf{X})/\partial x_{ij} - \partial\psi(\mathbf{X})/\partial x_{ik}] \leq 0$ for all $1 \leq j \neq k \leq N$.

Let $\alpha^1, \ldots, \alpha^n \in \mathcal{R}^N$ denote the n *rows* of the $n\times N$ matrix $\mathbf{A}$, $\alpha^i = (\alpha_1^i, \ldots, \alpha_n^i)$, $i = 1, \ldots, n$. We assume that the rows are similarly ordered, that is $(\alpha_j^i-\alpha_k^i)(\alpha_j^{i'}-\alpha_k^{i'}) \geq 0$ for all $1 \leq j,\ k \leq N,\ 1 \leq i,\ i' \leq n$. Note that if $\mathbf{T} = t\mathbf{I} + (1-t)\mathbf{P}$ where $\mathbf{P}$ is a permutation matrix that interchanges only two coordinates, then applied to these two coordinates, $\mathbf{T}$ operates like the matrix $\left(\begin{smallmatrix} t & 1-t \\ 1-t & t \end{smallmatrix}\right)$ which preserves the order if $t \geq ½$ and reverses the order if $t \leq ½$. If the rows of $\mathbf{A}$ are similarly ordered, then so are the rows of $\mathbf{AT}$ for any $\mathbf{T} \in \mathcal{T}$.

THEOREM 3. *Let* $\psi(\mathbf{A})$ *be defined by*

$$\psi(\mathbf{A}) = \Sigma_{j_1 \neq \ldots \neq j_n} \Pi_{k=1}^{n} \alpha_{j_k}^{k} \tag{2.1}$$

where the sum extends over all $\binom{N}{n}$ *vectors of n different indices between* 1 *and* N $\mathbf{A} = \|\alpha_j^i\|$ *is* $n \times N$ *satisfying* $\alpha_j^i \geq 0$, $i = 1, \ldots, n$, $j = 1, \ldots, N$, $n \leq N$, *and the rows of* $\mathbf{A}$, $\alpha^1, \ldots, \alpha^n$ *are similarly ordered. Then* $\psi(\mathbf{A}) \leq \psi(\mathbf{AT})$ *for all* $\mathbf{T} \in \mathcal{T}$

Proof. In view of Lemma 1, we compute

$$\partial\psi/\partial\alpha_1^1 - \partial\psi/\partial\alpha_2^1 = \Sigma_{1 \neq j_2 \neq \ldots \neq j_n} \alpha_{j_2}^2 \ldots \alpha_{j_n}^n - \Sigma_{2 \neq j_2 \neq \ldots \neq j_n} \alpha_{j_2}^2 \ldots \alpha_{j_n}^n .$$

Let

$$u_k = \Sigma_{3 \leq j_2 \neq \ldots \neq j_{k-1} \neq j_{k+1} \neq \ldots \neq j_n} \alpha_{j_2}^2 \ldots \alpha_{j_{k-1}}^{k+1} \alpha_{j_{k+1}}^{k+1} \ldots \alpha_{j_n}^n, \quad k = 1, \ldots, n.$$

Then

$$\begin{aligned} \partial\psi/\partial\alpha_1^1 - \partial\psi/\partial\alpha_2^1 &= (\Sigma_{k=2}^n \alpha_2^k u_k + u_1) - (\Sigma_{k=2}^n \alpha_1^k u_k + u_1) \\ &= \Sigma_{k=2}^n (\alpha_2^k - \alpha_1^k) u_k. \end{aligned} \tag{2.2}$$

Therefore

$$\Sigma_{k=1}^n (\alpha_1^i - \alpha_2^i)(\partial\psi/\partial\alpha_1^1 - \partial\psi/\partial\alpha_2^1) = \Sigma_{i=1}^n \Sigma_{k \neq i} (\alpha_1^i - \alpha_2^i)(\alpha_2^k - \alpha_1^k) u_k \leq 0$$

since similar ordering implies $(\alpha_1^i - \alpha_2^i)(\alpha_2^k - \alpha_1^k) \leq 0$, and replacing 1,2 by any pair of indices the required result follows from Lemma 1. □

Note that Theorem 3 involves Schur concavity with respect to $\mathcal{T}$ on the set of nonnegative $n \times N$ matrices having similarly ordered rows.

In the proof of Theorem 3 consider the subclass $\mathcal{S}$ of $\mathcal{T}$ consisting of finite products of matrices of the form $\mathbf{T} = t\mathbf{T} + (1-t)\mathbf{P}$, $\mathbf{P}$ a permutation matrix that interchanges only two adjacent coordinates and $\frac{1}{2} \leq t \leq 1$. Such a $\mathbf{T}$ preserves the ordering of the components when applied to a vector. The calculation in (2.2) implies $(\alpha_1^1 - \alpha_2^1)(\partial\psi/\partial\alpha_1^1 - \partial\psi/\partial\alpha_2^1) = (\alpha_1^1 - \alpha_2^1)\Sigma(\alpha_2^k - \alpha_1^k)u_k \leq 0$ and the same holds if we replace the pair of indices 1,2 by any pair. By the well known criterion of Ostrowski (1952) it follows that $\psi(\mathbf{A}) = \psi(\alpha^1, \ldots, \alpha^n)$ is Schur convex in α^1, when $\alpha^2, \ldots, \alpha^n$ are fixed and $\alpha^1, \ldots, \alpha^n$ are all similarly ordered. This implies

THEOREM 4. *Under the conditions of Theorem 3* $\psi(\alpha^1, \ldots, \alpha^n) \leq \psi(\alpha^1\mathbf{T}_1, \ldots, \alpha^n\mathbf{T}_n)$ *for all* $\mathbf{T}_1, \ldots, \mathbf{T}_n \in \mathcal{S}$.

As a special case of Theorem 3 we obtain the inequalities (1.2a)–(1.2b). This is given by

PROPOSITION 1. *Let* $0 \leq \varphi_i$ *be increasing functions,* $i = 1, \ldots, n$, *then*

$$E_{R(0)}\{\Pi_{i=1}^n \varphi_i(X_i)\} \leq E_{R(1)}\{\Pi_{i=1}^n \varphi_i(X_i)\}. \tag{2.3}$$

Proof. Set $\alpha_j^i = \varphi_i(a_j)$, $i = 1, \ldots, n$, $j = 1, \ldots, N$. Then $E_{R(0)}\{\Pi_{i=1}^n \varphi_i(X_i)\} = \psi(\mathbf{A})/(N(N-1) \ldots (N-n+1))$ where $\psi(\mathbf{A})$ is defined by (2.1), while $E_{R(1)}\{\Pi_{i=1}^n \varphi_i(X_i)\} = N^{-n}\Pi_{i=1}^n(\Sigma_{j=1}^N \alpha_j^i)$. It is easy to see that inequality of (2.3) is homogeneous and we can assume $\Sigma_{j=1}^N \alpha_j^i = 1$, $i = 1, \ldots, n$, without loss of generality. Then (2.3) becomes

$$\psi(\mathbf{A}) = N(N-1) \ldots (N-n+1)/N^n. \tag{2.4}$$

For $\mathbf{J} \in \mathcal{T}$ having all entries equal to N^{-1} we now have $\mathbf{AJ} = \mathbf{J}$, and a simple calculation shows that $\psi(\mathbf{J}) = N(N-1) \ldots (N-n+1)/N^n$. Schur concavity of ψ implies $\psi(\mathbf{A}) \leq \psi(\mathbf{AJ}) - \psi(\mathbf{J})$ and (2.4) follows. □

3. A Generalized Birthday Problem. We finally apply Theorem 3 to obtain an extension of the "*birthday problem*" (see Marshall and Olkin, 1979, p. 305). Consider a group of n individuals. Let α_j^i, $i = 1, \ldots, n$, $j = 1, \ldots, N = 365$ denote the probability that the ith person's birthday occurs on the jth day of the year, $1 \le j \le 365 = N$. Then for $\psi(\mathbf{A})$ defined in (2.1) we have for n independent persons

$$\psi(\mathbf{A}) = \text{Probability that the } n \text{ persons have } n \text{ distinct birthdays,}$$

i.e., no coincidences of birthdays occur.

This probability was studied in the case that the likelihood of a birthday on a particular day is the same for all persons. Here we allow different persons to have different distributions of birthdays as long as the vectors $(\alpha_1^i, \ldots, \alpha_{365}^i)$ are similarly ordered, which means that if day j has a higher probability of being one person's birthday than day k, then the same holds for all individuals. We have $\psi(\mathbf{A}) \le \psi(\mathbf{AT})$ for $\mathbf{T} \in \mathcal{T}$ and in particular $\psi(\mathbf{A}) \le \psi(\mathbf{AJ})$, which says that under the above assumptions the probability of no coincidence of birthdays is maximized if all days are equally likely birthdays for all individuals.

Added in proof: Theorem 2 can be derived from the theorem in Van Zwet (1983).

REFERENCES

BLACKWELL, D. (1953). Equivalent comparisons of experiments. *Ann. Math. Statist. 24* 265–272.

HARDY, G.H., LITTLEWOOD, J.E., and PÓLYA, G. (1934, 1952). *Inequalities*. 1st ed., 2nd. ed. Cambridge University Press, London and New York.

JOAG-DEV, K. and PROSCHAN, F. (1983). Negative association of random variables with applications. *Ann. Statist. 11* 286–295.

KARLIN, S. (1974). Inequalities for symmetric sampling plans I. *Ann. Statist. 2* 1065–1094.

KARLIN, S. and RINOTT, Y. (1981). Entropy inequalities for classes of probability distributions. II. The multivariate case. *Adv. Appl. Prob. 13* 325–351.

KARLIN, S. and RINOTT, Y. (1983). Comparison of measures, multivariate majorization, and applications to statistics. *Studies in Econometrics, Time Series, and Multivariate Statistics* S. Karlin, T. Amemiya, L. A. Goodman, eds., Academic Press, New York.

KEMPERMAN, J. H. B. (1973). Moment problems for sampling without replacement. I, II, III. *Indag. Math. 35* 149–188.

MARSHALL, A. W. and OLKIN, I. (1979). *Inequalities Theory of Majorization and Its Applications*. Academic Press, New York.

OSTROWSKI, A. M. (1952). Sur quelques applications des fonctions convexes et concaves au sens de I. Schur. *J. Math. Pures. Appl. 31* 253–292.

RINOTT, Y. (1973). Multivariate majorization and rearrangement inequalities with some applications to probability and statistics. *Israel J. Math. 15* 60–77.

ROSÉN, B. (1972). Asymptotic theory for successive sampling with varying probabilities without replacement I. *Ann. Math. Statist. 43* 373–397.

SEN, P. K. (1970). The Hájek-Rényi inequality for sampling from a finite population. *Sankhyā 32* 181–188.

SERFLING, R. J. (1973). Refinements of some probability inequalities of Hoeffding (abstract). *Bull. Inst. Math. Statist. 2* 104.

SHERMAN, S. (1951). On a theorem of Hardy, Littlewood, Pólya and Blackwell. *Proc. Natl. Acad. Sci. USA 37* 826–831.

TONG, Y. L. (1982). Rectangular and elliptical probability inequalities for Schur-concave random variables. *Ann. Statist. 10* 637–642.

VAN ZWET, W. R. (1983). An inequality for random replacement sampling plans. *A Festschrift for Erich L. Lehmann*. P. J. Bickel, K. Doksum, J. L. Hodges, Jr., eds. Wadsworth, CA.

Inequalities in Statistics and Probability
IMS Lecture Notes–Monograph Series Vol. 5 (1984), 41–53

EIGENVALUE INEQUALITIES FOR RANDOM EVOLUTIONS: ORIGINS AND OPEN PROBLEMS

BY JOEL E. COHEN
The Rockefeller University

This paper has three parts. The first give an expository review of a model from stochastic population biology. This model leads to eigenvalue inequalities for random evolutions. The second gives a proof by Charles M. Newman of one of these inequalities. The third gives conjectures and questions. Some of these have been previously stated; most are new.

1. Introduction. The growth of a population in a random environment can be modeled in a simple way by a stochastic process called a random evolution. Bounds on the growth rate of a population in a random environment can be expressed in terms of certain eigenvalues. The first purpose of this paper is to describe informally how a population can be modelled by a random evolution and why eigenvalue inequalities arise naturally (section 2). The main inequalities to be discussed have been proved by Cohen, Friedland, Kato and Kelly (1982). I shall henceforth refer to this paper as CFKK. The second purpose of this paper is to give a proof (section 3) of one of these inequalities that was discovered by Charles M. Newman of the University of Arizona during my talk at the symposium on Inequalities in Statistics and Probability. The third purpose of this paper is to state conjectures, open problems and questions concerning further inequalities (section 4).

2. Populations in Random Environments and Eigenvalue Inequalities. Suppose one has a vat of bacteria sitting in a laboratory. Suppose the number of bacteria is large enough so that there is no discomfort in taking $N(t)$, the number of bacteria at (real scalar) time t, to be a real variable rather than strictly integer valued. Suppose also that the number of bacteria is small compared to the number of bacteria that the nutrient medium in the vat can support, or that the medium is continuously refreshed. If the division cycles of the bacteria are unsynchronized, then the simplest model of the population is to suppose that the number of fissions that occur per unit time is directly proportional to the number of bacteria in the vat. Thus, for some real constant b,

$$dN(t)/dt = bN(t) \text{ for } t \geqslant 0, N(0) = N_0.$$

It is well known, even among biologists, that the solution of this equation is

$$N(t) = N_0 e^{bt}.$$

In this deterministic model, the long-run growth rate b may be computed from an observed trajectory $N(t)$ of the size of the population from the formula

$$\lim_{t\to\infty} t^{-1} \log N(t) = b.$$

The left side of this equation is referred to as a Liapunov characteristic number of the process.

AMS 1980 subject classifications: Primary 60H25, 15A42; Secondary 60J57, 92A15.

Key words and phrases. Random environment. Liapunov characteristic number. Multiplicative Markov functional. Feynman-Kac formula. Stochastic population biology. Discrete skeleton. Reversible Markov chain. Additive Markov functional.

If one interprets the growth rate b as a stochastic process that is degenerate at a single fixed value for all time t, one can take averages wherever one wishes in the Liapunov characteristic number, i.e.,

$$t^{-1} E \log N(t) = t^{-1} \log E N(t),$$

because both sides are equal with probability 1 to b. However, if b is a stochastic process that is not degenerate, the above equality must be replaced by

$$t^{-1} E \log N(t) \leqslant t^{-1} \log E N(t)$$

because of Jensen's inequality and the concavity of log.

Now suppose that the conditions of our vat of bacteria are not perfectly uniform in time, but fluctuate randomly because of changes in the outside weather, in the voltages of the power lines that drive the vat's heating bath and mixers, and in other factors affecting the growth rate of the bacteria. Let us model these fluctuations by supposing that

$$dN(t) / dt = b(t)N(t) \quad \text{for} \quad t \geqslant 0, N(0) = N_0,$$

where $b(t)$ is a real-valued functional of an n-state $(1 < n < \infty)$ continuous-time homogeneous irreducible Markov process $V(t)$. This means that the sample paths of $b(t)$ are piecewise constant, and each piece is constant at one of n real numbers $b_1, \ldots, b_n$. The Markov chain $V(t)$ on the state space $\{1, \ldots, n\}$ may be thought of as determining the subscript i of the growth rate b_i that is current at time t according to

$$b(t) = b_{V(t)}, \quad \text{with} \quad b(0) = b_1.$$

Given that the Markov chain $V(t)$ starts in state 1, the subsequent behavior is determined by the intensity matrix $\mathbf{Q}$ according to

$$P[V(t+s) = j \mid V(t) = i] = (e^{\mathbf{Q}s})_{ij} \quad \text{for} \quad s \geqslant 0, t \geqslant 0.$$

Recall that the intensity matrix $\mathbf{Q}$ is essentially nonnegative, i.e., $q_{ij} \geqslant 0$ if $i \neq j$, with the additional condition that $\Sigma_{j=1}^{n} q_{ij} = 0, i = 1, 2 \ldots, n$.

The process $N(t)$ is called a random evolution. It is a special case of the random evolutions studied by Griego and Hersh (1971), reviewed by Hersh (1974) and inspired by Kac (1957).

Since $V(t)$ and $b(t)$ are piecewise constant, a graph of a sample path of $\log N(t)$ as a function of t is piecewise linear, increasing when $b(t) > 0$, constant when $b(t) = 0$, and declining when $b(t) < 0$.

If $g_j(t)$ is the occupancy time in state j up to time t, i.e., the sum of the lengths of the time intervals up to time t such that $V(t) = j$, then an explicit formula for the random evolution $N(t)$ is

$$N(t) = N_0 \exp(\Sigma_{j=1}^{n} b_j g_j(t)).$$

This formula is obvious because, as long as $V(t) = j$, $N(t)$ grows exponentially at rate b_j as if $b(t)$ were fixed at b_j.

The formula makes it easy to compute one of the plausible measures of the long-run growth rate. First, since $\mathbf{Q}$ is irreducible (meaning that every growth rate is accessible from every other growth rate), there is an invariant or equilibrium probability vector π with positive elements $\pi_i, i = 1, 2, \ldots, n$, such that

$$\pi^T \mathbf{Q} = 0$$

and such that

$$g_j(t) / t \to \pi_j \quad \text{with probability 1, for} \quad j = 1, \ldots, n.$$

Then using the above formula for $N(t)$ gives

$$\lim_{t\to\infty} t^{-1} E \log N(t) = \lim_{t\to\infty} t^{-1} E(\Sigma_{j=1}^{n} b_j g_j(t)) = \Sigma_{j=1}^{n} b_j \pi_j.$$

In words, the mean of the growth rates of population size (averaged over sample paths) equals the mean growth rate (averaged over the growth rates of any single sample path).

The second plausible measure of the long-run growth rate, namely the growth rate of mean population size, is given by a Feynman-Kac formula for this random evolution (Cohen, 1979a) as

$$\lim_{t\to\infty} t^{-1} \log E\, N(t) = \log r(e^{\mathbf{Q}+\mathbf{B}}), \tag{2.1}$$

where r is the spectral radius, or maximum of the moduli of the eigenvalues, and $\mathbf{B}$ is the $n \times n$ diagonal matrix with jth diagonal element equal to b_j.

Whereas the mean of the growth rates of population size depends on the intensity matrix $\mathbf{Q}$ only through its leading left eigenvector π^T, and is therefore the same for any intensity matrix with the same π^T, the growth rate of mean population size depends on all of $\mathbf{Q}$.

While it is plausible to suppose that the bacteria grow in continuous time, it is equally plausible to suppose that a biologist observes them at discrete time intervals. Suppose he or she observes once a day the conditions (temperature, light, nutrient concentration) affecting bacterial growth and infers or records a time series $b(0)$, $b(1)$, $b(2)$, ... of instantaneous growth rates. Suppose the observer models $N^D(t)$ the number of bacteria (D for "discrete") by

$$N^D(t+1) = N_D(t)e^{b(t)},\ t = 0, 1, 2, \ldots, N^D(0) = N_0.$$
$$P[b(t+1) = b_j \mid b(t) = b_i] = (e^{\mathbf{Q}})_{ij} \equiv p_{ij},\ b(0) = b_1.$$

According to the first of these equations, if the observer sees growth rate $b(t)$ at the epoch of observation on day t, he supposes that this growth rate will continue without variation until the epoch of observation on the next day. Since he has no information to tell him otherwise, this seems a reasonable first approximation. According to the second of these equations, he takes the transition probability p_{ij} from growth rate b_i to growth rate b_j to be just the transition probability that would be estimated from any long sample path of $V(t)$ by sampling at unit intervals.

The biologist's purpose in constructing this discrete approximation $N^D(t)$ to the random evolution $N(t)$ is to estimate the growth rates of $N(t)$. If he computes $\lim t^{-1} E \log N^D(t)$ and $\lim t^{-1} \log E\, N^D(t)$, how will these rates relate to the corresponding rates for $N(t)$?

For the average growth rate of population size, it follows from the explicit formula

$$N^D(t) = N_0 \exp(\Sigma_{j=1}^{n} b_j g_j^D(t))$$

and the fact that

$$g_j^D(t)/t \to \pi_j \textit{ with probability } 1,\ j = 1, \ldots, n,$$

where $g_j^D(t)$ is the discrete occupancy time of state j prior to time t (= number of days fr m 0 up to and including t–1 such that $V(.) = j$) that

$$\lim_{t\to\infty} t^{-1} E \log N^D(t) = \Sigma_{j=1}^{n} b_j \pi_j$$
$$= \lim_{t\to\infty} t^{-1} E \log N(t).$$

Thus the discrete model for $N^D(t)$ give exactly the average growth rate of the continuous time random evolution $N(t)$.

What about the second measure, the growth rate of average population size? An elementary computation shows that

$$EN^D(t) = e^{b(0)} \times \text{first row sum of } (e^{\mathbf{Q}}e^{\mathbf{B}})^t.$$

Because $\mathbf{Q}$ is irreducible, $e^{\mathbf{Q}}$ is (elementwise) positive, so that by the Perron-Frobenius theorem

$$\lim_{t\to\infty} t^{-1} \log EN^D(t) = \log r(e^{\mathbf{Q}}e^{\mathbf{B}}). \tag{2.2}$$

After publishing this formula (Cohen, 1979b, p. 249), I discovered that it was known to LeBras (1974). I have recently learned from a secondary source (Iosifescu, 1980, pp. 162–163) that the formula should be credited to papers by O. Onicescu and G. Mihoc published during World War II.

The growth rate (2.2) of mean population size in the discrete model is an upper bound on the growth rate (2.1) of mean population size in the random evolution. More precisely, Theorem 2 of CFKK (p. 64) states: If $\mathbf{A}$ and $\mathbf{B}$ are two real $n \times n$ matrices, all $a_{ij} \geq 0$ if $i \neq j$, and all $b_{ij} = 0$ if $i \neq j$, then

$$r(e^{\mathbf{A}+\mathbf{B}}) \leqslant r(e^{\mathbf{A}}e^{\mathbf{B}}). \tag{2.3}$$

Moreover, this inequality is strict if $\mathbf{A}$ is irreducible and at least two diagonal elements of $\mathbf{B}$ are distinct. The next section gives a proof of the weak inequality.

Before giving that proof, let me informally show how (2.2) can be used to derive (2.1). (This does not pretend to be a rigorous proof.) We construct a sequence of discrete approximations $N_1^D(t) = N^D(t), N_2^D(t), N_3^D(t), \ldots$. In the kth approximation a unit interval of time is divided into k equal subintervals. The growth rate $b(t)$ is constrained to be constant on each subinterval but is permitted to change, with transition probability matrix $e^{\mathbf{Q}/k}$, from one subinterval to the next. Within each subinterval of length $1/k$, the long-run growth rate of mean population size, from (2.2), is $\log r(e^{\mathbf{Q}/k}e^{\mathbf{B}/k})$. Therefore, in one unit of time, which is k subintervals of length $1/k$, the growth rate of mean population size is

$$k \log r(e^{\mathbf{Q}/k}e^{\mathbf{B}/k}) = \log r^k(e^{\mathbf{Q}/k}e^{\mathbf{B}/k}) = \log r([e^{\mathbf{Q}/k}e^{\mathbf{B}/k}]^k).$$

But for any two $n \times n$ matrices $\mathbf{A}$ and $\mathbf{B}$, there is a formula attributed to Sophus Lie (can anyone tell me the original source?)

$$\lim_{k\to\infty} (e^{\mathbf{A}/k}e^{\mathbf{B}/k})^k = e^{\mathbf{A}+\mathbf{B}}.$$

If you accept that

$$\lim_{k\to\infty} \lim t^{-1} \log EN_k^D(t) = \lim t^{-1} \log EN(t),$$

so that the growth rate of mean population size in the discrete approximations approaches that of the continuous-time random evolution, then the two preceding formulas and an exchange of limits combine to yield the Feynman-Kac formula (2.1).

If the Markov chain $V(t)$ is reversible, CFKK proved (p. 62) that many more eigenvalue inequalities hold. For example, if the eigenvalues $\lambda_1, \ldots, \lambda_n$ of an arbitrary $n \times n$ complex matrix M are ordered so that $|\lambda_1| \geqslant |\lambda_2| \geqslant \ldots \geqslant |\lambda_n|$, and if

$$\varphi_k = \Sigma_{i=1}^k |\lambda_i|, \; k = 1, \ldots, n,$$

then for any reversible intensity matrix $\mathbf{Q}$, any diagonal real matrix $\mathbf{B}$ and $k = 1, \ldots, n$,

$$\varphi_k(e^{\mathbf{Q}+\mathbf{B}}) \leqslant \varphi_k(e^{\mathbf{Q}}e^{\mathbf{B}}).$$

The random evolution described here has also been found useful by economists to whom

I have suggested it as a model of price inflation in random (political, social and economic) environments. In this application, $N(t)$ is an index of price, $b(t)$ the instantaneous rate of inflation, and one of the interesting questions is: given $0 < N_1 < N_2$ and $t_1 < t_2$ and $N(t_1) < N_1$, what is the probability that $N(t) < N_2$ for all $t_1 \leq t \leq t_2$? Another question is: given $N(t_1) = N_1$, what is the distribution of the first time $t_2 > t_1$ such that $N(t_2) = N_2 > N_1$?

I hope it will both aid and promote people's interest in studying discrete and continuous random evolutions to observe that these random evolutions are special multiplicative functionals of Markov chains and processes. Markov's first paper in 1906 on these chains (again according to Iosifescu [1980]) aimed for a central limit theorem for additive functionals of a chain. Since then, many publications, scattered and apparently overlapping, have dealt with various functionals of Markov chains and processes. Among them are Doeblin (1937), Frechet (1952), Blanc-Lapierre and Fortet (1953 [1965]), Volkov (1958), Kemeny and Snell (1960 [1976]), Meyer (1962), Keilson and Wishart (1964, 1967) Fukushima and Hitsuda (1967), Pinsky (1968), Keilson and Rao (1970), Kertz (1974), O'Brien (1974), Ellis (1974), Serfozo (1975), Çinlar (1975), Keepler (1976), Wolfson (1977), Katz (1977), Durrett and Resnick (1978), Dagsvik (1978), primary papers cited in Iosifescu (1980), Friedland (1981) and Siegrist (1981). This work needs coherent synthesis.

3. Newman's Proof. After my lecture at the Nebraska Conference on Inequalities, Charles M. Newman presented me with a proof of (2.3). I give this proof with his kind permission.

LEMMA 1. *Let* $\mathbf{A}$ *and* $\mathbf{B}$ *be two real* $n \times n$ *matrices with* $a_{ij} \geq 0$ *if* $i \neq j$, *and* $b_{ij} = 0$ *if* $i \neq j$. *Then for any positive integer* m, *any real* x_i, $i = 1, \ldots, m$, *and* $a > 0$,

$$F(x_1, \ldots, x_m) = Tr(e^{a\mathbf{A}}e^{x_1\mathbf{B}}e^{a\mathbf{A}}e^{x_2\mathbf{B}} \ldots e^{a\mathbf{A}}e^{x_m\mathbf{B}})$$

is minimized on $\{x_i \geq 0, \Sigma x_i = ma\}$ *at* $x_i = a$ *for all* i.

Proof. We may express $\mathbf{A}$ as $\mathbf{A} = \mathbf{Q} + \mathbf{D}$ where $\mathbf{Q}$ is an intensity matrix and $\mathbf{D}$ is diagonal real. Let $V(t)$ be the Markov process determined by $\mathbf{Q}$ with an initial distribution that is uniform on the states $\{1, \ldots, n\}$ and conditioned so that $V(0) = V(T)$ where $T = am$. Thinking of $a = T/m$ as a discrete timestep for observing $V(t)$, we have by the Feynman-Kac formula that

$$KF(x_1, \ldots, x_m) = E\exp\{\textstyle\int_0^T d_{V(t)}dt + \Sigma_{j=1}^m x_j b_{V(ja)}\}$$

where $d_j = \mathbf{D}_{jj}$ and $b_j = \mathbf{B}_{jj}$ and $K = [Tr(e^{TQ})]^{-1}$.

To see that F is a convex function of $(x_1, \ldots, x_m)$ it suffices to show that the matrix with (i,j)th element equal to $\partial^2 F/\partial x_i \partial x_j$ is positive semi-definite.

If $c_1, \ldots, c_m$ are any m complex numbers it follows that

$$K\Sigma_{i,j}\bar{c}_i\{\partial^2 F/\partial x_i \partial x_j]c_j = E(|\Sigma_j c_j b_{V(ja)}|^2 \exp[\Sigma_h x_h b_{V(ha)} + \textstyle\int_0^T d_{V(t)}dt]) \geq 0.$$

Thus F is convex in $(x_1, \ldots, x_m)$.

On $\{x_1 \geq 0, \Sigma x_i = ma\}$, write $x_m = ma - \Sigma_{i=1}^{m-1} x_i$ so that $x_1, \ldots, x_{m-1}$ are independent. To prove Lemma 1, it suffices to show that

$$\partial F/\partial x_j = 0 \text{ at } (x_1, \ldots, x_m) = (a, a, \ldots, a).$$

But because x_m depends on all x_j, $1 \leq j \leq m-1$, we find

$$\begin{aligned}\partial F/\partial x_j = {} & Tr(e^{a\mathbf{A}}e^{x_1\mathbf{B}} \ldots e^{a\mathbf{A}}\mathbf{B}e^{x_j\mathbf{B}} \ldots e^{a\mathbf{A}}e^{x_m\mathbf{B}}) \\ & - Tr(e^{a\mathbf{A}}e^{x_1\mathbf{B}} \ldots e^{a\mathbf{A}}e^{x_j\mathbf{B}} \ldots e^{a\mathbf{A}}e^{x_m\mathbf{B}}\mathbf{B}).\end{aligned}$$

When $(x_1, \ldots, x_{m-1}) = (a, \ldots, a)$ the two Trace terms are equal and $\partial F/\partial x_j = 0$. □

LEMMA 2. *For all* $N = 1, 2, \ldots,$

$$Tr([e^{\mathbf{A}}e^{\mathbf{B}}]^N) \geq Tr(e^{(\mathbf{A}+\mathbf{B})N}).$$

Proof. Choosing $a = 1/K$, $m = KN$ as before, let $x_j = 0$ for all j that are not exactly divisible by K and let $x_j = 1$ if K divides j exactly. In this case,

$$F(x_1, \ldots, x_m) = Tr([e^{\mathbf{A}}e^{\mathbf{B}}]^N).$$

On the other hand

$$F(a, \ldots, a) = Tr([e^{a\mathbf{A}}e^{a\mathbf{B}}]^m) = Tr([e^{\mathbf{A}/K}e^{\mathbf{B}/K}]^{KN})$$

and Lemma 1 asserts that

$$Tr([e^{\mathbf{A}}e^{\mathbf{B}}]^N) \geq Tr([e^{\mathbf{A}/K}e^{\mathbf{B}/K}]^{KN})$$

Hold N fixed and let $K \to \infty$, using Sophus Lie's formula. □

Proof of (2.3). For any nonnegative $n \times n$ matrix $\mathbf{M}$,

$$r(\mathbf{M}) = \lim_{N\to\infty} [Tr(\mathbf{M}^N)]^{1/N}.$$

Apply this to the inequality of Lemma 2, using $\mathbf{M} = e^{\mathbf{A}}e^{\mathbf{B}}$ on the left and $\mathbf{M} = e^{\mathbf{A}+\mathbf{B}}$ on the right. □

The central idea in this proof of (2.3) is to use the Feynman-Kac formula. It seems to be harder to find a proof of the strict inequality in (2.3).

4. Open Problems. The concluding section is devoted to conjectures and open problems.

The first three conjectures, taken from CFKK (pp. 92–93), arose in attempts to find proofs of (2.3) and related results.

CONJECTURE 1. *Let* $\mathbf{A}$ *be an* $n \times n$ *essentially nonnegative matrix,* $\mathbf{B}$ *an* $n \times n$ *real diagonal matrix. Then* $F(t) = \log r(e^{\mathbf{A}t}e^{\mathbf{B}t})$ *is convex in the real variable* t. *If, in addition,* $\mathbf{A}$ *is irreducible and* $\mathbf{B}$ *is not a scalar matrix, then* $F(t)$ *is strictly convex in* t.

The conjecture is proved only for 2×2 matrices. I have checked it numerically with examples of 3×3 matrices, including matrices $\mathbf{A}$ with real and complex spectra. For $n \times n$ matrices, it is not hard to show that $F(t) + F(-t) \geq 0 = 2F(0)$. If r is replaced by Tr, then $F(t)$ is not convex in t for some 3×3 matrices $\mathbf{A}$ with complex spectra. If the stated conjecture is true, it provides another proof of both the weak and strict inequality (2.3) via Theorem 5 of CFKK (p. 78).

CONJECTURE 2. *Let* $\mathbf{A}_1, \ldots, \mathbf{A}_k$ *be nonnegative irreducible* $n \times n$ *matrices with positive diagonal elements, for some positive integer* k. *Let* $\mathbf{D}_1, \ldots, \mathbf{D}_k$ *be real diagonal* $n \times n$ *matrices with zero trace. Then*

$$f(\mathbf{D}_1, \ldots, \mathbf{D}_k) = \log r(\mathbf{A}_1 e^{\mathbf{D}_1} \ldots \mathbf{A}_k e^{\mathbf{D}_k})$$

is a strictly convex function of $(\mathbf{D}_1, \ldots, \mathbf{D}_k)$.

If true, this conjecture would provide sufficient conditions for strict inequality in a generalization (Theorem 3 of CFKK, pp. 71–72) of (2.3).

CONJECTURE 3. *Let* $\mathbf{A}$ *and* $\mathbf{B}$ *be* $n \times n$ *Hermitian matrices and* $a_i \geq 0$, $b_i \geq 0$, $i = 1, \ldots, n$. *Let* $a = \Sigma_i a_i$, $b = \Sigma_i b_i$. *Then*

$$\|e^{a_1\mathbf{A}}e^{b_1\mathbf{B}} \ldots e^{a_k\mathbf{A}}e^{b_k\mathbf{B}}\| \leq r(e^{2a\mathbf{A}}e^{2b\mathbf{B}})^{1/2},$$
$$r(e^{a_1\mathbf{A}}e^{b_1\mathbf{B}} \ldots e^{a_k\mathbf{A}}e^{b_k\mathbf{B}}) \leq r(e^{a\mathbf{A}}e^{b\mathbf{B}}).$$

The first of these inequalities is known to be true if, for some nonnegative scalar c, and

$i = 1, \dots, k$, we have $b_i = ca_i$. The second inequality is known to be true when the same conditions hold and in addition $\max a_i \leq a/2$.

Instead of the initial condition $P[V(0) = i] = 1$ assumed in section 2, assume that $P[V(0) = i] = \pi_i \geq 0$, $i = 1, \dots, n$, where $\pi = (\pi_i)$ is an equilibrium probability row vector of $\mathbf{Q}$, i.e. $\pi^T\mathbf{Q} = 0$, $\Sigma_i\pi_i = 1$. Denote the expectations of $N(t)$ and $N^D(t)$ under these initial conditions by $E_\pi N(t)$ and $E_\pi N^D(t)$. Let $\mathbf{1}$ be the n-vector with all elements equal to 1. It is known (CFKK, p. 76) that, for integral times t,

$$E_\pi N(t) = \pi^T e^{(\mathbf{Q}+\mathbf{B})t}\mathbf{1} \leq \mathrm{E}_\pi \mathrm{N}^\mathrm{D}(\mathrm{t}) = \pi^\mathrm{T}(e^{\mathbf{Q}}e^{\mathbf{B}})^t\mathbf{1},$$

with strict inequality if $t \geq 1$, $\mathbf{Q}$ is irreducible and at least two diagonal elements of $\mathbf{B}$ are distinct.

cohen

What about higher moments of $N(t)$ and $N^D(t)$? For simplicity, take $\mathbf{Q}$ to be irreducible and $N_0 = 1$. Then $\pi > 0$ elementwise. For any real c,

$$E_\pi N^c(t) = E\exp[\Sigma_j b_j g_j(t)]c = E\exp[\Sigma_j(cb_j)g_j(t)] = \pi^T e^{(\mathbf{Q}+c\mathbf{B})T}\mathbf{1},\ t \geq 0,$$

and similarly

$$E_\pi[N^D(t)]^c = \pi^T(e^{\mathbf{Q}}e^{c\mathbf{B}})^t\mathbf{1},\ t = 0, 1, 2, \dots$$

Since $\mathrm{Var}\,N(t) \geq 0$ and $\mathrm{Var}\,N^D(t) \geq 0$, with both inequalities strict provided $t > 0$ and at least two diagonal elements of $\mathbf{B}$ are distinct, we have

$$\pi^T e^{(\mathbf{Q}+2\mathbf{B})t}\mathbf{1} \geq (\pi^T e^{(\mathbf{Q}+\mathbf{B})t}\mathbf{1})^2,\ t \geq 0,$$

$$\pi^T(e^{\mathbf{Q}}e^{2\mathbf{B}})^t\mathbf{1} \geq (\pi^T(e^{\mathbf{Q}}e^{\mathbf{B}})^t\mathbf{1})^2,\ t = 0, 1, 2, \dots,$$

with strict inequalities under the conditions stated.

More generally, I can prove that if $\mathbf{A}$ is an essentially nonnegative $n \times n$ matrix, $\mathbf{B}$ is a diagonal real $n \times n$ matrix, and $\mathbf{x}$ and $\mathbf{y}$ are nonnegative n-vectors $(1 < n < \infty)$, then

$$(\mathbf{x}^T e^{(\mathbf{A}+2\mathbf{B})t}\mathbf{y})(\mathbf{x}^T e^{\mathbf{A}t}\mathbf{y}) \geq (\mathbf{x}^T e^{(\mathbf{A}+\mathbf{B})t}\mathbf{y})^2,\ t \geq 0,$$

$$(\mathbf{x}^T[e^{\mathbf{A}}e^{2\mathbf{B}}]^t\mathbf{y})(\mathbf{x}^T e^{\mathbf{A}t}\mathbf{y}) \geq (\mathbf{x}^T[e^{\mathbf{A}}e^{\mathbf{B}}]^t\mathbf{y})^2,\ t = 0, 1, 2, \dots,$$

with strict inequality if $t > 0$, $\mathbf{A}$ is irreducible, $\mathbf{x} > 0$, $\mathbf{y} > 0$, and at least two diagonal elements of $\mathbf{B}$ are distinct. Also,

$$r(e^{\mathbf{A}+2\mathbf{B}})r(e^{\mathbf{A}}) \geq r^2(e^{\mathbf{A}+\mathbf{B}}),$$

$$r(e^{\mathbf{A}}e^{2\mathbf{B}})r(e^{\mathbf{A}}) \geq r^2(e^{\mathbf{A}}e^{\mathbf{B}}).$$

If $\mathbf{A}$ is irreducible and at least two diagonal elements of $\mathbf{B}$ are distinct, the preceding inequalities are strict. The preceding inequalities hold if r is replaced throughout by Tr.

CONJECTURE 4. *Let* $\mathbf{A}$ *be an essentially nonnegative matrix and* $\mathbf{B}$ *a diagonal real matrix. Then*

$$r(e^{2\mathbf{B}}e^{\mathbf{A}})r(e^{\mathbf{A}}) - r^2(e^{\mathbf{B}}e^{\mathbf{A}}) \geq r(e^{2\mathbf{B}+\mathbf{A}})r(e^{\mathbf{A}}) - r^2(e^{\mathbf{B}+\mathbf{A}}), \tag{4.1}$$

If $\mathbf{u} \geq 0$, $\mathbf{v} \geq 0$ *are n-vectors such that*

$$\mathbf{u}^T e^{\mathbf{A}} = \mathbf{u}^T r(e^{\mathbf{A}}), \tag{4.2}$$

$$e^{\mathbf{A}}\mathbf{v} = r(e^{\mathbf{A}})\mathbf{v}, \tag{4.3}$$

then for positive integers θ,

$$(4.4)\qquad \begin{aligned}&[\mathbf{u}^T(e^{2\mathbf{B}}e^{\mathbf{A}})^\theta\mathbf{v}][\mathbf{u}^Te^{\mathbf{A}\theta}\mathbf{v}]-[\mathbf{u}^T(e^{\mathbf{B}}e^{\mathbf{A}})^\theta\mathbf{v}]^2\\ &\geq[\mathbf{u}^Te^{(2\mathbf{B}+\mathbf{A})\theta}\mathbf{v}][\mathbf{u}^Te^{\mathbf{A}\theta}\mathbf{v}]-[\mathbf{u}^Te^{(\mathbf{B}+\mathbf{A})\theta}\mathbf{v}]^2.\end{aligned}$$

When $\mathbf{A}$ *is irreducible and* $\mathbf{B}$ *is not a scalar matrix (i.e. at least two diagonal elements of* $\mathbf{B}$ *are distinct), both inequalities are strict.*

If either (4.2) or (4.3) fails to hold, then the conjectured inequality (4.4) need not hold. For example, let

$$\mathbf{A}=\begin{pmatrix}1&1\\1&1\end{pmatrix},\mathbf{B}=\begin{pmatrix}1&0\\0&0\end{pmatrix}$$
$$\mathbf{u}=\begin{pmatrix}1\\0\end{pmatrix},\mathbf{v}=\begin{pmatrix}1\\1\end{pmatrix}$$

so that (4.3) holds but (4.2) fails. Then with $\theta = 1$, the left member of (4.4) is 0 while the right member is approximately 18.6917.

If $\mathbf{A}$ and $\mathbf{B}$ are complex Hermitian matrices, (4.1) need not hold. Numerous numerical examples have failed to falsify Conjecture 4. Conjecture 4 implies that

$$\operatorname{Var}N(t)\leq\operatorname{Var}N^D(t),\, t=0,1,2,\,\ldots\,,$$

and gives (potentially) sufficient conditions for strict inequality.

CONJECTURE 5. *Under the assumptions of Conjecture 4, for every nonnegative integer* k,

$$r^k(e^{\mathbf{A}}e^{2\mathbf{B}})r^k(e^{\mathbf{A}})-r^{2k}(e^{\mathbf{A}}e^{\mathbf{B}})\geq r^k(e^{\mathbf{A}/2}e^{\mathbf{B}}e^{\mathbf{A}/2}e^{\mathbf{B}})r^k(e^{\mathbf{A}})-r^{4k}(e^{\mathbf{A}/2}e^{\mathbf{B}/2}),$$

with strict inequality under the additional conditions given in Conjecture 4.

The use of Lie's formula shows that Conjecture 5 implies (4.1). For

$$\begin{aligned}&r(e^{\mathbf{A}+2\mathbf{B}})r(e^{\mathbf{A}})-r^2(e^{\mathbf{A}+\mathbf{B}})\\ &=\lim_{k\to\infty}[r^{2^k}(e^{\mathbf{A}/2^k}e^{2\mathbf{B}/2^k})r^{2^k}(e^{\mathbf{A}/2^k})-r^{2^{k+1}}(e^{\mathbf{A}/2^k}e^{\mathbf{B}/2^k})]\end{aligned}$$

and (4.1) would hold if the sequence in brackets on the right were a decreasing function of k. The left side of (4.1) is the case when $k = 0$.

CONJECTURE 6. *Let* $\mathbf{E}$ *and* $\mathbf{F}$ *be* $n \times n$ *nonnegative matrices (elementwise),* $\mathbf{F}$ *diagonal. Then for every nonnegative integer* k, *the function*

$$H(t)=r^k(\mathbf{E}^2\mathbf{F}^{2(1+t)})r^k(\mathbf{E}^2\mathbf{F}^{2(1-t)})-r^{2k}(\mathbf{E}\mathbf{F}^{1+t})r^{2k}(\mathbf{E}\mathbf{F}^{1-t})$$

is convex in t *on* $[-1, +1]$. *If* $\mathbf{E} > 0$ *elementwise and all diagonal elements of* $\mathbf{F}$ *are positive and at least two of them are distinct, then* $H(t)$ *is strictly convex.*

I claim Conjecture 6 implies Conjecture 5. For

$$\begin{aligned}H(1)&=r^k(\mathbf{E}^2\mathbf{F}^4)r^k(\mathbf{E}^2)-r^{2k}(\mathbf{E}\mathbf{F}^2)r^{2k}(\mathbf{E})\\&=H(-1)\end{aligned}$$

and by convexity $(H(1) + H(-1))/2 = H(1) \geq H(0)$. Thus

$$r^k(\mathbf{E}^2\mathbf{F}^4)r^k(\mathbf{E}^2)-r^{2k}(\mathbf{E}\mathbf{F}^2)r^{2k}(\mathbf{E})\geq r^{2k}(\mathbf{E}^2\mathbf{F}^2)-r^{4k}(\mathbf{E}\mathbf{F}).$$

Letting $\mathbf{E} = e^{\mathbf{A}/2}$, $\mathbf{F} = e^{\mathbf{B}/2}$ and rearranging terms gives

$$r^k(e^{\mathbf{A}}e^{2\mathbf{B}})r^k(e^{\mathbf{A}})-r^{2k}(e^{\mathbf{A}}e^{\mathbf{B}})\geq r^{2k}(e^{\mathbf{A}/2}e^{\mathbf{B}})r^{2k}(e^{\mathbf{A}/2})-r^{4k}(e^{\mathbf{A}/2}e^{\mathbf{B}/2})$$

which is equivalent to Conjecture 5.

The illustrative inequalities in section 2 for a random evolution driven by a reversible Markov chain are special cases of Corollary 4 of CFKK (p. 62), which states: If $\mathbf{A} = \mathbf{D}\mathbf{S}\mathbf{D}^{-1}$ where $\mathbf{S}$ is symmetric and $\mathbf{D}$ is diagonal nonsingular, $\mathbf{B}$ is diagonal real, and φ is a real-valued continuous function of the eigenvalues of its matrix argument, finite when all elements of its argument are finite, such that

(4.5) $$\varphi([\mathbf{MM}^*]^k) \geqslant |\varphi(\mathbf{M}^{2k})|, k = 1, 2, \ldots ,$$

for every $n \times n$ complex matrix $\mathbf{M}$, then

(4.6) $$\varphi(e^{\mathbf{A}}e^{\mathbf{B}}) \geqslant \varphi(e^{\mathbf{A}+\mathbf{B}}).$$

Does this result have a converse? If true, Conjecture 7 would provide a new characterization of reversibility.

CONJECTURE 7. *Let* $\mathbf{A}$ *be an essentially nonnegative matrix such that* (4.6) *holds for every diagonal real matrix* $\mathbf{B}$ *and every* φ, *as just described, that satisfies* (4.5). *Then there exists a symmetric matrix* $\mathbf{S}$ *and a diagonal nonsingular matrix* $\mathbf{D}$ *such that*

$$\mathbf{A} = \mathbf{DSD}^{-1}$$

So far in this paper, I have considered only the case where $N(t)$ is a real scalar. However, if the vat of bacteria contains more than one species of bacteria, or more than one genotype of the same species, or subgroups of a species differentiated by physical or biochemical markers, it is natural to try to model the simultaneous evolution of all distinguishable types.

Consider the following k-dimensional random evolution, where k is a fixed positive integer greater than 1. Let $\mathbf{B}_1, \ldots, \mathbf{B}_n$ be $k \times k$ real matrices and (as before) $V(t)$ a homogeneous continuous-time irreducible Markov process on the state space $\{1, \ldots, n\}$, with $V(0) = 1$, and with intensity matrix $\mathbf{Q}$. Let $\mathbf{N}(t)$ be a k-vector that evolves according to

$$d\mathbf{N}(t)/dt = \mathbf{B}_{V(t)}\mathbf{N}(t) \quad \text{for} \quad t \geqslant 0, \mathbf{N}(0) = \mathbf{N}_0.$$

For biological applications, the vector $\mathbf{N}(t)$ of number of individuals of each type is required to be nonnegative. Given $\mathbf{N}_0 \geqslant 0$, a condition sufficient to guarantee that, for all $t \geqslant 0$, $\mathbf{N}(t) \geqslant 0$ is that each $\mathbf{B}_j$ is essentially nonnegative, i.e., every off-diagonal element of $\mathbf{B}_j$ is nonnegative, $j = 1, \ldots, n$. I henceforth assume that every $\mathbf{B}_j$ is essentially nonnegative.

This assumption makes the model more relevant to biological situations where an increase in the number of one type of bacteria leads to an increase (or no decrease) in the number of other types, as when types correspond to genotypes. It makes the model less relevant to biological situations where the types are different species, some of which consume other species.

Under these assumptions, I can prove that, if $||.||$ is any vector norm,

$$\lim t^{-1} \log || E\,\mathbf{N}(t) || = \log r(e^{\mathbf{A}+\mathbf{B}}),$$

where $\mathbf{A}$ and $\mathbf{B}$ are both $(kn) \times (kn)$ matrices defined by

$$\mathbf{A} = \mathbf{Q} \otimes \mathbf{I}_k, \ \mathbf{I}_k = k \times k \ \text{identity matrix}$$
$$\mathbf{B} = \operatorname{diag}(\mathbf{B}_1, \ldots, \mathbf{B}_n).$$

The same norm applied to matrices means the matrix norm induced by the chosen vector norm. It is easy to show that

$$\lim t^{-1} E \log || \mathbf{N}(t) || \leqslant \Sigma_{i=1}^{n} \pi_i || \mathbf{B}_i ||.$$

QUESTION 1. *Is there a simple exact expression for lim* t^{-1} *E log* $|| \mathbf{N}(t) ||$ *(analogous to that in the one-dimensional case)?*

As in the one-dimensional case, it is natural to suppose that a biologist who observed this k-dimensional random evolution would construct a discrete-time approximation $\mathbf{N}^D(t)$ according to

$$\mathbf{N}^D(t+1) = \exp(\mathbf{B}_{V(t)})\mathbf{N}^D(t),\ t = 0, 1, \ldots, \mathbf{N}^D(0) = \mathbf{N}_0.$$

$$P[\mathbf{B}_{V(t+1)} = \mathbf{B}_j | \mathbf{B}_{V(t)} = \mathbf{B}_i] = (e^{\mathbf{Q}})_{ij} = p_{ij},\ \mathbf{B}_{V(0)} = \mathbf{B}_1.$$

Thus $\mathbf{N}^D(t)$ is determined by a Markovian product of random nonnegative matrices according to

$$\mathbf{N}^D(t) = \exp(\mathbf{B}_{V(t-1)})\exp(\mathbf{B}_{V(t-2)}) \ldots \exp(\mathbf{B}_{V(0)})\mathbf{N}_0.$$

It follows from Furstenberg and Kesten (1960) that $\lim t^{-1} \log ||\mathbf{N}^D(t)||$ exists and equals $\lim t^{-1} E \log ||\mathbf{N}^D(t)||$ with probability 1.

QUESTION 2. *Is there a simple exact expression for* $\lim t^{-1} E \log ||\mathbf{N}^D(t)||$?

QUESTION 3. *What is the relation between* $\lim t^{-1} E \log ||\mathbf{N}(t)||$ *and* $\lim t^{-1} E \log ||\mathbf{N}^D(t)||$?

It is easy to show that

$$\lim t^{-1} \log ||E\, \mathbf{N}^D(t)|| = \log r(e^{\mathbf{A}}e^{\mathbf{B}}),$$

where $\mathbf{A}$ and $\mathbf{B}$ are the $(kn) \times (kn)$ matrices defined above. This follows from a formula for $E\, \mathbf{N}^D(t)$ that I published (Cohen, 1977) without knowing that it had been previously derived in a never-published (so far as I know) report of Bharucha (1960). The above formula is a Feynman-Kac formula for products of random matrices. The expression $\log r(e^{\mathbf{A}+\mathbf{B}})$ for the continuous-time random evolution may be derived by constructing a sequence of approximations to $\mathbf{N}(t)$ using ever finer subdivisions of time, as in the one-dimensional case.

What is the relation between $\lim t^{-1} \log ||E\, \mathbf{N}(t)||$ and $\lim t^{-1} \log ||E\, \mathbf{N}^D(t)||$? The following conjecture would cover the special case when all the $\mathbf{B}_j$ matrices commute with one another.

CONJECTURE 8. *If n essentially nonnegative $k \times k$ matrices $\mathbf{B}_j$ satisfy*

$$\mathbf{B}_i\mathbf{B}_j = \mathbf{B}_j\mathbf{B}_i, \quad i, j = 1, \ldots, n,$$

then

$$r(e^{\mathbf{A}+\mathbf{B}}) \leq r(e^{\mathbf{A}}e^{\mathbf{B}}),$$

where $\mathbf{R}$ is any essentially nonnegative $n \times n$ matrix and

$$\mathbf{A} = \mathbf{R} \otimes \mathbf{I}_k,$$
$$\mathbf{B} = \mathrm{diag}(\mathbf{B}_1, \ldots, \mathbf{B}_n).$$

If $\mathbf{R}$ and $\mathbf{B}_i$, $i = 1, \ldots, n$ are all irreducible, then the inequality is strict.

The conjectured inequality need not hold if the assumed commutativity is not true. For example, if

$$\mathbf{R} = \begin{pmatrix} 1 & 1 \\ 1 & 1 \end{pmatrix}, \mathbf{B}_1 = \begin{pmatrix} 1 & 1 \\ 2 & 1 \end{pmatrix}, \mathbf{B}_2 = \begin{pmatrix} 1 & 2 \\ 1 & 1 \end{pmatrix}.$$

then $\mathbf{B}_1\mathbf{B}_2 \neq \mathbf{B}_2\mathbf{B}_1$ and

$$r(\exp[\mathbf{R} \otimes \mathbf{I}_2 + \mathrm{diag}(\mathbf{B}_1, \mathbf{B}_2)]) = 85.583$$
$$> r(\exp[\mathbf{R} \otimes \mathbf{I}_2]\exp[\mathrm{diag}(\mathbf{B}_1, \mathbf{B}_2)]) = 84.671$$

(These numerical computations, and other tests of conjectures, were performed on the MATLAB system of Moler [1981] as implemented on SCORE in the Stanford University Computer Science Department.)

A sufficient condition for Conjecture 8 is:

CONJECTURE 9. *If* $\mathbf{R}$ *is an elementwise nonnegative* $n \times n$ *matrix and* $\mathbf{B}_i$, $i = 1, \ldots,$ n *are commuting (elementwise) nonnegative* $k \times k$ *matrices, then*

$$r(\mathbf{A}^2\mathbf{B}^2) \geq r^2(\mathbf{AB})$$

where $\mathbf{A}$ *and* $\mathbf{B}$ *are defined in Conjecture 8. If* $\mathbf{R}$ *and all* $\mathbf{B}_i$ *are (elementwise) positive, then the inequality is strict.*

The counterexample to Conjecture 8 without commutativity also shows that Conjecture 9 fails without commutivity. In this case

$$\mathbf{A}^2\mathbf{B}^2 = \begin{pmatrix} 6 & 4 & 6 & 8 \\ 8 & 6 & 4 & 6 \\ 6 & 4 & 6 & 8 \\ 8 & 6 & 4 & 6 \end{pmatrix}, \mathbf{AB} = \begin{pmatrix} 1 & 1 & 1 & 2 \\ 2 & 1 & 1 & 1 \\ 1 & 1 & 1 & 2 \\ 2 & 1 & 1 & 1 \end{pmatrix}.$$

Since all row sums of $\mathbf{A}^2\mathbf{B}^2$ are 24, $r(\mathbf{A}^2\mathbf{B}^2) = 24$. Since all row sums of $\mathbf{AB}$ are 5, $r(\mathbf{AB}) = 5$. Thus $r(\mathbf{A}^2\mathbf{B}^2) < r^2(\mathbf{AB})$, contrary to the desired inequality.

So far we have considered two measures of the long-run growth rate of the k-dimensional random evolution $\mathbf{N}(t)$ in continuous time, namely $\lim t^{-1} E \log \|\mathbf{N}(t)\|$ and $\lim t^{-1} \log \|E\, \mathbf{N}(t)\|$ (with corresponding measures for the discrete-time approximation $\mathbf{N}^D(t)$). Another plausible measure is $\lim t^{-1} \log E \|\mathbf{N}(t)\|$. By the triangle inequality for norms,

$$\lim t^{-1} \log \|E\,\mathbf{N}(t)\| \leq \lim t^{-1} \log E \|\mathbf{N}(t)\|.$$

I claim that this inequality is in fact an equality for any vector norm. For any real or complex k-vector x, the vector norm defined by

$$\|x\|_1 = \Sigma_j |x_j|$$

is the Hölder p-norm for $p = 1$. By construction, $\mathbf{N}(t) \geq 0$. So $\|\mathbf{N}(t)\|_1 = \Sigma_j N_j(t)$ and $\|E\, \mathbf{N}(t)\|_1 = \Sigma_j[E\mathbf{N}(t)]_j = \Sigma_j E[\mathbf{N}(t)]_j = E\, \Sigma_j N_j(t) = E\, \|\mathbf{N}(t)\|_1$. For any other vector norm $\|.\|$, there exist constants c and c_1 depending on $\|.\|$, such that for all x, $\|x\| \leq c_1\|x\|_1$ and $\|x\|_1 \leq c\|x\|$ (see e.g. Lancaster, 1977, pp. 199, 204). Therefore

$$\begin{aligned} \lim t^{-1} \log E \|\mathbf{N}(t)\| &\leq \lim t^{-1} \log c_1 \mathrm{E}\, \|\mathbf{N}(t)\|_1 \\ &= \lim t^{-1} \log E \|\mathbf{N}(t)\|_1 \\ &= \lim t^{-1} \log \|\mathrm{E}\mathbf{N}(t)\|_1 \\ &\leq \lim t^{-1} \log c\, \|E\mathbf{N}(t)\| \\ &= \lim t^{-1} \log \|E\mathbf{N}(t)\|. \end{aligned}$$

Combining this with the reverse inequality previously established shows that

$$\lim t^{-1} \log \|E\,\mathbf{N}(t)\| = \lim t^{-1} \log E \|\mathbf{N}(t)\|.$$

Acknowledgments. J. E. C. acknowledges fellowships from the John Simon Guggenheim Memorial Foundation, New York, the Center for Advanced Study in the Behavioral Sciences, Stanford, and the John D. and Catherine T. MacArthur Foundation, Chicago; National Science Foundation grant DEB80–11026 and grants to the Center for Advanced Study from the National Institute of Mental Health (5T32MH14581–06) and the Exxon Education Foundation; and Mr. and Mrs. William T. Golden of Olive Bridge, New York, for their hospitality during this work. I thank Charles M. Newman for permission to give his proof in section 3, and Cleve Moler for introducing me to MATLAB.

REFERENCES

BHARUCHA, B. H. (1960). Stability of random linear systems with Markov parameters (Part 1). Electronics Research Lab., Dept. of Electrical Engineering, University of California, Berkeley ERD TN-60-776.

BLANC–LAPIERRE, A. and FORTET, R. (1953). *Théorie des Fonctions Aléatoires*. Masson, Paris. English trans. (1965). *Theory of Random Functions, Vol. 1*. Gordon and Breach Science Publishers, New York.

ÇINLAR, E. (1975). Markov renewal theory: a survey. *Management Science 21* 727–752.

COHEN, J. E. (1977). Ergodicity of age structure in populations with Markovian vital rates, III: Finite-state moments and growth rates; illustration. *Advances in Appl. Probability 1* 462–475.

COHEN, J. E. (1979a). Random evolutions and the spectral radius of a non-negative matrix. *Math. Proc. Camb. Phil. Soc. 86* 345–350.

COHEN, J. E. (1979b). Random evolutions in discrete and continuous time. *Stochastic Processes Appl. 9* 245–251.

COHEN, J. E., FRIEDLAND, S., KATO, T., and KELLY, F. P. (1982). Eigenvalue inequalities for products of matrix exponentials. *Linear Algebra and Appl. 45* 55–95.

DAGSVIK, J. (1978). Stability and heavy traffic results for the general bulk queue. *Advances in Appl. Probability 10* 213–231.

DOEBLIN, W. (1937). On the asymptotic properties of movement governed by certain types of simple chains. (In French) *Bulletin Mathématique de la Société Roumaine des Sciences 39* 57–115.

DURRETT, R. and RESNICK, S. (1978). Functional limit theorems for dependent variables. *Ann. Probability 6* 829—846.

ELLIS, R. S. (1974). Limit theorems for random evolutions with explicit error estimates. *Z. Wahrscheinlichkeitstheorie und verw. Gebiete 28* 249–256.

FRECHET, M. (1952). *Méthode des fonctions arbitraires: théorie des événements en chaine dans le cas d'un nombre fini d'états possibles*. 2nd ed. Gauthier-Villars, Paris.

FRIEDLAND, S. (1981). Convex spectral functions. *Lin. and Multilin. Alg. 9* 299–316.

FUKUSHIMA, M. and HITSUDA, M. (1967). On a class of Markov processes taking values on lines and the central limit theorem. *Nagoya Math. J. 30* 47–56.

FURSTENBERG, H. and KESTEN, H. (1960). Products of random matrices. *Ann. Math. Statist. 31* 457–469.

GRIEGO, R. and HERSH, R. (1971). Theory of random evolutions with applications to partial differential equations. *Trans. Amer. Math. Soc. 156* 405–418.

HERSH, R. (1974). Random evolutions: a survey of results and problems. *Rocky Mountain J. Math. 4* 443–477.

IOSIFESCU, M. (1980). *Finite Markov Processes and Their Applications*. Wiley, New York.

KAC, M. (1956). Some stochastic problems in physics and mathematics. *Magnolia Petroleum Co., Lectures in Pure and Applied Sciences, no. 2*.

KATZ, R. W. (1977). An application of chain-dependent processes to meteorology. *J. Appl. Probability 14* 598–603.

KEEPLER, M. (1976). Perturbation theory for backward and forward random evolutions. *J. Math. Kyoto Univ. 16* 395–411.

KEILSON, J. and WISHART, D. M. G. (1964). A central limit theorem for processes defined on a finite Markov chain. *Proc. Camb. Phil. Soc. 60* 547–567.

KEILSON, J. and WISHART, D. M. G. (1967). Addenda to processes defined on a finite Markov chain. *Proc. Camb. Phil. Soc. 63* 187–193.

KEILSON, J. and RAO, S. S. (1970). A process with chain dependent growth rate. *J. Appl. Probability 7* 699–711.

KEMENY, J. G. and SNELL, J. L. (1976). *Finite Markov Chains*. Springer-Verlag, Heidelberg.

KERTZ, R. P. (1974). Limit theorems for discontinuous random evolutions with applications to initial value problems and to Markov processes on N lines. *Ann. Probability 2* 1046–1064.

LANCASTER, P. (1977). *Theory of Matrices*. Calgary, Alberta: Author's Reprint of 1969 Edition of Academic Press, New York.

LEBRAS, H. (1974). Populations stables aléatoires. *Population* (Paris) *29* 435–464.

MEYER, P. A. (1962). Fonctionnelles multiplicatives et additives de Markov. *Ann. Inst. Fourier* (Grenoble) *12* 125–230.

MOLER, C. (1981). *MATLAB Users' Guide*, revised June 1981. Dept. of Computer Science, University of New Mexico, Tech. Rep. CS81-1.

O'BRIEN, G. (1974). Limit theorems for sums of chain dependent processes. *J. Appl. Probability 11* 582–587.

PINSKY, M. (1968). Differential equations with a small parameter and the central limit theorem for functions defined on a finite Markov chain. *Z. Wahrscheinlichkeitstheorie und verw. Gebiete 9* 101–111.

SERFOZO, R. F. (1975). Functional limit theorems for stochastic processes based on embedded processes. *Advances in Appl. Probability 7* 123–239.

SIEGRIST, K. (1981). Random evolution processes with feedback. *Trans. Amer. Math. Soc. 265* 375–392.

VOLKOV, I. S. (1958). On the distribution of sums of random variables defined on a homogeneous Markov chain with a finite number of states. *Theor. Probability Appl. 34* 384–399.

WOLFSON, D. B. (1977). Limit theorems for sums of a sequence of random variables defined on a Markov chain. *J. Appl. Probability 14* 614–620.

Inequalities in Statistics and Probability
IMS Lecture Notes–Monograph Series Vol. 5 (1984), 54–58

ON TP_2 AND LOG-CONCAVITY

BY SOMESH DAS GUPTA and S. K. SARKAR
Indian Statistical Institute and Temple University

Inter-relations between the TP_2 property and log-concavity of density functions have been investigated. The general results are then applied to noncentral chi-square density functions and beta density functions.

1. Results on Density Functions

Definition 1. A function $f: \mathcal{R}^2 \to \mathcal{R}$ is said to be TP_2 (Karlin (1968)) if, for $x_1 < x_2$, $y_1 < y_2$

$$f(x_1,y_2)f(x_2,y_1) \leq f(x_1,y_1)f(x_2,y_2). \tag{1.1}$$

We shall say that $1/f$ is TP_2, if (1.1) holds for f with the inequality reversed.

Let X be a positive random variable having the p.d.f. $f(\cdot,\theta,\lambda)$ with respect to Lebesgue measure; $\theta > 0$, $\lambda \geq 0$.

Definition 2. The p.d.f. $f(x,\theta,\lambda)$ is said to have the reproductive property (RP) in θ, if there exists a distribution function $G(\cdot,s)$ on $\mathcal{R}^+$ $(s > 0)$ such that

$$\int_0^x f(x-y,\theta,\lambda)\, G(dy,s) = f(x,\theta+s,\lambda). \tag{1.2}$$

THEOREM 1. *Suppose $f(x,\theta,\lambda)$ has the RP in θ. Then* (i) $f(x,\theta,\lambda)$ TP_2 *in* $(x,\lambda) \to 1/f(x,\theta,\lambda)$ TP_2 *in* (θ,λ), (ii) $f(x,\theta,\lambda)$ TP_2 *in* $(x,\theta) \to f(x,\theta,\lambda)$ *log-concave in* θ.

Proof. (i) For $0 < x_1 < x_2$, $\lambda_1 < \lambda_2$ we have

$$f(x_2,\theta,\lambda_1)f(x_1,\theta,\lambda_2) \leq f(x_2,\theta,\lambda_2)f(x_1,\theta,\lambda_1). \tag{1.3}$$

Write $x_1 = x_2 - y$. Integrating (1.3) with respect to $G(dy,s)$ we get

$$f(x_2,\theta,\lambda_1)f(x_2,\theta+s,\lambda_2) \leq f(x_2,\theta,\lambda_2)f(x_2,\theta+s,\lambda_1), \tag{1.4}$$

which shows that $1/f(x,\theta,\lambda)$ is TP_2 in (θ,λ).

(ii) For $0 < x_1 < x_2$, $\theta_1 < \theta_2$, we have

$$f(x_1,\theta_2,\lambda)f(x_2,\theta_1,\lambda) \leq f(x_2,\theta_2,\lambda)f(x_1,\theta_1,\lambda). \tag{1.5}$$

Write $x_1 = x_2 - y$. Integrating (1.5) with respect to $G(dy,s)$ we get

$$f(x_2,\theta_2+s,\lambda)f(x_2,\theta_1,\lambda) \leq f(x_2,\theta_2,\lambda)f(x_2,\theta_1+s,\lambda), \tag{1.6}$$

which shows that $f(x,\theta,\lambda)$ is log-concave in θ. □

Definition 3. The p.d.f. $f(x,\theta,\lambda)$ is said to have the mixture property (MP) in (θ,λ) if there exists a non-negative random variable K with the distribution $H(\cdot,\tau)$ with $\tau > 0$ such that

$$\int_0^\infty f(x,\theta+k,\lambda)\, H(dk,\tau) = f(x,\theta,\lambda+\tau). \tag{1.7}$$

Suppose H in Definition 3 possesses a density function h with respect to a σ-finite measure ν.

[1] Supported by a grant from the National Science Foundation, Grant No. MSC8202209 at University of Minnesota.

AMS 1980 subject classifications. 60E15, 26D99.

Key words and phrases: Totally positive of order 2 (TP_2), log-concavity.

THEOREM 2. *Suppose* $f(x,\theta,\lambda)$ *has the MP in* (θ,λ). *Then* (i) $f(x,\theta,0)$ TP_2, $h(k,\tau)$ TP_2 $\rightarrow f(x,\theta,\tau)$ TP_2 *in* (x,τ), (ii) $1/f(x,\theta,\lambda)$ TP_2 *in* $(\theta,\lambda) \rightarrow f(x,\theta,\lambda)$ *log-concave in* λ, (iii) $f(x,\theta,\lambda)$ TP_2 *in* $(x,\theta) \rightarrow f(x,\theta,\lambda)$ TP_2 *in* (x,λ), (iv) $f(x,\theta,\lambda)$ *log-concave in* $\theta \rightarrow f(x,\theta,\lambda)$ *log-concave in* λ.

Proof: (i) This follows from Karlin (1968, p. 17). (ii) For $\theta_1 \leq \theta_2$, $\lambda_1 < \lambda_2$,

$$f(x,\theta_2,\lambda_2)f(x,\theta_1,\lambda_1) \leq f(x,\theta_2,\lambda_1)f(x,\theta_1,\lambda_2). \tag{1.8}$$

Write $\theta_2 = \theta_1 + k$. Integrating (1.8) with respect to $H(dk,\tau)$ we get

$$f(x,\theta_1,\lambda_2+\tau)f(x,\theta_1,\lambda_1) \leq f(x,\theta_1,\lambda_1+\tau)f(x,\theta_1,\lambda_2). \tag{1.9}$$

which shows that $f(x,\theta,\lambda)$ is log-concave in λ.

(iii) For $0 < x_1 < x_2$, $\theta_1 \leq \theta_2$ we have

$$f(x_2,\theta_1,\lambda)f(x_1,\theta_2,\lambda) \leq f(x_2,\theta_2,\lambda)f(x_1,\theta_1,\lambda). \tag{1.10}$$

Write $\theta_2 = \theta_1 + k$. Integrating (1.10) with respect to $H(dk,\tau)$ we get

$$f(x_2,\theta_1,\lambda)f(x_1,\theta_1,\lambda+\tau) \leq f(x_2,\theta_1,\lambda+\tau)f(x_1,\theta_1,\lambda), \tag{1.11}$$

which shows that $f(x,\theta,\lambda)$ is TP_2 in (x,λ).

(iv) For $\theta_1 \leq \theta_2$, $0 \leq s$ we have

$$f(x,\theta_2+s,\lambda)f(x,\theta_1,\lambda) \leq f(x,\theta_2,\lambda)f(x,\theta_1+s,\lambda). \tag{1.12}$$

Write $\theta_2 = \theta_1 + k$ and integrate (1.12) with respect to $H(dk,\tau_1)\,H(ds,\tau_2)$. Then we get

$$f(x,\theta_1,\lambda+\tau_1+\tau_2)f(x,\theta_1,\lambda) \leq f(x,\theta_1,\lambda+\tau_1)f(x,\theta_1,\lambda+\tau_2). \tag{1.13}$$

The above shows that $f(x,\theta,\lambda)$ is log-concave in λ. □

THEOREM 3. *Suppose* $f(x,\theta,\lambda)$ *is log-concave in* x. *Then*

$$f(x,\theta,\lambda) \textit{ has the RP in } \theta \rightarrow f(x,\theta,\lambda) \textit{ is } TP_2 \textit{ in } (x,\theta). \tag{1.14}$$

Proof. For $0 < x_1 < x_2$, $0 < y$ we have

$$f(x_1,\theta,\lambda)f(x_2+y,\theta,\lambda) \leq f(x_2,\theta,\lambda)f(x_1+y,\theta,\lambda). \tag{1.15}$$

Write $x_1 = x_2 - z$. Integrating (1.15) with respect to $G(dz,s)$ we get

$$f(x_2,\theta+s,\lambda)f(x_2+y,\theta,\lambda) \leq f(x_2,\theta,\lambda)f(x_2+y,\theta+s,\lambda), \tag{1.16}$$

which shows that $f(x,\theta,\lambda)$ is TP_2 in (x,θ). □

Define $C(x,\theta,\lambda)$ by

$$f(x,\theta,\lambda)\,C(x,\theta,\lambda) = f(\lambda,\theta,x), \tag{1.17}$$

LEMMA 1. *If both* $f(x,\theta,\lambda)$ *and* $C(x,\theta,\lambda)$ *are log-concave in* λ, *then* $f(x,\theta,\lambda)$ *is log-concave in* x.

The above lemma is a well-known fact; see Das Gupta (1976, 1980).

Combining the above results, we get the following:

THEOREM 4. Suppose the following conditions hold: (a) $f(x,\theta,\lambda)$ has the RP in θ, as defined in (1.2), (b) $f(x,\theta,\lambda)$ has the MP in (θ,λ), as defined in (1.7), (c) $C(x,\theta,\lambda)$, as defined in (1.17), is log-concave in λ. Then the following are equivalent: (i) $f(x,\theta,\lambda)$ is TP_2 in (x,λ), (ii) $1/f(x,\theta,\lambda)$ is TP_2 in (θ,λ), (iii) $f(x,\theta,\lambda)$ is log-concave in λ, (iv) $f(x,\theta,\lambda)$ is log-concave in x, (v) $f(x,\theta,\lambda)$ is TP_2 in (x,θ), (vi) $f(x,\theta,\lambda)$ is log-concave in θ.

Moreover, all the above results (i)–(vi), under the conditions (a)–(c), are implied by the condition (d) $f(x,\theta,0)$ is TP_2, $h(k,\tau)$ is TP_2.

2. Application to Noncentral Chi-Square Distribution. Suppose $f(x,\theta,\lambda)$ is the p.d.f. of the noncentral chi-square distribution with θ degrees of freedom and the noncentrality parameter λ. Then (1.2) holds with $G(\cdot,s)$ as the distribution of χ_s^2. Moreover, (1.7) holds if H is taken such that $K/2$ is distributed as Poisson with mean $\tau/2$. With this specification of h, condition (d) of Theorem 4 obtains. It can also be seen that $C(x,\theta,\lambda)$, as defined in (1.17), is given by

$$C(x,\theta,\lambda) = (\lambda/x)^{\theta/2-1}, \tag{2.1}$$

which is log-concave in λ if $\theta/2 \geqslant 1$. Hence (i)–(iii) of Theorem 4 hold when $\theta > 0$, and (iv)–(vi) hold when $\theta \geqslant 2$. It can be seen easily that $f(x,\theta,0)$ is log-concave in x when $\theta \geqslant 2$; also $f(x,\theta,0)$ is TP_2 in (x,θ), and $f(x,\theta,0)$ is log-concave in θ. Ghosh (1973) gave an alternative proof of the TP_2 property of $f(x,\theta,\lambda)$ in (x,θ) when $\theta > 2$. Karlin (1968) proved that $f(x,\theta,\lambda)$ is log-concave in x when $\theta > 2$.

Remark. The chain of arguments used in the above theorems can be used also for discrete random variables after minor modifications.

3. Results on C.D.F.'s. Let X be a positive r.v. with the p.d.f. $f(x,\theta,\lambda)$ with respect to Lebesgue measure. The c.d.f. of X is given by

$$F(C,\theta,\lambda) \equiv P[X \leqslant C] \equiv 1 - \bar{F}(C,\theta,\lambda). \tag{3.1}$$

LEMMA 2. (a) *If $f(x,\theta,\lambda)$ satisfies (1.2), then so does $F(x,\theta,\lambda)$.* (b) *If $f(x,\theta,\lambda)$ satisfies (1.7), then so does $F(x,\theta,\lambda)$*ς (c) *If $f(x,\theta,\lambda)$ is TP_2 in (x,θ)(or, in (x,λ)), then $F(x,\theta,\lambda)$ is also TP_2 in (x,θ) (or, in (x,λ)).* (d) *If $f(x,\theta,\lambda)$ is log-concave in x, then $F(x,\theta,\lambda)$ is also log-concave in x. The above results (b)–(d) also hold for $\bar{F}$.*

Proof. The results (a) and (b) are trivial. The results (c) follows from Karlin's (1968) theorem and the fact that the indicator function of the set $(-\infty,C]$ is TP_2 in (x,C). The result (d) follows from Prekopa's Theorem; see Das Gupta (1976, 1980). □

Remark. If f or F satisfies RP, then it trivially follows that $F(c,\theta,\lambda)$ is decreasing in θ; this fact also follows from the condition that $F(c,\theta,\lambda)$ is TP_2 in (c,θ).

THEOREM 5. (a) *Suppose $f(x,\theta,\lambda)$ or $F(x,\theta,\lambda)$ satisfies the RP in θ, as given in (1.2). Then* (i) $F(x,\theta,\lambda)$ TP_2 *in* $(x,\lambda) \to 1/F(x,\theta,\lambda)$ TP_2 *in* (θ,λ). (ii) $F(x,\theta,\lambda)$ TP_2 *in* $(x,\theta) \to F(x,\theta,\lambda)$ *log-concave in* θ. (iii) $F(x,\theta,\lambda)$ *log-concave in* $x \to F(x,\theta,\lambda)$ TP_2 *in* (x,θ).

(b) *If $f(x,\theta,\lambda)$ or $F(x,\theta,\lambda)$ satisfies the MP in (θ,λ) as given in (1.7), then* (i) $1/F(x,\theta,\lambda)$ TP_2 *in* $(\theta,\lambda) \to f(x,\theta,\lambda)$ *log-concave in* λ. (ii) $F(x,\theta,\lambda)$ TP_2 *in* $(x,\theta) \to F(x,\theta,\lambda)$ TP_2 *in* (x,λ). (iii) $F(x,\theta,\lambda)$ log-concave in $\theta \to F(x,\theta,\lambda)$ *log-concave in* λ. *The above results in (a) and (b) also hold if F is replaced by $\bar{F}$.*

This theorem can be proved following the proofs of Theorems 1 and 2. However, we need to note some additional facts in order to prove the results for $\bar{F}$. If $F(x,\theta,\lambda)$ satisfies (1.2), we get

$$\textstyle\int_0^c \bar{F}(c-y,\theta,\lambda)\, G(dy,s) = \bar{F}(c,\theta+s,\lambda) - \bar{G}(c,s). \tag{3.2}$$

So, in order for Theorem 5(a)(i) to hold for $\bar{F}$ we must have $\bar{F}(c,\theta,\lambda)$ increasing in λ; but this is implied by the condition that $\bar{F}(c,\theta,\lambda)$ is TP_2 in (c,λ). Also, for Theorem 5(a)(ii) to hold for $\bar{F}$ we need $\bar{F}(c,\theta,\lambda)$ increasing in θ; again this is implied by the condition that $\bar{F}(c,\theta,\lambda)$ is TP_2 in (c,θ).

4. Application to Chi-Square Distribution. (a) If $F(\cdot,\theta,\lambda)$ is the c.d.f. of the chi-square distribution with θ degrees of freedom and noncentrality parameter λ, then the results (i)–(iii) of Theorem 4 hold for F or $\bar{F}$ in place of f, and (iv)–(vi) hold for F or $\bar{F}$ in place of f when $\theta \geq 2$.

(b) If $f(\cdot,\theta,\lambda)$ is the p.d.f. of χ^2_θ, then it is well-known that $f(x,\theta,0)$ is TP_2. Hence, following the proof of Theorem 1(ii) it can be shown that both $F(c,\theta,0)$ and $\bar{F}(c,\theta,0)$ are log-concave in $\theta > 0$.

(c) It follows from Lemma 2 and the subsequent remark that both $F(c,\theta,\lambda)$ and $\bar{F}(c,\theta,\lambda)$ are log-concave in c when $\theta \geq 2$. However, a stronger result can be obtained when $\lambda = 0$ by appealing to Prekopa's Theorem.

Suppose $X \sim \chi^2_\theta$, and let f^* be the p.d.f. of $Y = \log X$. Then $f^*(y,\theta)$ is log-concave in y for $\theta > 0$. Using Prekopa's Theorem, we get

$$(4.1) \qquad F^*(\alpha_1 d_1 + \alpha_2 d_2, \theta) \geq [F^*(d_1,\theta)]^{\alpha_1}[F^*(d_2,\theta)]^{\alpha_2},$$

for any d_1, d_2 and $0 \leq \alpha_1, \alpha_2 \leq 1$, $\alpha_1 + \alpha_2 = 1$, where F^* is the c.d.f. corresponding to f^*. The above inequality is equivalent to

$$(4.2) \qquad F(c_1^{\alpha_1} c_2^{\alpha_2}, \theta) \geq [F(c_1,\theta)]^{\alpha_1}[F(c_2,\theta)]^{\alpha_2}$$

for any positive c_1, c_2, where F is the c.d.f. of χ^2_θ. Thus from the "arithmetic mean geometric mean" inequality we get

$$(4.3) \qquad F(\alpha_1 c_1 + \alpha_2 c_2, \theta) \geq [F(c_1,\theta)]^{\alpha_1}[F(c_2,\theta)]^{\alpha_2},$$

which shows that $F(c,\theta)$ is log-concave in c for $\theta > 0$. Incidently (4.2) also holds for $\bar{F}$ in place of F.

5. More Results on p.d.f.'s. Let X be a positive random variable with the p.d.f. $f(\cdot,\theta)$ with respect to Lebesgue measure.

Definition 4. The density $f(\cdot,\theta)$ is said to have the restricted reproductive property (RRP) in θ, if there exists a positive r.v. Y with the distribution $G(\cdot,\theta,\delta)$ such that

$$(5.1) \qquad \int_0^x f(x-y,\theta)\,G(dy,\theta,\delta) = f(x,\theta+\delta).$$

THEOREM 6. *Suppose the following conditions hold:* (a) $f(x,\theta)$ *is* TP_2, (b) f *satisfies the RRP, as given in (5.1),* (c) $G(\cdot,\theta,\delta)$, *as given in Definition 4, is stochastic decreasing in* θ. *Then* $\bar{F}(x,\theta)$ *is log-concave in* θ.

Proof. It follows from (a) that $\bar{F}(c,\theta)$ is TP_2. For $0 < c_1 < c_2$, $\theta_1 < \theta_2$ we have

$$(5.2) \qquad \bar{F}(c_2,\theta_1)\bar{F}(c_1,\theta_2) \leq \bar{F}(c_2,\theta_2)\bar{F}(c_l,\theta_1).$$

Write $c_2 = c_1 + y$. Integrating (5.2) with respect to $G(dy,\theta_2,\delta)$ we get

$$\begin{aligned}(5.3) \quad \bar{F}(c_2,\theta_1)\bar{F}(c_2,\theta_2+\delta) &\leq \bar{F}(c_2,\theta_2)\textstyle\int_0^\infty \bar{F}(c_2-y_1,\theta_1)\,G(dy,\theta_1,\delta)\\ &\leq F(c_2,\theta_2)\textstyle\int_0^\infty \bar{F}(c_2-y,\theta_1)\,G(dy,\theta_1,\delta)\\ &= \bar{F}(c_2,\theta_2)\bar{F}(c_2,\theta_1+\delta).\end{aligned}$$

6. Application to Beta Distribution. Suppose $U \sim \beta_{m,n}$ and $V \sim \beta_{\delta,m+n}$ are independently distributed. Then $UV \sim \beta_{m+\delta,n}$. Write $X = -\log U$, $Y = -\log V$, and $\theta = m$. Let $f(\cdot,\theta)$ be the density of X and $G(\cdot,\theta,\delta)$ be the c.d.f. of Y. Then the conditions (a)–(c) of Theorem 5 hold. Hence $P[U \leq c]$ is log-concave in m.

Remark. Some of the above results relating to chi-square distribution are given in the Ph.D. dissertation of Sarkar. Furthermore, following the ideas of Das Gupta and Perlman (1974), Sarkar (1982) has shown that $\chi^2_{m,\alpha}$ is log-concave in $m > 0$, where $P[\chi^2_m > \chi^2_{m,\alpha}] = \alpha$.

Remark. The only basic result relating the TP_2 properties and log-concavity available in the literature is the following (Karlin (1968)): A positive-valued function g is log-concave iff. $g(x-y)$ is TP_2 in (x,y). This result follows easily from the developments in Theorem 1; one has to consider a special G in Definition 2 which assigns the entire probability mass to a positive number.

Remark. The reproductive property (RP) stated in Definition 2 looks similar to the semigroup property introduced by Proschan and Sethuraman (1977). The semigroup property of a TP_2 function $f(\theta, x)$ is defined as follows:

$$f(\theta_1 + \theta_2, y) = \int f(\theta_1, x) f(\theta_2, y-x) d\nu(x),$$

In the above, (i) $\mathcal{X} = \mathcal{R}$, $\theta \in \Theta \subset \mathcal{R}$ is an interval, or (ii) $\mathcal{X} = \{\dots, -1, 0, 1, 2 \dots\}$. Θ is an interval or an interval of integers, and ν is some measure on $\mathcal{X}$. Proschan and Sethuraman (1977) have shown that

$$\psi(\theta_1, \dots, \theta_n) = \int \pi_{i=1}^n f(\theta_i, x_i) \phi(x_1 x_2, \dots, x_n) \pi_{i=1}^n d\mu(x_i)$$

is Schur-concex whenever ϕ is Schur-convex, μ being the Lebesgue measure for case (i) and the counting measure for case (ii). Some related results are given in the book by Marshall and Olkin (1979).

The RP is slightly more general than the above semigroup property, and its use, as illustrated in our main results, is also different.

Remark. It should be noted that $C(x,\theta,\lambda)$ in (1.17) is defined only for those (x,θ,λ) for which $f(x,\theta,\lambda) > 0$.

REFERENCES

Das Gupta, S. (1976). S-Unimodal function: Related inequalities and statistical applications. *Sankhya, Ser. B 38* 301–314.

Das Gupta, S. (1980). Brunn-Minkowski Inequalities and its aftermath. *J. Multivariate Anal. 10* 296–318.

Das Gupta, S. and Perlman, M. D. (1974). Power of the noncentral F-test: effect of additional variates on Hotelling's T^2-test. *J. Amer. Statist. Assoc. 69* 174–180.

Ghosh, B. K. (1973). Some monotonicity theorems for χ^2, F and t distributions with applications. *J. Royal Statist. Soc. B 35* 480–492.

Karlin, S. (1968). *Total Positivity*. Stanford University Press, Stanford, CA.

Marshall, A. W. and Olkin, I. (1979). *Inequalities: Theory of Majorization and Its Applications*. Academic Press, New York.

Proschan, F. and Sethuraman, J. (1977). Schur functions in statistics. I. The Preservation Theorem. *Ann. Statist. 5* 256–262.

Sarkar, S. K. (1984). Some results on χ^2 and a related distribution. To appear in *Sankhyā*.

Inequalities in Statistics and Probability
IMS Lecture Notes–Monograph Series Vol. 5 (1984), 59–67

A CLASS OF GENERALIZATIONS OF HÖLDER'S INEQUALITY

BY Marshall Freimer and Govind S. Mudholkar
University of Rochester

Let $a_1 \geqslant a_2 \geqslant \ldots \geqslant a_n \geqslant 0$, $b_1 \geqslant b_2 \geqslant \ldots \geqslant b_n \geqslant 0$ and consider the problem of maximizing $\Sigma_1^n a_i b_i$ subject to $\Sigma_1^m a_i^p = 1$, $m \leqslant n$. In this paper Kuhn-Tucker theory is used to solve the problem and consequently to obtain a generalization of Hölder's inequality. The reversal of the generalized inequality, its extension to the symmetric gauge functions and the continuous case are discussed. Some statistical applications and other work presently in progress are outlined.

1. Introduction and Summary. In an article published in 1889, O. Hölder presented two basic and now very well known results. The first of these is known as "Jensen's Inequality". In an addendum to his article J. L. W. V. Jensen (1906), who is credited with its discovery, acknowledges that the inequality is not "entirely new", that, after completing his work, through a monograph by A. Pringsheim he became aware of its earlier discovery by Hölder (1889). In the same 1906 paper, Jensen uses this Hölder-Jensen inequality for convex functions to derive in explicit form the second basic result only implicit in Hölder (1889), namely the "Hölder's inequality" bounding the inner products of vectors in terms of their norms. Specifically, if **a** and **b** are two vectors with nonnegative components a_1, $a_2, \ldots, a_n$ and $b_1, b_2, \ldots, b_n$ respectively, then Hölder's inequality asserts that

$$\Sigma_{i=1}^n a_i b_i \leqslant (\Sigma_1^n a_i^p)^{1/p} (\Sigma_1^n b_i^q)^{1/q}, \tag{1.1}$$

for any $p \geqslant 1$ and q satisfying $p^{-1} + q^{-1} = 1$. The inequality is reversed if $p < 1$, provided the components of **a** and **b** are strictly positive. Moreover, if these components are proportional, i.e. $a_i^p = cb_i^q$ for some c and $i = 1, 2, \ldots, n$ then in (1.1) and its reversal the equality holds. In this essay our interest centers on this classical inequality due to Hölder. Our objective is to present some recent generalizations of this inequality, to outline some statistical applications and to indicate the directions of further work which is in progess.

Although Hölder's inequality (1.1) was introduced as a theorem about the "mean values" it is now widely studied in its own right and is variously applied. In its better known applications in sciences, it is generally encountered as the particular case $p = 2$, i.e. the Cauchy-Schwarz inequality. In mathematics it appears in the theory of linear spaces in the context of identifying the conjugate or adjoint spaces and establishing their dual character. For discussions of various generalizations of (1.1) see Beckenbach and Bellman (1965), Hardy, Littlewood, and Polya (1952), Mitronović (1968) and Rockafellar (1970). The generalizations include sharp bounds on the sums of products of type $\Sigma_{i=1}^n a_i b_i c_i$ of the components of three or more vectors, and on integrals of type $\int a(x)\, b(x) dx$. Another approach to generalizing (1.1) is to use arbitrary norms $\phi(\mathbf{a}) = \phi(a_1, a_2, \ldots, a_n)$ leading to results of the type

$$\Sigma_{i=1}^n a_i b_i \leqslant \phi(\mathbf{a})\, \phi^\circ(b), \tag{1.2}$$

where

$$\phi^\circ(\mathbf{b}) = \max_{a \neq 0} \Sigma_{i=1}^n a_i b_i / \phi(\mathbf{a}), \tag{1.3}$$

AMS 1980 subject classifications. Primary, 47A30, 52A40; Secondary, 15A42.

Key words and phrases: Kuhn-Tucker theory, non-linear programming, Cauchy-Schwarz inequality, symmetric gauge functions, MANOVA multiple comparison, Cramér-Rao bounds.

is the polar of the norm ϕ. Clearly (1.2) is only a tautology unless more is known about either the polar ϕ° than (1.3), or about the inequality itself.

Let $a_{(1)} \geq a_{(2)} \geq \ldots \geq a_{(n)} > 0$ denote the ordered values of $a_1, a_2, \ldots, a_n$, $m \leq n$, $b_{[j]} = b_{(j)} + b_{(j+1)} + \ldots + b_{(n)}$, the tail sum of the smallest b's and $p \geq 1$. Then Mudholkar, Freimer, and Subbaiah (1983) consider the norm $\phi(\mathbf{a}) = (\Sigma_{i=1}^m a^p_{(i)})^{1/p}$ and show that

$$(1.4) \qquad \Sigma_{i=1}^n a_i b_i \leq \{\Sigma_{i=1}^n a^p_{(i)}\}^{1/p}\{\Sigma_{i=1}^n b^q_{(i)} + (m-k)(b_{[k+1]}/(m-k))^q\}^{1/q},$$

where $q^{-1} = 1 - p^{-1}$ and k is the integer given by Lemma 2.1. They also show that (1.4) is sharp and derive the reversal of (1.4) when $p < 1$. They prove the extension (1.4) of Hölder's inequality using arguments of convex analysis. In Section 2 we formulate an optimization problem and obtain (1.4) as its solution using a constructive method, namely the Kuhn-Tucker theory.

The polar ϕ° of an arbitrary norm ϕ defined by (1.3) can be described in alternative frameworks, e.g. geometrical, and may even be computed using numerical methods; but obviously there can be no "explicit formula" for it. Yet Hölder's inequality (1.1) can be generalized as (1.3) using general norms ϕ instead of the p-type norm. For a (symmetric) norm ϕ on R^m and $\mathbf{a} \in R^n$, $m \leq n$, let $\phi_m(\mathbf{a}) = \phi(a_{(1)}, a_{(2)}, \ldots, a_{(m)})$, where $a_{(1)} \geq a_{(2)} \geq \ldots \geq a_{(n)} \geq 0$ are the ordered values of the magnitudes, $|a_i|$, of the coordinates of $\mathbf{a}$. Then $\phi_m(\mathbf{a})$ defines a norm, derived by trimming from ϕ, on R^n. In Section 3 we obtain the polar ϕ°_m of the trimmed norm ϕ_m in terms of the polar ϕ° of ϕ. The result (1.4) is then obtained as a corollary of this construction.

Section 4 is given to the continuous case. Here we present a continuous version of the results in Section 2 and some of their implications. Section 5 is devoted to the outlines of some statistical applications which have been the main motivations for the generalized inequalities discussed in this paper. These include the multiple comparison procedures in statistical analysis and the variance bounds in the statistical estimation. We also present some new matrix inequalities which are relevant in such applications. The final section 6 contains miscellaneous remarks and indications of the further work presently in progress.

2. An Application of the Kuhn-Tucker Theory. Hölder's inequality (1889) which gives the maximum of an inner product may be regarded as the solution of an optimization problem. A general approach to obtaining the optimum of an objective function subject to constraints rests upon the Kuhn-Tucker conditions, a set of easily written down equations and inequalities which are both necessary and sufficient for the purpose. In practice these conditions are either solved to obtain the solution or used to verify the correctness of an otherwise obtained solution.

Given a vector $\mathbf{b} \in \mathcal{R}^n$ consider the problem of maximizing an objective function $\Sigma_{i=1}^n a_i b_i$ w.r.t. the a's subject to a constraint $(\Sigma_{i=1}^m a^p_{(i)})^{1/p} = 1$, where $m \leq n$ and $a_{(1)} \geq a_{(2)} \geq \ldots \geq a_{(n)} \geq 0$ denote the ordered values of the magnitudes $|a_i|$ of the coordinates of $\mathbf{a}$. Since the constraint involves only the magnitudes of the a's, and $\Sigma_{i=1}^n a_i b_i \leq \Sigma_{i=1}^n a_{(i)} b_{(i)}$ in view of the well known rearrangement theorem, see Hardy, Littlewood and Pólya (1952), we assume without any loss of generality that,

$$(2.1) \qquad a_1 \geq a_2 \geq \ldots \geq a_n \geq 0; b_1 \geq b_2 \geq \ldots \geq b_n \geq 0,$$

and for $p \geq 1$ consider the problem:

$$(2.2) \qquad \text{Maximize } \Sigma_{i=1}^n a_i b_i \text{ subject to } \Sigma_{i=1}^m a_i^p = 1.$$

Clearly the solution to (2.2) must satisfy $a_m = a_{m+1} = \ldots = a_n$. Thus the problem (2.2) is reduced to the nonlinear programming problem:

(2.3) $$\text{maximize } \Sigma_{i=1}^{m-1} a_i b_i + a_m \Sigma_{i=m}^{n} b_i \text{ subject to } \Sigma_{i=1}^{n} a_i^p = 1 \text{ and } a_1 \geq a_2 \geq \ldots \geq a_m \geq 0.$$

We still have $b_1 \geqslant b_2 \geqslant \ldots \geqslant b_{m-1} \geqslant 0$, but the coefficient of a_m is known only to be nonnegative. If it were zero the problem (2.3) would be trivial; a_i^p would be proportional to b_i^q, $q^{-1} = 1 - p^{-1}$, $i = 1,2, \ldots, m-1$ and a_m would be zero. Hence we assume that $\Sigma_{i=m}^{n} b_i > 0$.

In mathematical programming problems it is customary to use x's for the variables and express the problem in a standard format:

(2.4) $$\text{Minimize } f(x_1, x_2, \ldots, x_m) \text{ subject to } g_i(x_1, x_2, \ldots, x_m) \leqslant 0, i = 1,2, \ldots, s.$$

Then Lagrange multipliers are introduced to form the Lagrangian

(2.5) $$L(\mathbf{x}, \lambda) = f(x_1, x_2, \ldots, x_m) + \Sigma_{i=1}^{s} \lambda_i g_i(x_i, x_2, \ldots, x_m).$$

If f, g_1, g_2, $\ldots$, g_s are all convex then the solution to the problem is characterized by the Kuhn-Tucker conditions:

(2.6) $$\begin{gathered} \partial L / \partial x_i = 0, \quad i = 1,2, \ldots, m, \\ \lambda_i \geq 0 \quad \text{and } \lambda_i g_i(\mathbf{x}) = 0, \ i = 1,2, \ldots, s. \end{gathered}$$

In the standard format our problem (2.3) can be written down as:

(2.7) $$\text{Minimize } \Sigma_{i=1}^{m} (-c_j) x_j \text{ subject to } x_j - x_{j-1} \leqslant 0, j = 2,3, \ldots, m, -x_m \leqslant 0, \text{ and } \Sigma_{i=1}^{m} x_i^p - 1 \leqslant 0,$$

where $p \geqslant 1$, and $c_j = b_j$, $j = 1,2, \ldots, m-1$, $c_m = \Sigma_{i=m}^{n} b_i$ satisfy $c_1 \geqslant c_2 \geqslant \ldots \geqslant c_{m-1} \geq 0$ and $c_m > 0$. The Lagrangian may then be expressed as

(2.8) $$L(\mathbf{x}, \lambda, \mu) = \Sigma_{i=1}^{m} (-c_j) x_j + \Sigma_{i=1}^{m-1} \lambda_i (x_{i+1} - x_i) - \lambda_m x_m + \mu(\Sigma_{i=1}^{m} x_j^p - 1),$$

leading to

(2.9) $$0 = \partial L / \partial x_j = -c_j - \lambda_{j-1} + p\mu x_j^{p-1},$$

$j = 1,2, \ldots, m$, where $\lambda_0 = 0$. Solving (2.9) for the x_j we get

(2.10) $$x_j = ((c_j + \lambda_j - \lambda_{j-1})/p\mu)^{1/(p-1)}, j = 1,2, \ldots, m.$$

The multiplier μ is chosen to scale the x_i's so that $\Sigma_{i=1}^{m} x_i^p = 1$. This entails $\mu > 0$, and justifies the use of the inequality "$\leqslant$" instead of "$=$" in the constraint on $\Sigma_{i=1}^{m} x_i^p$ in (2.7).

We must now determine the λ's in (2.10) so that (2.6) and (2.7) hold. Suppose that for some integer k, $0 \leq k < m$, $\lambda_0 = \lambda_1 = \ldots = \lambda_k = 0$. Then $x_1 \geq x_2 \geq \ldots \geq x_k$ because the corresponding c_i's satisfy such inequalities. From (2.10), the remaining inequalities on the x_i's hold if

(2.11) $$c_k \geq c_{k+1} + \lambda_{k+1} = c_{k+2} + \lambda_{k+2} - \lambda_{k+1} = \ldots = c_m + \lambda_m - \lambda_{m-1} = D, \text{ say.}$$

Thus we have $(m-k)$ equations

(2.12) $$c_j + \lambda_j - \lambda_{j-1} = D, j = k+1, \ldots, m$$

Adding these we get

(2.11) $$\Sigma_{j=k+1}^{m} c_j + \lambda_m = (m-k)D,$$

which holds with $\lambda_m = 0$ provided

(2.14) $$D = \Sigma_{j=k+1}^{m} c_j / (m-k) = \Sigma_{j=k+1}^{n} b_j / (m-k).$$

Then $c_k \geqslant D$ requires that

(2.15) $$b_k \geqslant \Sigma_{j=k+1}^{n} b_j / (m-k),$$

and $D - c_{k+1} = \lambda_{k+1} \geqslant 0$ requires that

(2.16) $$b_{k+1} \leqslant \Sigma_{j=k+1}^{n} b_j/(m-k).$$

Finally we note that the c_j's are nonincreasing and that

(2.17) $$\lambda_{k+j+1} - \lambda_{k+j} = D - c_{k+j+1},$$

$j = 1,2, \ldots, (m-k-2)$. Hence $\lambda_{k+2}, \lambda_{k+3}, \ldots, \lambda_{m-1}$ are also nonnegative.

To complete the derivation of the solution, it is necessary to show the existence of k such that (2.15) and (2.16) hold. This is done in the following.

LEMMA 2.1. *If $b_1 \geqslant b_2 \geqslant \ldots \geqslant b_n \geqslant 0$ and m is an integer $1 \leqslant m \leqslant n$, then there exists a unique integer k, $0 \leq k < m$, such that $b_k > \Sigma_{j=k+1}^{n} b_j/(m-k)$ and $b_{k+1} \leq \Sigma_{j=k+1}^{n} b_j/(m-k)$, the first of the inequalities being inoperative if $k = 0$.*

Proof. For $r = 1,2, \ldots, m$ define

(2.18) $$\beta_r = (m-r)\, b_r - \Sigma_{j=r+1}^{n} b_j$$

Then it is sufficient to show existence of a unique k such that $\beta_k > 0 \geqslant \beta_{k+1}$. This is apparent from the facts that $\beta_r - \beta_{r+1} = (m-r)(b_r - b_{r+1}) \geqslant 0$, for $r = 1,2, \ldots, m-1$, and $\beta_m = -\Sigma_{j=m+1}^{n} b_j < 0$. □

Hence the solution to our optimization problem is given by

(2.19) $$\begin{aligned} x_j &= \{b_j/(p\mu)\}^{1/(p-1)}, \; j = 1,2, \ldots, k \\ &= \{D/(p\mu)\}^{1/(p-1)}, \; j = k+1, \ldots, m \end{aligned}$$

where D is as in (2.14) and, in virtue of the constraint $\Sigma_{j=1}^{m} x_j^p = 1$

(2.20) $$(p\mu)^{1/(p-1)} = \{\Sigma_{j=1}^{k} b_j^{p/(p-1)}) + (m-k) D^{p/(p-1)}\}^{1/p}.$$

The corresponding optimal value of the objective function is

(2.21) $$\Sigma_{j=1}^{m} c_j x_j = \{\Sigma_{j=1}^{k} b_j^q + (m-k)\, \bar{b}^q\}^{1/q},$$

where $q^{-1} = 1 - p^{-1}$ and $\bar{b} = \Sigma_{j=k+1}^{n} b_j/(m-k)$.

The findings of this section may be summarized as follows:

THEOREM 2.2. *Let $a_1 \geqslant a_2 \geqslant \ldots \geqslant a_n \geqslant 0$, $b_1 \geqslant b_2 \geqslant \ldots \geqslant b_n \geqslant 0$, $p \geqslant 1$ and $m \leqslant n$. Then we have the following sharp inequality:*

(2.22) $$\Sigma_{i=1}^{n} a_i b_i \leqslant (\Sigma_{i=1}^{m} a_i^p)^{1/p} \{\Sigma_{j=1}^{k} b_j^q + (m-k)\bar{b}^q\}^{1/p}.$$

where $q^{-1} = 1 - p^{-1}$, $\bar{b} = \Sigma_{j=k+1}^{n} b_j/(m-k)$, and k is as in Lemma 2.1

The Kuhn-Tucker approach of this section can also be used to establish the following reversal of (2.22).

THEOREM 2.3. *Let $0 < a_1 \leqslant a_2 \leqslant \ldots \leqslant a_n$, $b_1 \geqslant b_2 \geqslant \ldots \geqslant b_n > 0$, $p \leqslant 1$ and $m \leqslant n$. Then the inequality (2.22) is reversed, the analogous result being sharp.*

Particular Cases. Theorem 2.2 and Theorem 2.3 may be illustrated by taking special values of p and q.

(i) Take $p = 1$. Then for $a_1 \geqslant a_2 \geqslant \ldots \geqslant a_n > 0$, and $b_1 \geqslant b_2 \geqslant \ldots \geqslant b_n \geqslant 0$ we have

(2.23) $$\{\Sigma_{i=1}^{m} a_{n-i+1}\}\{\Sigma_{j=k+1}^{n} b_j/(m-k)\} \leqslant \Sigma_{i=1}^{n} a_i b_i \leqslant b_1 \Sigma_{j=1}^{m} a_i$$

if $k \geqslant 1$. If $k = 0$ then the lower bound on $\Sigma_{i=1}^{m} a_i b_i$ still holds, but the upper bound is replaced by

(2.24) $$\Sigma_{i=1}^{n} a_i b_i \leqslant (\Sigma_{i=1}^{m} a_i)(\Sigma_{i=1}^{n} b_i/m).$$

(ii) Now take limits as $p \to 0$. Then for $0 < a_1 \leqslant a_2 \leqslant \ldots \leqslant a_n$ and $b_1 \geqslant b_2 \geqslant \ldots \geqslant b_n > 0$,

$$\Sigma_{i=1}^{n} a_i b_i \geq m\{\Pi_{i=1}^{m} a_i\}^{1/m}\{(\Sigma_{i=k+1}^{n} b_j/(m-k))^{m-k}\Pi_{i=1}^{k} b_i\}^{1/m}. \tag{2.25}$$

(iii) Take $m=2$ and $p=2$. Then for $a_1 \geq a_2 \geq \dots \geq a_n \geq 0$ and $b_1 \geq b_2 \geq \dots \geq b_n \geq 0$ we get

$$\begin{aligned} \Sigma_{i=1}^{n} a_i b_i &\leq (a_1^2+a_2^2)^{1/2}\{b_1^2+(\Sigma_{i=2}^{n} b_i)^2\}^{1/2}, \text{ if } b_1 \geq \Sigma_{i=2}^{n} b_i \\ &\leq (a_1^2+a_2^2)^{1/2}\Sigma_{i=1}^{n} b_i/\sqrt{2}\ , \text{ if } b_1 < \Sigma_{i=2}^{n} b_i. \end{aligned} \tag{2.26}$$

3. The Polars of Trimmed Symmetric Gauge Functions. A symmetric gauge function (s.g.f.)ϕ is a real valued function such that (i) $\phi(\mathbf{x}) \geq 0$, and $\phi(\mathbf{x}) > 0$ if $\mathbf{x} \neq 0$, (ii) $\phi(\mathbf{x} + \mathbf{y}) \leq \phi(\mathbf{x}) + \phi(\mathbf{y})$, (iii) $\phi(c\mathbf{x}) = |c|\ \phi(\mathbf{x})$, c real, and (iv) $\phi(\epsilon_1 x_{i_1}, \epsilon_2 x_{i_2}, \dots, \epsilon_n x_{i_n}) = \phi(\mathbf{x})$ for any permutation $(i_1, i_2, \dots, i_n)$ of $(1, 2, \dots, n)$ and $\epsilon_i = \pm 1$, $i=1, 2 \dots, n$. In other words, an s.g.f. is a symmetric norm. Let Φ_n denote the class of s.g.f's on $\mathcal{R}^n$. For any $\phi \in \Phi_n$, $\phi^\circ(\mathbf{y}) = \sup_{\mathbf{x}\neq 0}\Sigma x_i y_i/\phi(\mathbf{x})$ is also an s.g.f., i.e. $\phi^\circ \in \Phi_n$. ϕ° is variously known as the conjugate, the associate or the polar of ϕ. $\phi(\mathbf{x}) = (\Sigma_{i=1}^{n} x_i^p)^{1/p}$, $\phi^\circ(\mathbf{y}) = (\Sigma_{i=1}^{n} y_i^q)^{1/q}$, $p^{-1} + q^{-1} = 1$, is the best known illustration of an s.g.f. and its polar.

The term s.g.f. was first used by J. von Neumann (1937) in the context of metrizing the spaces of matrices. He showed that the class of unitarily invariant norms of $(n \times n)$ complex matrices coincides with the class of s.g.f.'s of their singular values. His results have since been extensively generalized and utilized by other authors. The s.g.f.'s are used to define the norms for operators on Hilbert and Banach spaces and they play a crucial role in the study of function spaces and function algebras. For a general discussion, see Hewitt and Ross (1969).

For any $\mathbf{x} \in \mathcal{R}^n$ let $x_{(1)} \geq x_{(2)} \geq \dots x_{(n)} \geq 0$ denote the ordered values of the magnitudes $|x_i|$ of the coordinates of $\mathbf{x}$. Then for any $\phi \in \Phi_m$, $m \leq n$ and $\chi \in \mathcal{R}^n$ it can be shown that $\phi_m(\mathbf{x}) = \phi(x_{(1)}, x_{(2)}, \dots, x_{(n)})$ defines an s.g.f. on $\mathcal{R}^n$, i.e. $\phi_m \in \Phi_n$. By analogy with the "trimmed means" we call ϕ_m an s.g.f. derived by trimming, or simply a trimmed s.g.f. The following result proved in Mudholkar and Freimer (1983) describes the polar $\phi_m^\circ \in \Phi_n$ in terms of $\phi^\circ \in \Phi_m$.

THEOREM 3.1. *Let $\phi_m \in \phi_n$ be the trimmed s.g.f. on $\mathcal{R}^n$ derived from $\phi \in \phi_m$, $m \leq n$. Then the polar $\phi_m^\circ \in \Phi_n$ of Φ_m is given by*

$$\phi_m^\circ(\mathbf{y}) = \phi^\circ(y_{(1)}, y_{(2)}, \dots, y_{(k)}, \bar{y}, \bar{y}, \dots, \bar{y}), \tag{3.1}$$

where $\phi^\circ \in \phi_m$ is the polar of ϕ. $y_{(1)} \geq y_{(2)} \geq \dots \geq y_{(n)} \geq 0$ are the ordered values of the magnitudes $|y_i|$ of the coordinates of $\mathbf{y}$ *and $\bar{y} = \Sigma_{j=k+1}^{n} y_{(j)}/(m-k)$.*

The proof of Theorem 3.1 is based upon the symmetry and convexity properties of the s.g.f.'s. It is easy to see that Theorem 2.2 is a particular case of this theorem with $\phi(\mathbf{x}) = (\Sigma x_i^p)^{1/p}$.

4. The Continuous Case. This section contains a continuous analogue of the results in Section 2, i.e., an upper bound on $\int_0^N a(t)\ b(t)\ dt$, for $a \in L_p(0, N)$, $b \in L_q(0, N)$ with $p^{-1} + q^{-1} = 1, p, q \geq 1$. Parallel to the discrete case let $\bar{a}$, $\bar{b}$ be the nonincreasing re-arrangements of $|a|$, $|b|$, respectively, as discussed in Hardy, Littlewood, and Pólya (1952). Then

$$\int_0^N a(t)b(t)\,dt \leq \int_0^N |a(t)|\,|b(t)|\,dt \leq \int_0^N \bar{a}(t)\bar{b}(t)\,dt. \tag{4.1}$$

Hence with no loss of generality, we assume that $a(t)$ and $b(t)$ are nonincreasing nonnegative functions.

Now let $0 < M < N$. Then from Lemma 2.1 by taking limits, or otherwise, it can be shown that there exists a K, $0 \leq K < M$, such that

(4.2) $$\int_0^N a(t)b(t)\,dt \leq \{\int_0^M a(t)^p\,dt\}^{1/p}\{\int_0^M \hat{b}(t)^q dt\}^{1/q},$$

where

(4.3) $$\begin{aligned}\hat{b}(t) &= b(t), \quad 0 \leq t \leq K \\ &= \{\textstyle\int_K^N b(t)\,dt\}/(M-K), K \leq t \leq M.\end{aligned}$$

The defining equation for K is analogous to that in the discrete case, namely

(4.4) $$b(K) = \int_K^N b(t)dt/(M-K).$$

The existence of K may be seen directly by noting a couple of points. First, if $b(0) \leq 1/M \int_0^N b(t)\,dt$ then $K = 0$. Second, if the opposite inequality holds for $b(0)$, and we define the nonincreasing function $B(r) = (M-r)\,b(r) - \int_r^N b(t)\,dt$, for $0 \leq r \leq M$ then we have $B(0) \geq 0$ and $B(M) \leq 0$. Thus for continuous B there exists a K such that $B(K) = 0$; otherwise B would have a jump through 0. In this latter case K is defined by $\lim_{r\to K^-} B(r) \geq 0 \geq \lim_{r\to K^+} B(r)$.

The inequality (4.2) may be used to obtain simple inequalities such as

(4.5) $$\{\int_0^1 (1-t)\,g(t)\,dt\}^2 \leq \tfrac{1}{2}\int_0^{1/2} g^2(t)\,dt,$$

for any nonnegative nonincreasing g. Such inequalities can often be established more directly.

5. Applications. The main results of this paper were motivated by a problem in multivariate statistical analysis. This and some other applications are now outlined.

1. Multiple Comparisons Among Mean Vectors. First consider the classical ANOVA setup in canonical form. Let X_i be k independently normally distributed random variables with means θ_i, $i = 1,2, \ldots, k$ and common variance σ^2. Also let s^2 be an independently distributed estimate of σ^2. The ANOVA problem is to test

(5.1) $$H_o: \theta_1 = \theta_2 = \ldots = \theta_k,$$

and to identify the nature of departure from H_o in case of its rejection. Fisher's variance ratio F and Tukey's studentized range are the two best known tests of H_o. These two tests and the associated multiple comparisons can be obtained using S. N. Roy's union-intersection approach and the following modification of Hölder's inequality, (e.g. see Subbaiah and Mudholkar (1983)):

(5.2) $$\max_{\mathbf{c}'\mathbf{1}=0} \mathbf{c}'\,\mathbf{x}/\|\mathbf{c}\|_p = \min_{\eta} \|\mathbf{x}-\eta\mathbf{1}\|_q,$$

where $p \geq 1$, $p^{-1} + q^{-1} = 1$, $\|\mathbf{c}\|_p = (\Sigma_{i=1}^k |c_i|^p)^{1/p}$, $\theta' = (\theta_1, \theta_2, \ldots, \theta_k)$, and $\mathbf{1}' = (1, 1, \ldots, 1)$. Specifically by taking $p = 1$ and $p = 2$, respectively, we get

(5.3) $$|\mathbf{c}'\mathbf{x}| \leq s\Sigma|c_i|\{\max_{i,j}(x_i - x_j)/s\},$$

(5.4) $$\text{and} \quad |\mathbf{c}'\,\mathbf{x}| \leq s(\Sigma_{i=1}^k c_i^2)^{1/2}\,\{(\Sigma_{i=1}^k x_i^2)^{1/2}/s\},$$

for all $\mathbf{c}$ such that $\Sigma_{i=1}^k c_i = 0$. Replacing $\mathbf{X}$ by $(\mathbf{X}-\theta)$ in (5.3) and (5.4) we get, respectively, the T-method and S-method multiple comparisons, i.e. the simultaneous confidence intervals for all contrasts $\Sigma_{i=1}^k c_i\theta_i$, $\Sigma_{i=1}^k c_i = 0$, given by the F-test and the studentized range test.

The multivariate ANOVA, i.e. MANOVA, hypothesis in canonical form is H_0: $\Theta = 0$ where Θ is a $(p \times k)$ matrix of the mean-vector of k p-variate normal populations with a common covariance matrix Σ. The invariant tests, see Lehmann (1959), of H_o depend upon the eigenvalues $\lambda_1 \geq \lambda_2 \geq \ldots \geq \lambda_p$ of $S_H S_E^{-1}$, where S_H and S_E are the matrices of the sums of squares and products due to the hypothesis and errors respectively. A class of such statistics especially suited to multiple comparisons introduced by Muldholkar (1965, 1966), see also Mudholkar, Davidson, and Subbaiah (1974), and Wijsman (1980), are the unitarily invariant norms $\|\Theta\ S_E^{1/2}\|_\phi = \phi(\lambda_1^{1/2}, \lambda_2^{1/2}, \ldots, \lambda_p^{1/2})$ generated by the s.g.f.'s $\phi \in \Phi_p$. The largest root statistic λ_1 due to S. N. Roy and Hotelling's trace criterinon $\Sigma_{i=1}^p \lambda_i$ which belong to this class are analogous to the univariate studentized range and the F statistics, in that the former yield shorter confidence intervals whereas the latter have superior overall power. This suggests trimmed s.g.f.'s $\phi(\lambda_1, \lambda_2, \ldots, \lambda_m) = \phi_m(\lambda)$, $m < p$, $\phi \in \Phi_m$, ϕ_m, $\in \Phi_p$, as the compromise statistic which would capture most of the noncentrality in the problem without serious sacrifice in the shortness of the confidence intervals.

Now the construction of simultaneous confidence intervals in the MANOVA setting using the s.g.f. statistics $\phi_m(\lambda)$ rests upon inequalities of the form

(5.5) $$tr(\mathbf{AB}) \leq \|\mathbf{A}\|_{\phi_m} \|\mathbf{B}\|_{\phi_m^\circ},$$

which are analogous to the Hölder's inequality. This takes us to the second application.

2. *Some Matrix Inequalities.* The inequalities involving matrix functions such as singular values, eigenvalues, traces, determinants, etc. are of broader interest than the multiple comparisons discussed above, e.g. see Beckenbach and Bellman (1971), Marshall and Olkin (1979) or Mitrinović (1970). The following two results, which bound the trace functions in terms of sums resembling inner products, may be found in Marshall and Olkin (1979, ch. 20).

THEOREM 5.1. (von Neumann, 1937). *If* $\mathbf{A}$, $\mathbf{B}$ *are* $(n \times n)$ *complex matrices, and* $\mathbf{U}, \mathbf{V}$ *are unitary then*

(5.6) $$Re(tr\,\mathbf{UAVB}) \leq |tr\,(\mathbf{UAVB})| \leq \Sigma_{i=1}^n \sigma_i(\mathbf{A})\sigma_i(\mathbf{B}),$$

where $\sigma_i(\mathbf{A})$, $\sigma_i(\mathbf{B})$ *are the singular values of* $\mathbf{A}$ *and* $\mathbf{B}$ *arranged in decreasing order,* $i = 1, 2, \ldots, n$.

THEOREM 5.2. *Let* $\mathbf{H}$ (nxn) *be a Hermitian matrix with eigenvalues* $\lambda_1 \geq \lambda_2 \geq \ldots \geq \lambda_n$ *and* $\mathbf{U}$ (kxn) *be a complex matrix such that the eigenvalues of* $\mathbf{UU^*}$ *are* $\beta_1 \geq \beta_2 \geq \ldots \geq \beta_k \geq 0$. *Then for all* $k = 1, 2, \ldots, n$

(5.7) $$\Sigma_{i=1}^k \lambda_{n-i+1}\, \beta_i \leq tr\,\mathbf{UHU^*} \leq \Sigma_{i=1}^k \lambda_i \beta_i$$

Clearly application of Theorem 2.2 to (5.6) and of Theorem 2.2 and 2.3 to (5.7) result in numerous inequalities involving $Retr(\mathbf{UAVB})$ and $tr(\mathbf{UHU^*})$.

3. *Cramér-Rao Information Inequality.* As illustrated in Section 2, Theorem 2.2 is a generalization of the well known Cauchy-Schwartz inequality. Hence it is potentially useful in establishing extensions of results generated using the Cauchy-Schwartz bound. One such basic result in statistical inference is the lower bound on the variance of an estimator due to H. Cramér and C. R. Rao, see e.g. Rao (1973).

Let $X_1, X_2, \ldots, X_N$ be a random sample from a population with probability density function $f(x;\theta)$ depending on a real valued parameter θ. Then $V = \mathrm{Var}(T)$ of an estimator T such that $E(T) = \theta + b(\theta)$ satisfies

(5.8) $$V \geq (1 + b')/NI,$$

where $b' = \partial/\partial\theta(b\theta)$, and $I = I(\theta) = E(\partial/\partial\theta \log f(x;\theta))^2$ is the information per observation in the sample.

The result (5.8) can be extended in several directions by applying the inequalities of this paper. As a simple example, consider n such problems with analogous quantities $N_j, f_j(x;\theta)$, V_j, $b_j(\theta)$ and $I_j(\theta)$, $j = 1, 2, \ldots, n$. Then from (5.8) we get

$$\Sigma_{j=1}^{n}(1 + b_j') \leq \Sigma_{j=1}^{n}(N_jI_j)V_j. \tag{5.9}$$

If we apply Theorem 2.2 to the right hand side of (5.9) then we obtain tight lower bounds on risk functions of type $\Sigma_{j=1}^{m} V_{(j)}$, $m < n$, the sum of the m largest variances. These lower bounds can be used to identify good common estimators for parameters of the same type, for example location, for different distributions.

6. Remarks.

1. Nonlinear programming, which is now a well developed field, provides a new constructive approach for generating inequalities. Kuhn-Tucker theory, and Lagrangian duality, are the two underpinnings of this subject. Bazaraa and Shetty (1979) Chapters 4 and 6 provide an excellent summary of these topics. Pourciau's (1980) essay entitled "Modern Multiplier Rules" is a nice expository survey.

2. In this paper we have focused upon inequalities involving convex functions and their multiplicative duals called polars. If f is a real valued convex function on $\mathcal{R}^n$ then $f^c(\mathbf{y}) = \sup_{\mathbf{x}} [\mathbf{y}'\mathbf{x} - f(\mathbf{x})]$, known as the Fenchel conjugate of f, yields inequality $\mathbf{y}'\mathbf{x} \leq f(\mathbf{x}) + f^c(\mathbf{y})$. Analogues of the result in Section 3 for Fenchel conjugates exist.

3. In Section 3 we deal with s.g.f.'s, the symmetric homogeneous norms, which include the p-norms $p \geq 1$. It is possible to develop the analogue of the reversed inequality given in Theorem 2.3 in the general setup using concave functions.

4. The work on the results of Section 2 for infinite sequences is in progress.

5. Section 4 gives the continuous version of results in Section 2. Investigation of the integrals of functions defined on the entire real line and the continuous version of the result in Section 3 is also continuing.

6. It is well known that the von Neumann norms based upon the s.g.f.'s play a crucial role in the theory of function spaces. The analysis of the normed linear spaces using the trimmed norms is likely to be interesting.

REFERENCES

Bazaraa, M. S. and Shetty, C. M. (1979). *Nonlinear Programming Theory and Algorithms*, Wiley, New York.

Beckenbach, E. F. and Bellman, R. (1971). *Inequalities*, Springer-Verlag, New York.

Hardy, G. H., Littlewood, J. E., and Pólya, G. (1952). *Inequalities*, 2nd ed. Cambridge University Press, London and New York.

Hewitt, E. and Ross, K. A. (1969). *Abstract Harmonic Analysis, Vol. 2*, Springer-Verlag, New York.

Hölder, O. (1889). Uber einen Mittelwertsatz. *Gottener Nachr.*, 38–47.

Jensen, J. L. W. V. (1906). Sur les fonctions convexes et les inegalities entre les valeurs moyennes. *Acta Math. 30* 175–193.

Marshall, A. W. and Olkin, I. (1979). *Inequalities: Theory of Majorization and Its Applications*. Academic Press, New York.

Mitrinović, D. S. (1970). *Analytic Inequalities*. Springer-Verlag, New York.

Mudholkar, G. S. (1965). A class of tests with monotone power functions for two problems in multivariate statistical analysis. *Ann. Math. Statist. 36* 1794–1801.

MUDHOLKAR, G. S. (1966). On confidence bounds associated with multivariate analysis of variance and non-independence between two sets of variates. *Ann. Math. Statist.* *37* 1736–1746.

MUDHOLKAR, G. S., DAVIDSON, M. L., and SUBBAIAH, P. (1974). A note on the union-intersection character of some MANOVA procedures. *J. Multivariate Anal.* *4* 486–493.

MUDHOLKAR, G. S., FREIMER, M., and SUBBAIAH, P. (1984). An extension of Hölder's inequality. To appear in *J. Math. Analy. Appl.*

MUDHOLKAR, G. S., and FREIMER, M. (1984). A structure theorem for the polars of unitarily invariant norms. Submitted for publication.

POURCIAU, B. H. (1980). Modern multiplier rules. *Amer. Math. Monthly, 87* 433–452.

RAO, C. R. (1973). *Linear Statistical Inference and Its Applications,* 2nd ed., Wiley, New York.

ROCKAFELLAR, R. T. (1970). *Convex Analysis.* Princeton University Press, Princeton, N.J.

SUBBAIAH, P. and MUDHOLKAR, G. S. (1983). The union-intersection principle and the simultaneous test procedures for means. To appear in *Applied Statistics*.

VON NEUMANN, J. (1937). Some matrix inequalities and metrization of matric spaces. *Tomsk. Univ. Rev.* *1* 286–300 (also the collected works Vol. IV).

WIJSMAN, R. A. (1979). Constructing all smallest confidence sets in a given class, with Applications to MANOVA. *Ann. Statist.* *7* 1003–1018.

Inequalities in Statistics and Probability
IMS Lecture Notes–Monograph Series Vol. 5 (1984), 68–77

CONVEXITY PROPERTIES OF ENTROPY FUNCTIONS AND ANALYSIS OF DIVERSITY

BY C. RADHAKRISHNA RAO
University of Pittsburgh

Some natural conditions which a diversity measure (variability) of a probability distribution should satisfy imply that it must have certain convexity properties, considered as a functional on the space of probability distributions. It is shown that some of the well known entropy functions, which are used as diversity measures do not have all the desirable properties and are, therefore, of limited use. A new measure called the quadratic entropy has been introduced, which seems to be well suited for studying diversity.

Methods for apportioning diversity (APDIV) at various levels of a hierarchically classified set of populations are described. The concept of analysis of diversity (ANODIV), as a generalization of ANOVA, applicable to observations of any type, is developed and its use in the analysis of cross classified data is demonstrated. The choice of a suitable measure of diversity for the above purpose is discussed.

1. Introduction. There is an extensive literature on the measurement and analysis of diversity. A unified approach to these problems is given in Rao (1982a), and a complete bibliography of papers on this subject is complied by Denis, Patil, Rossi and Taille (1979). The choice of a diversity (DIV) measure for the analysis of given data poses a serious problem. An attempt is made in this paper to lay down some natural conditions which a diversity measure should satisfy (Section 2) and discuss the methodology for data analysis through an appropriate diversity measure. Some of the situations where such an analysis is needed are as follows.

Geneticists are interested in comparing populations by the diversity exhibited in certain measurements (Karlin, Kennett and Bonne-Tamir (1979)), and in apportioning diversity (APDIV) in a substructured population as due to between and within groups (Lewontin (1972), Nei (1973), Chakraborty (1974), Rao (1982a) and Rao and Boudreau (1982)).

In analysis of variance (ANOVA) of quantitative data, we choose the *variance* as a measure of diversity and partition it into a number of additive components. Of particular practical interest is the analysis of data classified by the levels of a number of factors, where the total variability is partitioned as due to main effects and interactions of factors. A natural question arises as to whether other measures of diversity such as *mean absolute deviation* could be used for this purpose. Further, what is the natural extension of ANOVA to observations which are not quantitative in nature?

In this paper, the concept of ANOVA is extended to more general analysis of diversity (ANODIV) applicable to observations belonging to *any sample space* by an appropriate choice of a diversity measure satisfying some convexity properties.

The choice of the well known entropy functions due to Shannon (1948), Havrda and Charvát (1967) and Rényi (1961) as diversity measures have only limited use as they do

AMS 1980 subject classifications, 94A17, 62J10.
Key words and phrases: Analysis of diversity, Complete convexity, Diversity measure, Entropy, Jensen difference, Rao's Quadratic entropy.

not have strong convexity properties (Section 3). A new measure called *quadratic entropy* which is *completely convex* (Section 4) is introduced and shown to be a useful tool in APDIV (Section 5) and ANODIV (Sections 6 and 7).

2. Measures of Diversity. Consider a measurable space $(\mathcal{X},\mathcal{B})$ and a convex set $\mathcal{P}$ of probability measures defined on it. A function H mapping $\mathcal{P}$ into the real line R is said to be a measure of diversity if

$$C_0\colon H(P) \geqslant 0 \quad \text{for every } P \in \mathcal{P} \text{ and } H(P) = 0 \text{ iff } P \text{ is degenerate.} \tag{2.1}$$

The condition C_0 is a natural one since a measure of diversity should preferably be non-negative and take the value zero only when all the individuals of a population are identical.

Consider two measures P_1 and P_2 and a mixture $\lambda_1 P_1+\lambda_2 P_2$, $(\lambda_1 \geqslant 0, \lambda_2 \geqslant 0, \lambda_1+\lambda_2=1)$. It is again a natural requirement that the amount of diversity in a mixture of populations should not be smaller than the average of the diversities within the individual populations. We may formulate this requirement (diversity is possibly increased by mixing populations) as

$$C_1\colon H(\lambda_1 P_1+\lambda_2 P_2) \geqslant [\lambda_1 H(P_1)+\lambda_2 H(P_2)] \tag{2.2}$$

with $>$ sign if $P_1 \neq P_2$. The condition C_1 is equivalent to saying that H is a strictly concave function or $-H$ is a strictly convex function.

Let us denote $J^{(0)} = -H$ and define

$$J^{(1)}(P_1,P_2\colon\lambda_1,\lambda_2) = \lambda_1 J^{(0)}(P_1) + \lambda_2 J^{(0)}(P_2) - J^{(0)}(\lambda_1 P_1 + \lambda_2 P_2) \tag{2.3}$$

as the first *Jensen difference* between P_1 and P_2. From the Condition C_1, the first Jensen difference $J^{(1)}$ is positive if the measures P_1 and P_2 are different, and hence may be considered as a measure of dissimilarity (distance) between P_1 and P_2. Now consider pairs of measures (P_{11},P_{12}), (P_{21},P_{22}) and the mixture $(\mu_1 P_{11}+\mu_2 P_{21}, \mu_1 P_{12}+\mu_2 P_{22})$ all belonging to $\mathcal{P}^2 = \mathcal{P} \times \mathcal{P}$. It is a natural requirement that the distance between the two mixed populations $\mu_1 P_{11}+\mu_2 P_{21}$ and $\mu_1 P_{12}+\mu_2 P_{22}$ should not be larger than the average of the distances between P_{11} and P_{12} and between P_{21} and P_{22}. This requirement (dissimilarity is possibly decreased by mixing) leads to

$$\begin{aligned} C_2\colon & \mu_1 J^{(1)}(P_{11},P_{12}\colon \lambda_1,\lambda_2)+\mu_2 J^{(1)}(P_{21},P_{22}\colon \lambda_1,\lambda_2) \\ & -J^{(1)}(\mu_1 P_{11}+\mu_2 P_{21}, \mu_1 P_{12}+\mu_2 P_{22}\colon \lambda_1,\lambda_2) \geqslant 0 \end{aligned} \tag{2.4}$$

where the left side of (2.4) is defined on $\mathcal{P}^4=\mathcal{P}^2\times\mathcal{P}^2$. We denote this expression by

$$J^{(2)}(\{P_{ij}\}\colon \{\lambda_i \mu_j\}) \tag{2.5}$$

observing that it can be alternatively written as

$$\begin{aligned} & \lambda_1 J^{(1)}(P_{11},P_{21}\colon \mu_1,\mu_2)+\lambda_2 J^{(1)}(P_{12},P_{22}\colon \mu_1,\mu_2) \\ & -J^{(1)}(\lambda_1 P_{11}+\lambda_2 P_{12}, \lambda_1 P_{21} + \lambda_2 P_{22}\colon \mu_1,\mu_2) \end{aligned} \tag{2.6}$$

exhibiting row and column symmetry. The condition C_2 means that the first Jensen difference $J^{(1)}$ is a convex function on $\mathcal{P}^2$.

Generalizing the above concepts, the need for which is demonstrated in Section 6, we lay down a series of conditions

$$C_i\colon J^{(i)} \geqslant 0, i = 1,2, \ldots \tag{2.7}$$

where $J^{(i)}$ is defined on $\mathcal{P}^{2^i}$ in a recursive way.

We use the following definitions:

Definition 2.1. A diversity measure H satisfying C_0 is said to be completely convex *if* $J^{(0)} = -H_1, J^{(1)}, \ldots$ *defined on appropriate spaces are all convex, or the conditions* C_i *are satisfied for all* i.

Definition 2.2. A diversity measure is said to be j-th order convex if the conditions $C_0, C_1, \ldots, C_j$ *are satisfied.*

3. Entropy as a Diversity Measure. A number of diversity measures have been introduced through the concept of entropy and information and applied in different areas of research. When $\mathcal{P}$ is the simplex of all multinomial distributions in k cells

$$\mathcal{P} = \{p = (p_1, \ldots, p_k)', p_i \geq 0, \Sigma p_i = 1\}, \tag{3.1}$$

some of the well known measures of entropy are

$$H_S(p) = -\Sigma p_i \log p_i, \quad \text{(Shannon (1948))}, \tag{3.2}$$

$$H_\alpha(p) = \frac{1 - \Sigma p_i^\alpha}{\alpha - 1}, \alpha > 0, \alpha \neq 1, \quad \text{(Havrda and Charvát (1967))}, \tag{3.3}$$

$$H_R(p) = \frac{\log \Sigma p_i^\alpha}{1 - \alpha}, \alpha > 0, \alpha \neq 1, \text{(Rényi (1961))}. \tag{3.4}$$

All these measures are non-negative and take the value zero only when one of the p_i is unity and the rest are zero. They all attain the maximum when $p_i = 1/k$ for every i. Thus they satisfy the condition C_0.

It is easy to verify that H_S and H_α for any $\alpha > 0$ satisfy the concavity condition C_1, while H_R satisfies C_1 only for $0 < \alpha < 1$.

Burbea and Rao (1982a, 1982b) have shown: (i) H_S satisfies C_2 but not $C_3, C_4, \ldots$, and (ii) H_α satisfies C_2 for $1 \leq \alpha \leq 2$ when $k > 2$ and for $\alpha \in [1,2] \bigcup [3, 11/3]$ when $k = 2$ and does not satisfy $C_3, C_4, \ldots$ except when $\alpha = 2$. It is not known whether H_R satisfies C_2 or not.

It may be noted that the continuous analogues of (3.2)–(3.4) are not necessarily nonnegative functionals and their interpretation as diversity measures poses some difficulties.

4. Quadratic Entropy. Let us consider a measurable space $(\mathcal{X}, \mathcal{B})$ and a function d defined on $\mathcal{X}^2$ such that

$$\begin{aligned} d(X_1, X_2) = d(X_2, X_1) &> 0 \quad \text{if } X_1 \neq X_2 \\ &= 0 \quad \text{if } X_1 = X_2. \end{aligned} \tag{4.1}$$

Using such a nonnegative function (kernel), we may define the diversity of a probability measure P defined on $(\mathcal{X}, \mathcal{B})$ by

$$H(P) = \int d(X_1, X_2) P(dX_1) P(dX_2). \tag{4.2}$$

The motivation for such a definition in a biological context was given in Rao (1982a). The expression (4.2) is the average difference (as defined by 4.1) between two individuals (observations) drawn at random from the population specified by the probability measure P.

The functional (4.2) satisfies the condition C_0 for a diversity measure. The condition C_1 requires

$$J^{(1)}(P_1, P_2 \colon \lambda_1, \lambda_2) = 2\lambda_1\lambda_2[H(P_1, P_2) - \tfrac{1}{2}H(P_1) - \tfrac{1}{2}H(P_2)] \geq 0 \tag{4.3}$$

where $H(P_1, P_2)$ is the average difference between two randomly drawn individuals, one from P_1 and another from P_2. The concavity of H or nonnegativity of $J^{(1)}$ reflects the intuitive requirement that two individuals drawn from different populations are on the average more different than those coming from the same population. The expression

$$D(P_1, P_2) = H(P_1, P_2) - \tfrac{1}{2}[H(P_1) + H(P_2)] \tag{4.4}$$

which provides the excess variability, represents the amount of dissimilarity between the populations defined by P_1 and P_2.

Not all nonnegative kernels, $d(x_1,x_2)$, lead to concave diversity measures $H(P)$ as defined in (4.2). We shall investigate the conditions under which this happens. For this purpose we introduce the concept of a conditionally negative definite (C.N.D.) function studied by Schoenberg (1938), Wells and Williams (1975) and Parthasarathy and Schmidt (1972).

Definition. *A nonnegative function* $d(\cdot,\cdot)$ on $\mathcal{X}^2$ *is said to be conditionally negative definite (C.N.D.) if*

$$\Sigma_1^n\Sigma_1^n d(X_i,X_j)a_ia_j \leq 0 \tag{4.5}$$

for any n and choices of real numbers $a_1, \ldots, a_n$ *such that* $\Sigma\, a_i = 0$. *If* (4.5) *is satisfied for all* $a_1, \ldots, a_n$, *then d is said to be a negative definite function.*

The following is a well known theorem of Schoenberg (1938).

THEOREM 4.1. *Let* $d(\cdot,\cdot)$ *be a nonnegative symmetric function on* $\mathcal{X}^2$. *Then* (i) *d is C.N.D. iff* $\rho = d^{1/2}$ *is a metric (i.e., satisfies the postulates of a distance function), and there is an isometry which embeds the metric space* $(\mathcal{X},\rho)$ *into a Hilbert space.* (ii) *If d is C.N.D., then* d^β *is also C.N.D. for* $0\leq\beta\leq1$. (iii) *If* $\mathcal{X}$ *is a set of not more than four points, then for any metric d,* $(\mathcal{X}, d^\beta)$, $0\leq\beta\leq\frac{1}{2}$, *can be isometrically embedded into the Euclidean space of dimension* 3.

The following theorem is given in Parthasarathy and Schmidt (1972).

THEOREM 4.2. *Let* $d(\cdot,\cdot)$ *be a nonnegative symmetric function on* $\mathcal{X}^2$. *Then the following conditions are equivalent:* (i) *d is C.N.D.* (ii) *For any fixed* $X_0\in\mathcal{X}$, *the kernel defined by*

$$d_0(X_1,X_2)=d(X_1,X_2)-d(X_1,X_0)-d(X_0,X_2)+d(X_0,X_0) \text{ for every } X_1,X_2$$

is negative definite. (iii) *For every* $t>0$, e^{-td} *is positive definite.*

The following theorem gives the conditions under which the diversity measure (4.2) is concave.

THEOREM 4.3. *The diversity measure H defined in (4.2) is concave if d is a C.N.D. function.*

The result follows by observing (see Rao (1982a, 1982b)) that

$$\begin{aligned}&J^{(1)}(P_1,P_2{:}\,\lambda_1,\lambda_2)\\&= -\lambda_1\lambda_2\int d(X_1,X_2)[P_1(dX_1)-P_2(dX_1)][P_1(dX_2)-P_2(dX_2)].\end{aligned}$$

Theorem 4.3 shows that when d is chosen as a conditionally negative definite function, the condition C_1 for the diversity measure H defined in (4.2) is satisfied. What further property should d satisfy in order that $C_2,C_3, \ldots$ hold? Fortunately, no further condition seems to be necessary as demonstrated in the following theorem (see Rao (1982a)).

THEOREM 4.4. *If d is C.N.D., then* $-H$, *where H is as defined in* (4.2), *is completely convex, i.e., H satisfies all the conditions* $C_1,C_2, \ldots$.

Consider two different pairs of probability measures $P_{11},P_{12}{:}P_{21},P_{22}$. Then $J^{(2)}$ as defined in (2.4) is seen to be, apart from a constant,

$$= -\int d(X_1,X_2)P_{(1-2)(1-2)}(dX_1)P_{(1-2)(1-2)}(dX_2)$$

where $P_{(1-2)(1-2)} = P_{11}+P_{22}-P_{12}-P_{21}$. Then $J^{(2)}\geq0$, since the total measure of $P_{(1-2)(1-2)}$ is zero and d is C.N.D. Similarly, all higher order Jensen differences are convex, which proves the theorem.

A diversity measure H defined as in (4.2) and completely convex in the sense of Theorem 4.4 is called *quadratic entropy*.

The quadratic entropy seems to have better properties as a measure of diversity than the traditional entropy measures considered in Section 3. Now, we raise the question as to whether a completely convex diversity measure is a quadratic entropy. The following theorem due to Lau (1982) provides the answer in the affirmative, and thus gives a characterization of the quadratic entropy of Rao (1982a).

THEOREM 4.5. *Let $\mathcal{X}$ be a normed topological space and $\mathcal{P}$ be the space of probability measures on $(\mathcal{X},\mathcal{B})$ with weak topology. Let $H: P \to R_+$ be a continuous function such that* (i) $H(P)=0$ *if P is degenerate, and* (ii) $-H$ *is completely convex. Then there exists a unique C.N.D. function d on $\mathcal{X}^2$ such that*

$$H(P) = \int d(X_1,X_2)P(dX_1)P(dX_2) \tag{4.6}$$

i.e., H is a quadratic entropy.

Note 1. For example, if $\mathcal{X} = R^1$, then

$$d(X_1,X_2) = (X_1-X_2)^2 \text{ for every } X_1,X_2, \in R^1 \tag{4.7}$$

is a C.N.D. function, and the diversity measure associated with it

$$\int (X_1-X_2)^2 P(dX_1)P(dX_2) = 2\sigma_p^2 \tag{4.8}$$

is the variance functional of P. It is well known that the variance functional is completely convex.

Note 2. It follows from Theorem 4.1, result (ii), that if $\mathcal{X} = R^1$,

$$H(P) = \int |X_1-X_2|^\beta P(dX_1)P(dX_2) \tag{4.9}$$

is completely convex for $0 \leq \beta \leq 2$, and in particular the *city block distance* functional

$$\int |X_1-X_2| P(dX_1)P(dX_2) \tag{4.10}$$

is completely convex.

Note 3. If $\mathcal{X}$ is a space of not more than four points, then the quadratic entropy based on any function d which is a metric on $\mathcal{X}$ is completely convex. [Note that this may not be true when $\mathcal{X}$ has more than four points.]

Note 4. Let $\mathcal{X}$ be a space of n points and

$$d(X_1,X_2) = 1, \quad \text{if } X_1 \neq X_2 \quad \text{and} = 0, \text{ if } X_1 = X_2. \tag{4.11}$$

Then the diversity measure (4.2) based on (4.11) is completely convex.

5. Apportionment of Diversity (APDIV). Biologists are interested in apportioning the total diversity in a population as due to differences between and within subpopulations. A concave diversity measure H is ideally suited for this purpose. If $P_1, \ldots, P_k$ are probability distributions in k subpopulations with prior probabilities $\lambda_1, \ldots, \lambda_k$, then we have the decomposition of the total diversity (T).

$$\begin{aligned} H(\Sigma \lambda_i P_i) &= \Sigma \lambda_i H(P_i) + J^{(1)}(\{P_i\}:\{\lambda_i\}) \\ T \quad &= \quad W \quad + \quad B \end{aligned} \tag{5.1}$$

where the components W and B are nonnegative. In (5.1), W is the average diversity within subpopulations and B may be interpreted as the diversity between the subpopulations. The ratio

$$G = \frac{B}{B+W} \tag{5.2}$$

called the index of diversity between subpopulations compared to the total has been used in genetic studies (see for instance Lewontin (1972), Nei (1973) and Chakraborty (1974)). Different diversity measures give different values of G, which raises the problem of choosing an appropriate measure in practical applications. For a discussion of this problem and some illustrative examples, the reader is referred to Rao (1982a) and Rao and Boudreau (1982). It is seen that if the object is APDIV (apportionment of diversity), we need only a concave diversity measure, i.e., one which satisfies only the condition C_1.

More generally, let us consider a number of populations grouped in a hierarchical classification such as populations within regions and regions within species and so on. If the distributions within populations and their apriori probabilities are known, then the distributions of groups at any level of classification and the associated apriori probabilities can be computed. This would enable us to compute the average diversity within groups at any level of classification. Then we have the apportionment of the total diversity H_0 (for all populations mixed together) as is shown in Table 1.

TABLE 1. APDIV for a hierarchical classification

due to	*diversity*	*ratio*
within populations	H_P	
between populations (within regions)	$H_R - H_P$	$(H_R - H_P)/H_R$
within regions	H_R	
between regions (within species)	$H_S - H_R$	$(H_S - H_R)/H_S$
within species	H_S	
between species	$H_0 - H_S$	$(H_0 - H_S)/H_0$
Total	H_0	

The only property required of a diversity measure for APDIV is concavity.

Note 1. It is interesting to note that if we use a quadratic entropy for APDIV, the decomposition (5.1) can be written as

$$H(\Sigma\,\lambda_i P_i) = \Sigma\,\lambda_i H(P_i) + \Sigma\Sigma\,\lambda_i\lambda_j D_{ij} \tag{5.3}$$

where

$$\begin{aligned} D_{ij} &= 2H\Big(\frac{P_i+P_j}{2}\Big) - H(P_i) - H(P_j) \\ &= H(P_i,P_j) - \tfrac{1}{2}H(P_i) - \tfrac{1}{2}H(P_j) \end{aligned}$$

is the dissimilarity between the populations i and j. The second term on the right hand side of (5.3) is the average dissimilarity between populations. Such an interpretation is available only if a quadratic entropy is used.

Note 2. Let $P_1 = N(\mu_1,\sigma^2)$ and $P_2 = N(\mu_2,\sigma^2)$. If we use the variance functional (4.8), then $H(P_1) = 2\sigma^2 = H(P_2)$ and $H(P_1,P_2) = (\mu_1-\mu_2)^2+2\,\sigma^2$ so that the dissimilarity between P_1 and P_2 is

$$D_{12} = H(P_1,P_2) - \tfrac{1}{2}H(P_1) - \tfrac{1}{2}H(P_2) = (\mu_1-\mu_2)^2 = \delta^2. \tag{5.4}$$

On the other hand, if we use the city block functional (4.10)

$$H(P_1) = 2\sigma/\sqrt{\pi} = H(P_2)$$

$$H(P_1,P_2) = (2\sigma/\sqrt{\pi})e^{-\delta^2/4\sigma^2} + 2\delta\phi(-\delta/\sigma\sqrt{2}) - \delta$$

and the dissimilarity is

$$D_{12} = (2\sigma/\sqrt{\pi})e^{-\delta^2/4\sigma^2} - 1) + 2\delta\phi(-\delta/\sigma\sqrt{2}) - \delta \tag{5.5}$$

6. ANODIV: Generalization of ANOVA. Apportionment of diversity corresponds to analysis of variance (ANOVA) of one-way classified data, where the populations are identified by the levels of a single factor. Let us now consider two factors A_1 and A_2 and represent the probability distribution associated with the i-th level of A_1 and the j-th level of A_2 by P_{ij} with apriori probability $\lambda_i^{(1)}\lambda_j^{(2)}$, $i=1, \ldots, p$; $j=1, \ldots, q$ ($\Sigma\ \lambda_i^{(1)}=1$, $\Sigma\ \lambda_j^{(2)} = 1$). Define

$$P_{..} = \Sigma\Sigma\ \lambda_i^{(1)}\lambda_j^{(2)}P_{ij}$$

$$P_{i.} = \Sigma_j\lambda_j^{(2)}P_{ij},\ P_{.j} = \Sigma_i\lambda_i^{(1)}P_{ij}$$

where $P_{..}$ is the overall distribution and $P_{i.}$ and $P_{.j}$ are the marginal distributions for the levels of the individual factors A_1 and A_2 respectively. Consider the analysis of diversity (ANODIV), i.e., a decomposition of the overall diversity, $H(P_{..})$, as in Table 2, using any measure of diversity.

TABLE 2. ANODIV for two way data

row no.	*due to*	*diversity*
1	factor(A_1)	$J^{(1)}(\{P_{i.}\}:\{\lambda_i^{(1)}\})$
2	factor(A_1)	$J^{(1)}(\{P_{.j}\}:\{\lambda_j^{(2)}\})$
3	interaction (A_1A_2)	$J^{(2)}(\{P_{ij}\}:\{\lambda_i^{(1)}\lambda_j^{(2)}\})$
4	between populations	$J^{(1)}(\{P_{ij}\}:\{\lambda_i^{(1)}\lambda_j^{(2)}\})$
5	within populátions	$\Sigma\Sigma\ \lambda_i^{(1)}\lambda_j^{(2)}H(P_{ij})$
6	Total	$H(P_{..})$

Rows (4) and (5) provide the analysis of diversity as between and within populations defined by the pq cells of the two way classification.

Rows (1) and (2) measure the diversities in the marginal distributions or the main effects of the factors A_1 and A_2. The residual diversity in row (3) represents the interaction between the factors A_1 and A_2. Thus the rows (1)-(3) and (5) provide an analysis of (6), the total diversity, as assignable to different causes. For a two way ANODIV, the diversity measure (H) need only satisfy the conditions C_1 and C_2 to ensure that $J^{(1)}$ and $J^{(2)}$ representing the main effects and interation are nonnegative.

If we have three way data, we can obtain a similar decomposition. The main effects are computed from one factor marginal distributions, two factor interactions from two factor marginal distributions, while the three factor interaction, the third order Jensen difference, is obtained by the formula

$$\begin{aligned} &J^{(3)}(\{P_{ijk}\}:\{\lambda_i^{(1)}\lambda_j^{(2)}\lambda_k^{(3)}\}) \\ &= H(P_{...}) - \Sigma\Sigma\Sigma\ \lambda_i^{(1)}\lambda_j^{(2)}\lambda_k^{(3)}H(P_{ijk}) \\ &-(A_1)-(A_2)-(A_3)-(A_1A_2)-(A_2A_3)-(A_3A_1) \end{aligned} \tag{6.1}$$

as in the case of analysis of variance with balanced data.

In order that all main effects and interactions are nonnegative in the ANODIV of three way classified data it is necessary to choose a diversity measure for which $J^{(1)}$, $J^{(2)}$ and $J^{(3)}$ are nonnegative, i.e., which satisfies the conditions C_1, C_2 and C_3. The results are easily generalized to analysis of m-way data for which we need a diversity measure whose Jensen differences up to order m are nonnegative. Note that $J^{(m)}$ represents the m factor interaction.

The generalization of ANOVA to ANODIV suggested above is quite general and can be applied on *any type of data* with observations *in any space*, by choosing a diversity measure with the *appropriate order* of convexity. Thus, if we have two way classified qualitive data, Shannon's entropy could be used, but not for higher order classified data.

7. Sampling Problems. In Section 6, we have discussed ANODIV in terms of population distributions, which provided various diversity components. In practice we have only observations from different populations, in which case we have problems of estimating the diversity components and testing hypotheses concerning them. We shall briefly describe how the appropriate methodology could be developed for this purpose.

To indicate how ANODIV provides a unified approach to the analysis of different types of data, let us consider the familiar analysis of variance of one way classified data as in Table 3.

TABLE 3. Populations and observations

1	2	...	k
x_{11}	x_{21}	...	x_{k1}
.	.		.
.	.		.
.	.		.
x_{1n_1}	x_{2n_2}	...	x_{kn_k}

We estimate the probability distribution function F_i for the i-th population by the empirical distribution function $\hat{F}_i$ based on the observations $x_{i1}, \dots, x_{in_i}$. Let us choose $\lambda_i = n_i/n.$ (where $n. = \Sigma n_i$) as the apriori probability of the i-th population. Further let us consider the variance functional

$$H(F) = \int (X_1-X_2)^2 dF(X_1)dF(X_2) \tag{7.1}$$

as the diversity measure. Substituting $\hat{F}_i$ for F_i *and* $\lambda_i = n_i/n.$ in the basic decomposition formula (5.1), we have

$$H(\Sigma(n_i/n.)\hat{F}_i) = \Sigma(n_i/n.)H(\hat{F}_i)+J^{(1)}(\{\hat{F}_i\}:\{\lambda_i\}). \tag{7.2}$$

Computing the various expressions in (7.2), using (7.1) for the H function, we obtain

$$n.^{-1}\Sigma_i\Sigma_j(x_{ij}-\bar{x}_{..})^2 = \Sigma_i(n_i/n.)n_i^{-1}\Sigma_j(x_{ij}-\bar{x}_{i.})^2+n.^{-1}\Sigma_i n_i(\bar{x}_{i.}-\bar{x}_{..})^2 \tag{7.3}$$

which is the usual ANOVA as within and between populations. The decomposition (7.3) is used in testing the hypothesis that the populations are the same, and in estimating the magnitude of differences between populations when the null hypothesis is rejected.

Instead of the variance functional, we can also use the city block distance functional

$$H(F) = \int |X_1-X_2| dF(X_1)dF(X_2). \tag{7.4}$$

This leads to the decomposition

$$n.^{-2}\Sigma_{ij}\Sigma_{rs}|x_{ij}-x_{rs}| = \Sigma_i(n_i/n.)n_i^{-2}\Sigma_r\Sigma_s|x_{ir}-x_{is}| + J^{(1)}\mathrm{L}(\{\hat{F}_i\}:\{\lambda_i\})$$

which could provide valid tests of significance for non-normal populations.

The ANODIV for one way classified categorical data (two way contingency table) using the second order entropy $(1 - \Sigma\, p_i^2)$, also called the Gini-Simpson index, is already illustrated in a paper by Light and Margolin (1971). We shall extend the analysis to two way classified categorical data using a more general quadratic entropy for a multinomial distribution with $p_1, \ldots, p_k$ as cell probabilities

$$H(p) = \Sigma\Sigma\, p_i p_j d_{ij} \tag{7.5}$$

where d_{ij} are chosen such that the $(k-1) \times (k-1)$ matrix

$$(d_{ik}+d_{jk}-d_{ij}-d_{kk}),\ i,j = 1, \ldots, k-1$$

is nonnegative definite to ensure complete convexity of the diversity measure (7.5). For further details and a characterization of (7.5), reference may be made to Rao (1982c).

Let us represent the observed numbers for k different categories in the (i,j)-th cell by n_{ijr}, $r = 1, \ldots, k$, and the estimated probabilities by $p_{ijr} = n_{ijr}/n_{ij.}$ where $n_{ij.} = n_{ij1}+\ldots+n_{ijk}$. If the cell numbers satisfy the conditons

$$n_{ij.} = n_{\ldots}\lambda_i^{(1)}\lambda_j^{(2)},\ \lambda_i^{(1)} = n_{i..}/n_{\ldots},\ \lambda_j^{(2)} = n_{.j.}/n_{\ldots} \tag{7.6}$$

then we can obtain the ANODIV using the diversity measure (7.5) as shown in Table 4.

TABLE 4. ANODIV: Two way categorical data

due to	*diversity*
factor (A_1)	$\Sigma\Sigma\, d_{rs}p_{..r}p_{..s} - \Sigma\,\lambda_i^{(1)}\Sigma\Sigma\, d_{rs}p_{i.r}p_{i.s}$
factor (A_2)	$\Sigma\Sigma\, d_{rs}p_{..r}p_{..s} - \Sigma\,\lambda_j^{(2)}\Sigma\Sigma\, d_{rs}p_{.jr}p_{.js}$
interaction (A_1A_2)	* (by subtraction)
between populations	* (by subtraction)
within populations	$\Sigma\Sigma\,\lambda_i^{(1)}\lambda_j^{(2)}\Sigma\Sigma\, d_{rs}p_{ijr}p_{ijs}$
total	$\Sigma\Sigma\, d_{rs}p_{..r}p_{..s}$

If $n_{ij.}$ do not satisfy the conditions (7.6), we can still carry out the ANODIV by choosing appropriate values of $\lambda_i^{(1)}$ and $\lambda_j^{(2)}$ (see Rao (1982c) for example).

The sampling distributions of the various expressions in Table 4 are likely to be complicated even in large samples. Their use in tests of significance and estimation of diversity components is under investigation.

REFERENCES

BURBEA, J. and RAO, C. R. (1982a). On the convexity of some divergence measures based on entropy functions. *IEEE Trans. Inf. Theory 28* 489–495.

BURBEA, J. and RAO, C. R. (1982b). On the convexity of higher order Jensen differences based on entropy functions. *IEEE Trans. Inf. Theory 28* 961–963.

CHAKRABORTY, R. (1974). A note on Nei's measure of gene diversity in a substructured population. *Humangenetik 21* 85–88.

DENNIS, B., PATIL, G. P., ROSSI, O. and TAILLE, C. (1979). A bibliography of literature on ecological diversity and related methodology. In *Ecological Diversity in Theory and Practice 1 CPH* 319–354.

HAVRDA, J. and CHARVÁT, F. (1967). Quantification method in classification processes: concept of structural α-entropy. *Kybernetika 3* 30–35.

KARLIN, S., KENNETT, R. and BONNE-TAMIR, B. (1979). Analysis of biochemical genetic data on Jewish populations: II Results and interpretations of heterogeneity indices and distance measures with respect to standards. *Ann. J. Hum. Genet. 31* 341–365.

LAU, K.-S. (1983). Characterization of the quadratic entropies of Rao. *Sankhyā Ser.* (in press).

LEWOTIN, R. C. (1972). The apportionment of human diversity. *Evolutionary Biology 6* 381-398.

LIGHT, R. J. and MARGOLIN, B. H. (1971). An analysis of variance for categorical data. *J. Amer. Statist. Assoc. 66* 534–544.

NEI, M. (1973). Analysis of gene diversity in subdivided populations. *Proc. Nat. Acad. Sci. 70* 3321–3323.

PARTHASARATHY, K. R. and SCHMIDT, K. (1972). *Positive Definite Kernels, Continuous Tensor Products and Central Limit Theorems of Probability Theory*. Springer-Verlag, Lecture Notes in Mathematics, 272.

RAO, C. RADHAKRISHNA (1982a). Diversity and dissimilarity coefficients: a unified approach. *Theoret. Pop. Biol. 21* 24–43.

RAO, C. RADHAKRISHNA (1982b). Diversity: its measurement, decomposition, apportionment and analysis. *Sankhyā Ser. A. 44* 1–22.

RAO, C. RADHAKRISHNA (1982c). Gini-Simpson index of diversity: a characterization, generalization and applications. *Utilitas Mathematica 218* 273–282.

RAO, C. RADHAKRISHNA and BOUDREAU, R. (1982). Diversity and cluster analyses of blood group data on some human populations. *Tech. Report 82–35,* Department of Mathematics and Statistics, University of Pittsburgh, Pittsburgh, PA.

RÉNYI, A. (1961). On measures of entropy and information. *Proc. Fourth Berkeley Symposium 1* 547–561.

SCHOENBERG, I. J. (1938). Metric spaces and positive definite functions. *Trans. Amer. Math. Soc. 44* 522–536.

SHANNON, C. E. (1948). A mathematical theory of communications. *Bell System Tech J. 27* 379–423, 623–656.

WELLS, J. and WILLIAMS, L. (1975). *Embeddings and Extensions in Analysis*. Springer-Verlag, Lecture Notes in Mathematics.

Inequalities in Statistics and Probability
IMS Lecture Notes–Monograph Series Vol. 5 (1984), 78–83

SOME SHARP MARTINGALE INEQUALITIES RELATED TO DOOB'S INEQUALITY

BY David C. Cox
Battelle Memorial Institute

Let $p > 1$. The best constant $C = C_{n,p}$ in the inequality $E(\max_{1 \leq i \leq n} |Y_i|)^p \leq C\,E|Y_n|^p$, where $Y_1, \ldots, Y_n$ is a martingale, is determined. For each n and p, the method allows one to construct a martingale attaining equality. As $n \to \infty$, $K_p n^{2/3}(q^p - C_{n,p}) \to 1$, where K_p is a known constant. As an application, the classical inequality of Doob is sharp. It is shown that equality cannot be attained by a non-zero martingale.

1. Introduction. Let $Y_1, Y_2, \ldots$ be a martingale with difference sequence $X_1 = Y_1$, $X_i = Y_i - Y_{i-1}$, $i = 2, 3, \ldots$ Thus, $E(X_i|X_1, \ldots, X_{i-1}) = 0$, $i = 2, 3, \ldots$. Let $p > 1$ and define $q = p/(p-1)$. The principal purpose of this paper is to determine the best constant $C = C_{n,p}$ in the inequality

$$E(\max_{1 \leq i \leq n} |Y_i|)^p \leq CE|Y_n|^p. \tag{1.1}$$

Although $C_{n,p}$ is found in implicit form, it can be easily approximated. For each n and p, the method allows one to construct a martingale attaining equality in (1.1), with $C = C_{n,p}$. Once the distribution of Y_1 is fixed, such a martingale is uniquely determined.

Furthermore, as $n \to \infty$, $C_{n,p} \to q^p$ at a rate proportional to $n^{-2/3}$. Specifically, $K_p n^{2/3}$ $(q^p - C_{n,p}) \to 1$, where K_p is a known constant. As an application, this provides a new proof that Doob's inequality (1953, p. 317)

$$E(\sup_{i \geq 1}|Y_i|)^p \leq q^p \sup_{i \geq 1} E|Y_i|^p \tag{1.2}$$

is sharp. An example to that effect was given previously by Dubins and Gilat (1978). Inequality (1.2) is strengthened to

$$E(\sup_{i \geq 1}|Y_i|)^p \leq q^p \sup_{i \geq 1} E|Y_i|^p - q\,E|Y_1|^p. \tag{1.3}$$

It follows from (1.3) that equality cannot be attained in (1.2) by a non-zero martingale. The sharpness of Doob's inequality for $p = 1$ (1953, p. 317)

$$E(\sup_{i \geq 1}|Y_i|) \leq [e/(e-1)](1 + E(\sup_{i \geq 1}|Y_i| \log^+ \sup_{i \geq 1}|Y_i|)],$$

is still an open question.

The method of this paper is based on results from the theory of moments (Kemperman (1968)), together with induction and the device of conditioning. Where applicable, it always leads to a sharp inequality and provides an example of a martingale attaining equality or nearly so. In principle, the method can be applied to many other martingale inequalities. For example, the author used it (Cox (1982)) to find the best constant in Burkholder's weak-L^1 inequality (Burkholder (1979)) for the martingale square function. The method does have the drawback of computational complexity, which sometimes makes it difficult or impossible to push the calculations through.

Section 2 contains statements of the results, together with comments and some proofs. In section 3, some needed analytic lemmas are established. Section 4 contains the main proofs, and an example for the case $p = 2$, $n = 3$ of (1.1).

AMS 1980 subject classifications. Primary 60E15, 60G42; secondary, 42B25.
Key words and phrases: Martingales, sharp inequalities, theory of moments, Doob's inequality.

2. Results. Throughout the paper, p is fixed. Dependence on p will often be suppressed in the notation. Let s, t be real numbers with $|t| \leq |s|$. For $0 < A \leq 1$ and $n = 1, 2, 3, \ldots$, define

$$\phi_n(s, t, A) = \inf E[|t + \Sigma_{i=1}^n X_i|^p - A \max(|s|^p, |t + X_1|^p, \ldots, |t + \Sigma_{i=1}^n X_i|^p)],$$

where the infimum is taken over all martingale difference sequences $X_1, \ldots, X_n$ with $EX_1 = 0$. The idea here is that $C_{n,p}^{-1}$ can be defined as the largest A for which $\phi_{n-1}(t, t, A) \geq 0$ for all t. Thus, to determine the best constant in (1.1), only the case $s = t$ of ϕ_n is needed. However, the inductive step (2.2) below requires knowing the value of ϕ_n for $|s| > |t|$ also. Note that $\phi_n(s, t, A) = |t|^p \phi_n(s/t, 1, A)$, for $t \neq 0$. This reduction does not, however, simplify the calculations made in the paper.

One has

$$\phi_1(s, t, A) = \inf\{E[|t + X|^p - A \max(|s|^p, |t + X|^p)]: EX = 0\}, \tag{2.1}$$

and by induction, conditioning on $X_1 = X$,

$$\phi_{n+1}(s, t, A) = \inf\{E\phi_n(\max(|s|, |t + X|), t + X, A): EX = 0\}, \tag{2.2}$$

for $n = 1, 2, \ldots$. Both (2.1) and (2.2) involve evaluating $\inf Ef(X)$ over all random variables X with $EX = 0$, where f is a given function. This is a standard problem of the theory of moments (Kemperman (1978), Cox (1982)) and can be solved graphically as shown in the proof of Theorem 1 given in Section 4.

THEOREM 1. *For $n = 1, 2, \ldots$, there exists $A_n \in (q^{-p}, 1]$, together with a function $g_n: (0, A_n] \to [0, 1)$ for which*

$$\begin{aligned}\phi_n(s, t, A) &= |t|^p - A|s|^p \quad \text{if } |t| \leq g_n(A)|s|,\\ &= p\, g_n(A)^{p-1}|t|\,|s|^{p-1} - [A + (p-1)g_n(A)^p]|s|^p \quad \text{if } g_n(A)|s| \leq |t| \leq |s|,\end{aligned}$$

for $0 \leq A \leq A_n$, while $\phi_n(s, t, A) = -\infty$ if $A > A_n$.

The constant A_n and the function g_n are defined inductively as follows. Let

$$\phi(y,x) = \{x/((p-1)[(py^{p-1} - (p-1)y^p - x)^{1-q} - 1])\}^{1/p} \tag{2.3}$$

for $0 \leq x, y \leq 1$ and $0 \leq \gamma(y,x) \equiv py^{p-1} - (p-1)y^p - x < 1$. Define $g_0(x) \equiv 1$, $g_{n+1}(x) = \phi(g_n(x), x)$, $n = 0, 1, \ldots$. Then, for $n = 1, 2, \ldots$, there is a unique $q^{-p} < A_n \leq 1$ with $g_n(A_n) = 0$; the domain of g_n is precisely $(0, A_n]$. One has $A_n > A_{n+1} > q^{-p}$ and $\lim_{n\to\infty} A_n = q^{-p}$. More precisely,

THEOREM 2. $\lim_{n\to\infty} n^{2/3}(A_n - q^{-p}) = (2\pi^2 q^{1-3p})^{1/3}$.

For $0 < x \leq q^{-p}$, the sequence $\{g_n(x)\}$ is strictly decreasing with limit $g(x) = y$, the larger of the two roots of the equation

$$\phi(y, x) \equiv (p-1)(y^p - y^{p-1}) + x = 0.$$

In particular, $g(q^{-p}) = q^{-1}$.

COROLLARY 1. *Let $n \geq 2$ and $0 < A \leq A_{n-1}$. Suppose that $Y_1, \ldots, Y_n$ is a martingale. The following inequality is sharp*

$$E|Y_n|^p \geq AE(\max_{1 \leq i \leq n} |Y_i|)^p + \gamma(g_{n-1}(A), A)E|Y_1|^p.$$

Proof: Let $X_i = Y_{i+1} - Y_i$, $i = 1, \ldots, n-1$. Then, conditional on $Y_1 = t$, $X_1, \ldots X_{n-1}$ is a martingale difference sequence with $EX_1 = 0$. Now apply Theorem 1 with $s = t$, and then integrate with respect to the distribution of Y_1. □

COROLLARY 2. *The best constant $C = C_{n,p}$ in (1.1) is $C_{n,p} = A_n^{-1}$.*

Proof. One has $\gamma(g_{n-1}(A), A) \geq 0$ for $0 < A \leq A_n$ with equality iff $g_n(A) = 0$, i.e., iff $A = A_n$. Now apply Corollary 1. □

The proof of Theorem 1, presented in Section 4, shows how a martingale attaining equality in (1.1), with $C = A_n^{-1}$, may be constructed. Moreover, once the distribution of Y_1 is fixed, such a martingale is uniquely determined. An example is given after the proof of Theorem 1.

From Theorem 2, the asymptotic behavior of $C_{n,p}$ can be characterized.

COROLLARY 3. $\lim_{n\to\infty} n^{2/3}(q^p - C_{n,p}) = (2\pi^2 q^{3p+1})^{1/3}$.

Letting $n\to\infty$ in Corollary 1, one obtains

COROLLARY 4. *Let* $Y_1, Y_2, \ldots$ *be a martingale, and* $0 < A \leq q^{-p}$. *The following inequality is sharp*

$$\sup_{i\geq 1} E|Y_i|^p \geq A\, E(\sup_{i\geq 1}|Y_i|)^p + g(A)^{p-1}E|Y_1|^p.$$

In particular, letting $A = q^{-p}$, one obtains (1.3).

Proof. Just note that $\gamma(g(A), A) = g(A)^{p-1}$, see Theorem 2. □

COROLLARY 5. *Doob's inequality (1.2) is sharp. However, equality cannot be attained by a non-zero martingale.*

Proof. Sharpness follows from $C_{n,p} \to q^p$. Equality in (1.2) forces $Y_1 \equiv 0$, from (1.3). Applying the same argument to the martingale $Y_2, Y_3, \ldots$, one finds $Y_2 = 0$, etc. □

3. Analytic Preliminaries. The object of this section is to establish some needed results concerning the functions g_n.

LEMMA 1. *The function* ϕ, *defined by (2.3), has the following properties.*

(3.1) $$\phi(y, x) \leq y \text{ with equality iff } \theta(y, x) = 0$$

(3.2) $$\delta\phi/\delta y > 0, \text{ for } 0 < x, y < 1, 0 < \gamma(y, x) < 1.$$

(3.3) $$\delta\phi/\delta x < 0, \text{ for all } y, \text{ if } q^{-p} < x < 1.$$

Proof. First consider (3.1). One has $\phi(y, x) \leq y$ iff

(3.4) $$(p-1)y^p[\gamma(y, x)^{1-q} - 1] - x \geq 0.$$

The derivative of the LHS of (3.4) with respect to x is given by $y^p\gamma(y, x)^{-q} - 1$. Since $\gamma(y, x) + \theta(y, x) = y^{p-1}$, it follows that the minimum value of the LHS of (3.4) is 0, taken when $\theta(y, x) = 0$. Since $\theta(y, x) > 0$ for all $y > 0$ when $x > q^{-p}$, one has $\phi(y, x) < y$ for all y in this case. Next, a straightforward calculation gives

(3.5) $$\delta\phi/\delta y = [(q-1)x y^{p-2}(1-y)] / [\phi^{p-1}\gamma^q(\gamma^{1-q} - 1)^2]$$

where $\phi = \phi(y, x)$, $\gamma = \gamma(y, x)$, which establishes (3.2). Finally,

(3.6) $$\delta\phi/\delta x = [\gamma - \gamma^q - (q-1)x] / [p(p-1)\phi^{p-1}\gamma^q(\gamma^{1-q} - 1)^2]$$

The numerator in (3.6) is $(1 - q)\theta(\gamma^{q-1}, x)$, which yields (3.3). □

LEMMA 2. *There is a unique* $q^{-p} < A_n \leq 1$ *with* $g_n(A_n) = 0$, $n = 1, 2, \ldots$; *the domain of* g_n *is* $(0, A_n]$. *One has* $\theta(g_n(x), x) > 0$ *for* $0 < x \leq A_n$. *For* $0 < x \leq q^{-p}$, $g_n(x) \downarrow g(x)$ *as* $n\to\infty$.

Proof. First consider $0 < x \leq q^{-p}$. I claim that $1 \geq g_n(x) > g(x)$ for all $n = 0, 1, 2, \ldots$. Since this is trivial for $n = 0$, assume that it holds for some $n \geq 0$. Then,

$$1 > \gamma(g_n(x), x) > \gamma(g(x), x) = g(x)^{p-1} > 0,$$

so that $g_{n+1}(x)$ is defined. Next,

$$(p-1)g(x)^p[\gamma(g_n(x),x)^{1-q}-1] < (p-1)g(x)^p[\gamma(g(x),x)^{1-q}-1] = x.$$

It follows that $g_{n+1}(x) > g(x)$. From Lemma 1, $g_{n+1}(x) < g_n(x)$, so that $g_n(x)\downarrow g(x)$ as $n \to \infty$, since $y = \lim_{n\to\infty} g_n(x)$ must satisfy $\phi(y,x) = y$. Suppose next that, for some $n \geq 1$, it has been established that the domain of g_n is $(0, A_n]$ with $g_n(A_n) = 0$, where $q^{-p} < A_n \leq 1$. I claim that $g_n'(x) < 0$ for $q^{-p} < x < A_n$. This is clear from (3.3) for $n = 1$, since $g'_1(x) = \delta\phi/\delta x$. Since $g'_{j+1}(x) = \delta\phi/\delta y\, g_j'(x) + \delta\phi/\delta x$, for $j = 1, 2, \ldots$, the claim follows by induction from (3.2) and (3.3). By the same argument, $g_{n+1}(x)$ is strictly decreasing on its domain, for $x > q^{-p}$. Since $g_{n+1}(x) < g_n(x)$ where both are defined, and $g_{n+1}(q^{-p}) > g(q^{-p}) > 0$, the existence and uniqueness of $A_{n+1} < A_n$ follow. Finally, $\theta(g_n(x), x) > 0$ for $0 < x \leq A_n$ follows from Lemma 1. □

4. Main Proofs.

Proof of Theorem 1. The properties of g_n and A_n a relevant to this proof have been established in Section 3. If one defines $\phi_0(s, t, A) = |t|^p - A|s|^p$, then the theorem holds for $n = 0$. Moreover, see (2.1) and (2.2), the inductive relation between ϕ_n and ϕ_{n+1} remains valid for $n = 0$. Assume by induction, therefore, that the theorem is true for some $n \geq 0$. Let $0 < A \leq A_n$ (where $A_0 = 1$), and, without loss of generality, $t \geq 0$. From (2.2) one finds

$$\phi_{n+1}(s, t, A) = \inf\{Eh(X):EX = t\},$$

where $h(x)$ is given by

$$\begin{aligned} &|x|^p - A|s|^p \quad \text{if } |x| \leq g_n(A)|s| \\ &p\, g_n(A)^{p-1}|s|^{p-1}|x| - [A + (p-1)g_n(A)^p]\,|s|^p \quad \text{if } g_n(A)|s| \leq |x| \leq |s| \\ &\gamma(g_n(A), A)|x|^p \quad \text{if } |x| > |s|. \end{aligned}$$

It is well-known (Kemperman (1968), Cox (1982)) that the required infimum is given by the height, at location $x = t$, of the lower boundary of the convex hull of the graph of h. For $A_{n+1} < A \leq A_n$, $\gamma(g_n(A), A) < 0$ so the infimum is $-\infty$. Now suppose $0 < A \leq A_{n+1}$. Clearly, $h'(x)$ is continuous at $x = \pm\, g_n(A)|s|$ so that $h(x)$ is convex for $|x| < |s|$, and also for $|x| > |s|$. Moreover, $h'_+(|s|) = \gamma(g_n(A), A)\, p\, |s|^{p-1} < p\, g_n(A)^{p-1}\, |s|^{p-1} = h'_-(|s|)$, since $\theta(g_n(A), A) > 0$. It follows that the convex hull of the graph of $h(x)$ for $x \geq 0$ is formed by drawing a common tangent from the part for $0 \leq x \leq g_n(A)|s|$ to the part for $x > |s|$. The tangent to $y = |x|^p - A|s|^p$ at $x = x_0 > 0$ has equation

$$y = x_0^p(1-p) - A|s|^p + p\, x_0^{p-1}x. \tag{4.1}$$

The slope of $h(x)$ for $x > |s|$ is $p\, \gamma(g_n(A), A)x^{p-1}$, which coincides with the slope of (4.1) iff $x_0 = \gamma(g_n(A), A)^{q-1}x$. It follows that the required common tangent has a point of tangency at $x_0 = \phi(g_n(A), A)|s| = g_{n+1}(A)|s|$ with the graph of $h(x)$ for $0 \leq x \leq g_n(A)|s|$. Using (4.1) one immediately obtains the required formula for $\phi_{n+1}(s, t, A)$. This completes the inductive step and proves Theorem 1. □

Remark 1. It is clear from the above proof that $\phi_{n+1}(s, t, A) = \inf\{Eh(X)\colon EX = t\}$ is attained by a unique random variable X, for each s and t. Specifically, $X \equiv t$ if $|t| \leq g_{n+1}(A)|s|$, while X takes the two values $g_{n+1}(A)|s|$ sgn t, $\gamma(g_n(A), A)^{1-q}\, g_{n+1}(A)|s|$ sgn t, if $g_{n+1}(A)|s| \leq |t| \leq |s|$. By working backwards, then, the unique martingale attaining the value $\phi_n(s, t, A)$ can always be constructed, see Example 1 below. Further, once the distribution of Y_1 is fixed, a unique martingale attaining equality in (1.1) with $C = C_{n,p}$ is determined.

Example 1. Let $p = 2$, so that $\phi(y, x) = [x(x + (1-y)^2)^{-1} - x)]^{1/2}$. A calculation shows that $A_3 = 16/25$. Hence, if X_1, X_2, X_3 is a martingale difference sequence, the following inequality is sharp.

(4.2) $\quad E\max[X_1^2, (X_1 + X_2)^2, (X_1 + X_2 + X_3)^2] \leqslant (25/16)E(X_1 + X_2 + X_3)^2.$

The following martingale difference sequence attains equality. Let

$X_1 \equiv 1, P[X_2 = 1] = 3/8, P[X_2 = -3/5] = 5/8$. Then

$P[X_3 = 4/3|X_2 = 1] = 3/8, P[X_3 = -4/5|X_2 = 1] = 5/8,$

$P[X_3 = 0|X_2 = -3/5] = 1.$ □

Note that equality can be attained in (4.2) with an *arbitrary* distribution for X_1. Namely, multiply the difference sequence given above by any variable X independent of (X_1, X_2, X_3). However, once the distribution of X_1 is fixed, a unique martingale attaining equality in (4.2) is defined.

Proof of Theorem 2. From results of Section 2, it is clear that $A_n \to q^{-p}$. After all, $\lim_{n\to\infty} A_n \geqslant q^{-p}$ exists. Moreover, $\lim_{n\to\infty} A_n > q^{-p}$ is impossible because the equation $\phi(y,x) = y$ has no solution if $x > q^{-p}$

It follows from (3.5) that $\delta\phi/\delta x$ is continuous at the point (q^{-1}, q^{-p}), where it takes the value 1. Let $0 < \epsilon < 1/4$ be otherwise arbitrary and choose $\delta > 0$ such that $|y - q^{-1}| < \delta$, $|x - q^{-p}| < \delta \Rightarrow |\delta\phi/\delta y - 1| < \epsilon$. Choose n_0 so that $n \geqslant n_0 \Rightarrow A_n - q^{-p} < \delta$. Then, for $n \geqslant n_0, j = 0, \ldots, n$, one has $g_j(A_n) \leqslant g_j(q^{-p})$. Also, $g_j(A_n) \geqslant g_j(A_{n_0})$, for $j = 0, \ldots, n_0$. Since $g_j(q^{-p}) \downarrow q^{-1}$ as $j \to \infty$, the above two inequalities taken together imply that there exists n_1, independent of n, such that $|g_j(A_n) - q^{-1}| < \delta, j = 0, \ldots, n$, with the possible exception of n_1 values of j, i.e., all but finitely many members of the sequence $g_j(A_n), j = 0, \ldots, n$, lie within δ of q^{-1} independently of n. Now fix $n \geqslant n_0$ and let $y_j = g_j(A_n)$. Thus,

$$(y_j - y_{j-1})/(\phi(y_{j-1}, A_n) - y_{j-1}) = 1, j = 1, \ldots n.$$

Now

$$\int_{y_{j-1}}^{y_j} dy/(\phi(y, A_n) - y) = 1 + \rho_j, j = 1, \ldots n{-}1.$$

(where $j = n$ is excluded since the corresponding integral is not finite). One has $|\rho_j| \leqslant 1/2\, M_j(1 - M_j)^{-2}$, provided $M_j = \sup\{|\delta\phi/\delta y - 1| : y_j \leqslant y \leqslant y_{j-1}\} < 1$. Hence, all but n_1 of the $|\rho_j|$ are smaller than ϵ. Since ϵ is arbitrary it follows that

$$\lim_{n\to\infty} 1/n \int_{g_{n-1}(A_n)}^{1} dy/(y - \phi(y, A_n)) = 1$$

As $n\to\infty$, $g_{n-1}(A_n) \to U$, where $U < q^{-1}$ is the solution of the equation $\gamma(U, q^{-p}) = 0$. Therefore,

(4.3) $$\lim_{n\to\infty} 1/n \int_U^1 dy/(y - \phi(y, A_n)) = 1,$$

Next, examine the asymptotic behavior of the integral $I(x) = \int_U^1 dy/(y - \phi(y, x)$, as $x \downarrow q^{-p}$. Clearly, $I(x) \to \infty$ as $x \downarrow q^{-p}$. It is well-known that its asymptotic behavior is determined by the behavior of $y - \phi(y, x)$ near its minimum (as a function of y). For x close to q^{-p}, this minimum is attained at a value of y close to q^{-1}. Recalling the definition of $\theta = \theta(y, x)$, one has, for (y, x) close to (q^{-1}, q^{-p}),

$$\begin{aligned}(y - \phi(y, x))^{-1} &= 2q^{2-2p}(q-1)^{-1}\theta^{-2} + o(\theta^{-2})\\ &= 8/(qp^3[(y - q^{-1})^2 + 2(x - q^{-p})(p-1)^{-2}q^{p-3}]^2)\end{aligned}$$

on expanding θ in a Taylor series about (q^{-1}, q^{-p}). It follows that

$$\lim_{x \downarrow q^{-p}} (2\pi^2)^{-1/2} (q^{3p-1}(x-q^{-p})^3)^{1/2} I(x) = 1. \tag{4.4}$$

The conclusion of Theorem 2 follows from (4.3) and (4.4). □

REFERENCES

BURKHOLDER, D. L. (1979). A sharp inequality for martingale transforms. *Ann. Probability 7* 858–863.

COX, D. C. (1982). The best constant in Burkholder's weak-L^1 inequality for the martingale square function. *Proc. Amer. Math. Soc. 85* 427–433.

DOOB, J. L. (1953). *Stochastic Processes*. Wiley, New York.

DUBINS, L. E., and GILAT, D. (1978). On the distribution of maxima of martingales. *Proc. amer. Math. Soc. 68* 337–338.

KEMPERMAN, J. H. B. (1968). The general moment problem, a geometric approach. *Ann. Math. Statist. 39* 93–122.

Inequalities in Statistics and Probability
IMS Lecture Notes–Monograph Series Vol. 5 (1984), 84–103

LEAST ABSOLUTE VALUE AND MEDIAN POLISH

BY J. H. B. Kemperman
University of Rochester

We are interested in best L_p-approximations $\Sigma_r\beta_r g_r(x)$ to a given finite array of numbers $z^{(o)}(x)$, $(x \in X)$. For the case $p > 1$, a natural iterated polishing method is shown to converge to the unique optimal solution. Let $p = 1$. Several conditions are obtained, each of which is necessary and sufficient for a given array of residuals $z(x)$ $(x \in X)$ to be optimal. Detailed results are derived for the case of a two-way $m\times n$ layout, allowing several observations $z^{(o)}_{ijk}$ in cell (i, j). For instance, a set of residuals is optimal if and only if there exists a solution to an associated moment problem with given marginals, which depends only on the signs σ_{ijk} of the residuals z_{ijk}. This criterion leads to an elegant and efficient max-flow-min-cut type of algorithm for calculating a best L_1-approximation. For the case of a single observation in each cell, it is also determined precisely which pairs (m, n) are 'safe' for Tukey's median polish, in the sense that the endproduct of an $m\times n$ polish is necessarily a best L_1-approximation. The answer depends on the type of allowable medians.

1. Introduction. Let the $m \times n$ matrix $\mathbf{Z}^{(o)} = (z^{(o)}_{ij})$ represent a two-way table of observations. An elementary way of arriving at a reasonable additive approximation $\alpha_i + \beta_j$ is by means of median polish, as developed by Tukey (1977); see Section 4 for further details. An algorithm in APL and further comments can be found in Anscombe ((1981) p. 106, 382).

One motivating idea behind median polish is that it *might* minimize the L_1-norm of the matrix $\mathbf{Z} = (z_{ij})$ of residuals $z_{ij} = z^{(o)}_{ij} - \alpha_i - \beta_j$. However, this is not always true as follows already from the well-known fact that the norm of the final endproduct of a median polish or mid-median polish may not be the same when starting with a polishing of the rows as when starting with a polishing of the columns.

These endproducts will be called an EMP or EMMP, respectively. More generally, an $m \times n$ matrix $\mathbf{Z}$ will be called an EMP or EMMP, respectively, when 0 is a median or mid-median, respectively of each row and each column.

The matrix $\mathbf{Z}$ of residuals will be said to be *optimal* if its norm cannot be further reduced. For this it is necessary that $\mathbf{Z}$ be an EMP. It is shown (Theorem 6) that for each choice of (m, n) there exist non-optimal EMP's. There even exist non-optimal EMMP's, unless (m, n) is one of the special pairs $(2, n)$; $(3, 4)$; $(4, 4)$; $(4, 5)$ and $(4, 6)$, (assuming that $2 \leq m \leq n$). Thus, if $m = 4$ and $n = 6$ then the endproduct of a convergent mid-median polish process is always optimal. This is false when $m = 3$ and $n = 3$ or 5.

The main purpose of the present paper is to derive efficient tests for optimality together with explicit procedures for improving a given non-optimal matrix. Many of our results lead to an explicit algorithm, usually safer and faster than median polish, though that algorithm may not be spelled out in any great detail. For, our principal goal is to achieve a good theoretical understanding of the main problem.

Most results are developed for the general regression problem, where one wants to minimize the L_p-norm (2.1) by a suitable choice of the free parameters β_r. Median polish carries over to this general problem in a natural way. We show in Theorem 1 that this

---

[1] Research supported in part by the National Science Foundation.
AMS 1980 subject classifications. Primary, 62J05; secondary, 41A50, 65D10.
Key words and phrases: Linear regression, least absolute value, best L_1 approximation, criterion for optimality, two-way layout, median polish, given marginals, max-flow-min-cut, algorithm.

generalized polish always converges to the unique optimal solution, provided $p > 1$.

From Section 5 on, it is assumed that $p = 1$. Then the optimality of $\mathbf{Z} = \{z(x); x \in X\}$ depends only on the associated sign pattern $\{\text{sgn } z(x); x \in X\}$ and one may speak of optimal sign patterns. It turns out that an optimal sign pattern remains optimal when one or more of the elements $+1$ or -1 are replaced by 0. In the different criteria for optimality, a central role is played by the set $D = \{x \in X: z(x) = 0\}$. For instance, in the case of an n-way layout, one necessary and sufficient condition for optimality requires the existence of a measure on D having preassigned marginals, see Section 7. The sections 8 and 9 are concerned with a two-way layout, allowing several observations per cell.

There is a large literature on explicit algorithms leading to an optimal L_1-approximation. Each of these algorithms amounts to a descent method of some type, often restricted to the finite set of basic solutions of the associated linear programming problem. See the surveys by Gentle (1977) and Kennedy and Gentle (1980) pp. 515–559.

A selected list of such papers has been included in the bibliography. Space did not allow us to give an adequate discussion of the many cross relations which exist between these papers and the present one.

2. Stating the problem. In this paper, X is a fixed finite index set and $Z^{(o)} = \{z^{(o)}(x): x \in X\}$ a given collection of real numbers or observations. Further, $g_r : X \to R$ $(r = 1, \ldots, M)$ is a given set of linearly independent functions on X. We will be interested in diferent aspects of the problem to

$$\text{(2.1)} \qquad \text{minimize:} \quad S = \Sigma_{x\epsilon X}\, \omega(x)|z^{(o)}(x) - \Sigma^M_{r=1}\beta_r g_r(x)|^p.$$

Here, $p \geqslant 1$ and only the β_r are unknown. The weights $\omega(x) > 0$ may indicate the multiplicity or importance of the corresponding observations.

This problem arises in many ways, for instance, as a maximum likelihood problem when

$$z^{(o)}(x) = \Sigma^M_{r=1}\, \beta_r g_r(z) + \eta(x), \qquad (x \in X)$$

where the errors $\eta(x)$ $(x \in X)$ are independent with a density type $c(\omega)\exp(-\omega(x)|y|^p)$, $(y \in R)$.

We will especially be interested in the case $p = 1$. One situation we have in mind, see Sections 8 and 9, is that of a two-way $m \times n$ layout with observations $z^{(o)}_{ijk}$ in cell (i, j), with $i \in Y_1 = \{1, \ldots, m\}$; $j \in Y_2 = \{1, \ldots, n\}$ and $k = 1, \ldots, k_{ij}$, where one wants to minimize the L_1-norm

$$\text{(2.2)} \qquad S = \Sigma_{i,j,k}|z^{(o)}_{ijk} - \alpha_i - \beta_j|.$$

Here, X is taken as the set of triplets (i, j, k) with $i \in Y_1, j \in Y_2, k \in \{1, \ldots, k_{ij}\}$ while $\omega(x) = 1$. Further, $M = m + n$ and

$$\text{(2.3)} \qquad \begin{aligned} g_r(i, j, k) &= \delta^r_i \qquad \text{for } r = 1, \ldots, m; \\ &= \delta^{r-m}_j \qquad \text{for } r = m + 1, \ldots, M; \end{aligned}$$

($\delta^r_q = 1$ or 0 if $q = r$ or $q \neq r$, respectively). The general layout problem is to minimize a norm of the form

$$\text{(2.4)} \qquad S = \Sigma_{x\epsilon X}\, \omega(x)|z^{(o)}(x) - \Sigma_{t\epsilon T}\, \beta_t(\phi_t(x))|.$$

Here, for each $t\epsilon T$, $\phi_t: X \to Y_t$ is a given function, while the $\beta_t(y)$ are unknown parameters. Often X is a subset of $\mathcal{R}^n$ while $\phi_t(y)$ is expressed in terms of the coordinates x_h of x, e.g., $\phi_t(x) = x_1$ or $\phi_t(x) = (x_1, x_2)$.

In this illustration, the function g_r in (2.1) becomes

$$\text{(2.5)} \qquad g_{t,y}(x) = \delta^y_{\phi_T(x)}, \quad (t\epsilon T; y\epsilon Y_t)$$

while $\beta_t(y)$ plays the role of β_r. The support of $g_{t,y}$ is

(2.6) $$L_t(y) = \{x \in Z: g_{t,y}(x) \neq 0\} = \{x \in X: \phi_t(x) = y\}$$

and will also be referred to as a 'layer'. In the case (2.2) one would have $T = \{1, 2\}$, $\phi_1(i, j, k) = i$ and $\phi_2(i, j, k) = j$. Moreover, $L_1(i)$ and $L_2(j)$, respectively, would be the set of points $x = (i, j, k)$ having a fixed component i or j, respectively.

The general layout problem is equivalent to having, for each $t \in T$, a partition of X into disjoint sets $L_t(y)$ with associated additive parameters $\beta_t(y)$.

Remark. As will be shown in a subsequent paper, most results of the present paper carry over to the case where the exponent p in (2.1) is replaced by a function $p(x) \geq 1$ $(x \in X)$.

3. The p-Centre of a Mass Distribution. Be given a mass distribution on the reals having mass $q_i > 0$ at y_i $(i = 1, \ldots, n)$ and suppose one wants to minimize $\psi(s) = \Sigma_{i=1}^n q_i|s - y_i|^p$. This is somewhat comparable to least squares relative to the (variable) weights $q_i|y_i - s|^{p-2}$, as observed by Mosteller and Tukey (1977) p. 365. If $p < 2$ then relatively less weight is given to the very large observations.

If $p > 1$ then $\psi(s)$ is strictly convex and the minimum on hand is uniquely achieved at the so-called p-centre $s^\circ = \text{mean}_p\{y_i: q_i\}$ of the mass distribution. It is the unique solution of the equation

$$\Sigma_{i=1}^n \text{sgn}(s^\circ - y_i) q_i |s^\circ - y_i|^{p-1} = 0,$$

(where $\text{sgn}(u) = -1$, 0 or $+1$, depending on the sign of u). In particular, $\text{mean}_2\{y_i: q_i\}$ is nothing but the ordinary mean.

If $0 < p < 1$ then $\psi(s)$ would be strictly concave as long as s differs from the y_i and, thus, $\psi(s)$ takes its minimal value (only) at one of the points y_i. If $p = 1$ then $\psi(s)$ is piecewise linear and convex and, hence, attains its minimal value precisely at the values s where the nondecreasing derivative $\psi'(s) = \Sigma_{i=1}^n q_i \text{sgn}(s - y_i)$ changes sign from negative to positive; this always includes one of the points y_i. Such a 1-mean or median s° may also be defined by the inequalities

(3.1) $$\Sigma_{y_i < s^\circ} q_i \leq Q/2 \leq \Sigma_{y_i \leq s^\circ} q_i, \quad \text{where } Q = \Sigma_{i=1}^n q_i.$$

This median is unique unless the second sum can take the value $Q/2$. The latter happens, for instance, when n is even and $q_i = 1$ for all i. The notation

$$s^\circ = \text{mean}_1\{y_i: q_i\} = \text{med}\{y_i: q_i\} \quad \text{or} \quad \text{mean}_1\{y_i: q_i\} = s^\circ$$

will simply indicate that s° is a median for the distribution on hand. The set of all medians is a finite closed interval $[s', s'']$. Its midpoint $(s' + s'')/2$ is called the mid-median and will be denoted as $\text{Med}\{y_i: q_i\}$.

LEMMA 1. *In order that $\beta^* = (\beta^*_1, \ldots, \beta^*_M)$ achieves the minimum in (2.1), it is necessary that the residuals*

(3.2) $$z(x) = z^{(o)}(x) - \Sigma_{r=1}^M \beta^*_r g_r(x) \qquad (x \in X)$$

satisfy

(3.3) $$\text{mean}_{p,r}\{z(x) = 0\} \quad \textit{for } r = 1, \ldots, M.$$

The latter mean is defined as

(3.4) $$\text{mean}_{p,r}\{z(x)\} = \text{mean}_p\{z(x)/g_r(x) : \omega(x)|g_r(x)|^p\}.$$

If $p > 1$ then the minimum in (2.1) is achieved at a unique point β^* and condition (3.3) is also sufficient.

Proof. Fix $1 \leqslant r \leqslant M$. A necessary condition is that

(3.5) $$\Sigma_{x \epsilon X} \omega(x)|z(x) - sg_r(x)|^p \qquad (s \epsilon R)$$

takes its smallest value at $s = 0$. Since one may as well restrict $x \epsilon X$ to the points with $g_r(x) \neq 0$, this is the same as saying that $s = 0$ minimizes the sum

$$\Sigma_{x \epsilon X} \omega(x)|g_r(x)|^p|s - z(x)/g_r(x)|^p.$$

From the remarks preceding the Lemma, the latter in turn is equivalent to (3.3).

If $p \geqslant 1$ then the sum in (2.1) defines a nonnegative and convex function $F(\beta) = F(\beta_1, \ldots, \beta_M)$ on R^M. Condition (3.3) says precisely that F is minimal at β^* relative to changes in a single variable only. If $p > 1$ then F is of class C^1 and strictly convex (since the g_r are linearly independent). This easily yields the last assertion. □

Comments. Also note that the sum in (3.5) assumes its smallest value at $s°$ if and only if $s° = \text{mean}_{p,r}\{z(x)\}$, as defined in (3.4). In the case $p = 1$ this means that

(3.6) $$\min_s \Sigma_{x \epsilon X} \omega(x)|z(x) - sg_r(x)|$$

is achieved at $s°$ if and only if $s°$ is a median of the set of all numbers $z(x)/g_r(x)$ with $g_r(x) \neq 0$ having corresponding weights $\omega(x)|g_r(x)|$.

The last two assertions of Lemma 1 would be false when $p = 1$. Namely, medians and, in general, optimal L_1-approximations are often not unique. The following example shows that condition (3.3) is not sufficient for optimality when $p = 1$.

Choose $X = \{1, 2, 3\}$ and $\omega(x) = 1$. Further, $M = 2$; $g_1(1) = g_2(2) = 0$ while $g_r(x) = 1$, otherwise. Finally, $z(1) = -1$; $z(2) = +1$ and $z(3) = 0$. Then condition (3.3) holds with $\beta^*_1 = \beta^*_2 = 0$. Namely, it then says that 0 is a median of the set of numbers $z(2) = +1$ and $z(3) = 0$, with weight 1 each, and also that 0 is a median of the set of numbers $z(1) = -1$ and $z(3) = 0$, with weight 1 each. Which is true. Nevertheless, the minimum in (2.1) is not achieved at $\beta = (0, 0)$. For, the L_1-norm (equal to 2) of $\{z(x)\}$ can be reduced to 0 since $z(x) - g_1(x) + g_2(x) = 0$ for all $x \epsilon X$. Essentially, one is confronted here with the simple function

$$f(u, v) = |u - 1| + |v + 1| + |u + v|.$$

It is convex and satisfies $f(0, 0) \leqslant f(u, 0)$, for all u, and $f(0, 0) \leqslant f(0, v)$, for all v. Nevertheless, $f(1, -1) = 0 < 2 = f(0, 0)$.

4. Median Polish. Tukey (1977) developed in detail the idea of calculating a reasonably good additive approximation to a given n-way layout of observations by so-called median polish. For a two-way layout as in (2.2), this additive approximatioon $\alpha_i + \beta_j$ to $z^{(o)}_{ijk}$ is derived as follows.

One starts with the 'matrix' $\mathbf{Z}^{(o)} = (z^{(o)}_{ijk})$ and applies a (median) row polish, yielding the matrix

$$\mathbf{Z}^{(1)}_{ijk} = (\mathbf{z}^{(1)}_{ijk} = z^{(o)}_{ijk} - \beta^{(1)}_i).$$

Here, the adjustment $\beta^{(1)}_i$ of the i-th row is taken as a fixed median (often the mid-median) of the set of numbers $z^{(o)}_{ijk}$ in i-th row (i fixed). Next, one applies a column polish to $\mathbf{Z}^{(1)}$ yielding

$$\mathbf{Z}^{(2)} = (z^{(2)}_{ijk} = z^{(1)}_{ijk} - \beta^{(2)}_j).$$

Here, $\beta^{(2)}_j$ is a fixed median of the set of numbers $z^{(1)}_{ijk}$ with j fixed. Polishing the rows of $\mathbf{Z}^{(2)}$ one obtains $\mathbf{Z}^{(3)}$ and so on. In general.

(4.1) $z_{ijk}^{(2h+1)} = z_{ijk}^{(2h)} - \beta_i^{(2h+1)}; \quad z_{ijk}^{(2h+2)} = z_{ijk}^{(2h+1)} - \beta_i^{(2h+2)}, (h = 0,1, \dots).$

As observed by Gabriel (1983), a more efficient and self-correcting procedure would be to keep the original matrix $\mathbf{Z}^{(o)}$ and store not the $\mathbf{Z}^{(g)}$ but only the current cumulative sums $a_i^{(h)} = \Sigma_{g=1}^{h} \beta_i^{(2g-1)}$ and $b_j^{(h)} = \Sigma_{g=1}^{h} \beta_j^{(2g)}$ of the row and column adjustments. For, these can be recursively computed: (i) $a_i^{(h)}$ is the median of the numbers $z_{ijk}^{(o)} - b_i^{(h-1)}$, ($i$ fixed; $b_j^{(o)} = 0$); (ii) $b_j^{(h)}$ is the median of the set of numbers $z_{ijk}^{(o)} - a_i^{(h)}$, ($j$ fixed). Here, one typically uses the mid-median.

One continues (4.1) or the latter process until either convergencew is nearly obtained or else the norm of the matrix $Z^{(g)}$ seems to have reached its minimal value.

GENERALIZED POLISH. We now introduce the following analogous calculation for the general minimization problem (2.1). Here $p \geqslant 1$.

Start out by selecting a fixed infinite sequence of integers $\{r_n; n \geqslant 1\}$ taking values in $\{1, \dots , M\}$, with the property that there exists a possibly large integer N, such that, for all $n \geqslant n_o$, each $r = 1, \dots , M$ occurs in the finite subsequence $\{r_n, r_{n+1}, \dots , r_{n+N}\}$. For instance, if $r_n \equiv n \pmod M$ then $N = M - 1$.

Now calculate recursively, for $n = 1, 2, \dots ,$

(4.2) $$z^{(n)}(x) = z^{(n-1)}(x) - s_n g_{r_n}(x), \qquad (x \in X)$$

where

(4.3) $$s_n = \text{mean}_{p,r}\{z^{(n-1)}(x)\},$$

(as defined in (3.4)). Therefore,

(4.4) $$\Sigma_{x \in X} \nu(x)|z^{(n)}(x)|^p = \min_s \Sigma_{x \in X} \nu(x)|z^{(n-1)}(x) - s g_{r_n}(x)|^p,$$

while s_n attains the latter minimum. If $p = 1$, this means that s_n is a median (to be specified) of the numbers $z^{(n-1)}(x)/g_{r_n}(x)$ with $g_{r_n} \neq 0$, relative to the weights $\omega(x)|g_{r_n}(x)|$.

As an illustration, in example (2.4) each index r_n is of the form $r_n = (t_n, y_n)$ with $t_n \in T$ and $y_n \in Y_{t_n}$. In the above generalization, the numbers $z^{(n)}(x)$ would be derived from the $z^{(n-1)}(x)$ by adjusting only the numbers with $x \in L_{t_n}(y_n)$. For instance, if $p = 1$ and $\omega(x) = 1$ then one substracts from the $z^{(n-1)}(x)$ with $x \in L_{t_n}(y_n)$ a fixed median of this same set of numbers.

In the remaining part of the present section, we consider the general minimization problem (2.1) with $p > 1$. Let $\beta^* = (\beta^*_1, \dots , \beta^*_M)$ denote the unique vector attaining the minimum (2.1), see Lemma 1, and let

(4.5) $$\mathbf{Z}^* = \{z^*(x) = z^{(o)}(x) - \Sigma_{r=1}^{M} \beta^*_r g_r(x); x \in X\}$$

denote the corresponding optimal 'matrix' of residuals, the unique adjustment having the smallest possible L_p-norm.

THEOREM 1. *Suppose $p > 1$. Then the sequence $\mathbf{Z}^{(n)} = \{z^{(n)}(x); x \in X\}$ $(n = 0, 1, 2, \dots)$ defined by the above generalized polish always converges to the optimal matrix $\mathbf{Z}^*$.*

Remark. The special case $p = 2$ can also be deduced from results due to Amemiya and Ando (1965) concerning projections in a Hilbert space. Smith, Solmon and Wagner ((1977) p. 1229) even established an exponential rate of convergence when $p = 2$ and $\{r_n\}$ is periodic.

Note from (4.2) that $\mathbf{Z}^{(n)} = \{z^{(n)}(x); x \in X\}$ is of the form

(4.6) $$z^{(n)}(x) = z^{(o)}(x) - \Sigma_{r=1}^{M} \beta_{n,r} g_r(x).$$

Here, the $\beta_{n,r}$ are unique since $g_1, \dots , g_M$ are linearly independent. Moreover, from (4.4), in going from the $z^{(n-1)}(x)$ to the $z^{(n)}(x)$, one takes $\beta_{n,r} = \beta_{n-1,r}$ for all $r \neq r_n$, while β_{n,r_n} is chosen so as to minimize the norm of $\mathbf{Z}^{(n)}$.

A GENERALIZATION. One can even replace the linear combination $\beta_1 g_1(x) + \ldots + \beta_M g_M(x)$ by a general continuous function $g(x; \beta_1, \beta_2, \ldots, \beta_M)$, thus the residuals take the form

$$d(x) = d(x; \beta_1, \ldots, \beta_M) = z^{(o)}(x) - g(x; \beta_1, \ldots, \beta_M).$$

The main goal would normally be to minimize some explicitly given expression in terms of these residuals, for instance, of the type $\Sigma_{x \in X}\omega(x)|d(x)|^p$. It measures the 'norm' or 'size' of the present set of residuals. Keeping the data $z^{(o)}(x)(x \in X)$ fixed, this expression becomes a known function $F(\beta_1, \ldots, \beta_M)$.

We will *assume* that this resulting function F is a *strictly convex* function on R^M of class C^1 and such that $F(\beta)$ tends to $+\infty$ as $\beta = (\beta_1, \ldots, \beta_M)$ tends to infinity. In the above special case with $p > 1$, the function F would be given by the right hand side of (2.1) and does indeed have the above properties.

The generalized polish starts again with $\mathbf{Z}^{(o)} = \{z^{(o)}(x); x \in X\}$. After $n - 1$ steps, one obtains a set of residuals

$$z^{(n-1)}(x) = z^{(o)}(x) - g(x; \beta_{n-1,1}, \ldots, \beta_{n-1,M}) \tag{4.7}$$

($x \in X$) and having its 'size' equal to $F(\beta_{n-1,1}, \ldots, \beta_{n-1,M})$. We now derive that $z^{(n)}(x)$ from the $z^{(n-1)}(x)$ by choosing $\beta_{n,r} = \beta_{n-1,r}$ for all $r \neq r_n$ and choosing β_{n,r_n} such that $F(\beta_{n,1}, \ldots, \beta_{n,M})$ is as small as possible.

It follows from the above assumptions that $\beta_{n,r}$ is uniquely determined. We will assume as before that there exists a positive integer N such that, for all $n \geqslant n_o$, each $r = 1, \ldots, M$ occurs among $\{r_n, r_{n+1}, \ldots, r_{n+N}\}$. The following result generalizes Theorem 1. Here $b_n = (\beta_{n,1}, \ldots, \beta_{n,N}), (b_o = 0)$.

THEOREM 2. *The sequence $\{b_n\}$ converges to the unique point $\beta^* = (\beta^*_1, \ldots, \beta^*_M)$ where F assumes its smallest value. Hence, for each $x \in X$, $z^{(n)}(x)$ converges to*

$$z^*(x) = z^{(o)}(x) - g(x; \beta^*_1, \ldots, \beta^*_M). \tag{4.8}$$

Proof. It follows from the properties of F and the choice of β_{n,r_n} that, for all $n \geqslant 1$,

$$(\delta/\delta\beta_{r_n})F(\beta) = 0 \quad \text{at } \beta = b_n, \tag{4.9}$$

and further

$$F(b_n) \leqslant F((b_{n-1} + b_n)/2) \leqslant F(b_{n-1}). \tag{4.10}$$

Hence, $\lim_n F(b_n)$ exists and all the b_n belong to the compact set $K = \{\beta \in R^M: F(\beta) \leqslant F(b_o)\}$. We further claim that

$$\lim_n (n_n - b_{n-1}) = 0. \tag{4.11}$$

For, suppose not. Then there would exist integers $1 < n_1 < n_2 < \ldots$ such that $b_{n_k} \to u$ and $b_{n_k-1} \to v$ as $k \to \infty$, with $u, v \in K$ and $u \neq v$. Then $F(u) = \lim_n F(b_n) = F(v)$ while, from (4.10), $F(u) \leqslant F((u+v)/2) \leqslant F(v)$. Hence, all the equality signs hold here which contradicts the strict convexity of F.

Let $\{b_{n_k}\}$ be a convergent subsequence of $\{b_n\}$ with limit b^*. From (4.11), one also has for each fixed m that $\lim_k b_{n_k+m} = b^*$. And we conclude from (4.9) that

(4.12) $$(\delta/\delta\beta_r)F(\beta) = 0 \quad (r = 1, \dots, M) \quad \text{at } \beta = b^*.$$

After all, F is of class C^1 while, for each k, one has that each $r = 1, \dots, M$ occurs among $r_{n_k+m}(m = 0, 1, \dots, N)$. However, property (4.12) uniquely determines the single point β^* where F takes its smallest value, thus $b^* = \beta^*$. This proves that $\{b_n\}$ converges to β^*. The last assertion follows from (4.7) and (4.8). □

5. Optimal Sign Patterns. In the remaining part of the paper, we restrict ourselves to the case $p = 1$ of the general minimization problem (2.1). Let $\mathbf{Z} = \{z(x); x \in X\}$ be a fixed 'matrix'; it usually arises as a matrix of residuals. We associate to $\mathbf{Z}$ the function

(5.1) $$F(\beta) = \Sigma_{x\in X}\omega(x)|z(x) - \Sigma_{r=1}^{M}\beta_r g_r(x)|,$$

where $\beta = (\beta_1, \dots, \beta_M)$. Clearly, F is a piecewise linear, continuous and convex function on $\mathcal{R}^M$ tending to $-\infty$ as β tends to infinity. Let K_Z denote the non-empty compact polyhedral convex set of points $\beta \in \mathcal{R}^M$ where F assumes its smallest value.

The matrix $\mathbf{Z}$ will be said to be *optimal* if its norm cannot be reduced by substracting from $z(x)$ a linear combination of the $g_r(x)$ $(r = 1, \dots, M)$. Equivalently, $0 \in K_Z$. Similarly,

(5.2) $$Z_\beta = z_\beta(x) - \Sigma_r\beta_r g_r(x); x \in X$$

is optimal if and only if $\beta \in K_Z$. It is known, see Kennedy and Gentle (1980) p. 515, and can easily be proved by an induction on M, that at least one of these optimal matrices $\mathbf{Z}_\beta$ has M or more of its components $z_\beta(x)$ equal to 0.

We will derive several different criteria which are necessary and sufficient for $\mathbf{Z}$ to be optimal. An important role is played by the set

(5.4) $$D = \{x \in X: z(x) = 0\} = \{x \in X: \zeta(x) = 1\}$$

of zero locations. Here, $\zeta(x) = 1 - |\sigma(x)|$ with

(5.4) $$\sigma(x) = \operatorname{sgn} z(x), \qquad (x \in X).$$

We further associate to $\mathbf{Z}$ the set of M constants

(5.5) $$u_r = \Sigma_{x\in X}\theta(z(x))\omega(x)g_r(x), \qquad (r = 1, \dots, M).$$

Here, $\theta(z(x)) = \sigma(x) + \zeta(x)$ equals -1 if $z(x)$ is negative and $+1$, otherwise. We always define $v_+ = \max(0, v)$, $v_- = \max(0, -v)$.

LEMMA 2. *In order that $\mathbf{Z} = \{z(x); x \in X\}$ be optimal, it is necessary and sufficient that*

(5.6) $$u_1\beta_1 + \dots + u_M\beta_M \leq \Sigma_{x\in D}2\omega(x)[\beta_1 g_1(x) + \dots + \beta_M g_M(x)]_+$$

holds for each choice of the real constants $\beta_1, \dots, \beta_m$. In fact, given $\beta \in \mathcal{R}^M$, (5.6) fails to hold if and only if $F(\lambda\beta) < F(0)$ for all sufficiently small scalars $\lambda > 0$.

Proof. Optimality of $\mathbf{Z}$ means that the associated convex function F defined by (5.1) has the origin as a local minimum along *each* half line through the origin of $\mathcal{R}^M$; for, this implies global minimality. Equivalently, one must have for all β that

(5.7) $$\lim_{\lambda\downarrow 0}(F(\lambda\beta) - F(0))/\lambda \geq 0.$$

Given $\beta \in R^M$, we have from (5.1) that (5.7) is equivalent to

(5.8) $$\Sigma_{x\notin D}\sigma(x)\omega(x)\,\Sigma_{r=1}^{M}\beta_r g_r(x) \leq \Sigma_{x\in D}\omega(x)|\Sigma_{r=1}^{M}\beta_r g_r(x)|.$$

Adding $\Sigma_{x\in D}\omega(x)\Sigma_{r=1}^{M}\beta_r g_r(x)$ to both sides of (5.8), one obtains condition (5.6). This proves Lemma 2. □

COROLLARY. Whether or not the matrix $\mathbf{Z} = \{z(x); x \in X\}$ is optimal depends only on the associated sign pattern $\{\sigma(x); x \in X\}$ and not on the values $z(x)$ themselves. Therefore, we will also speak of optimal and non-optimal sign patterns.

A sufficient condition for optimality is that $u_r = 0$ for all r. If $z(x) \neq 0$ for all $x \in X$, (that is, if D is empty) then the latter condition is also necessary.

It is convenient to introduce the set

(5.9) $\quad B = B(\sigma) = \{\beta\colon \Sigma_{x \in D} \omega(x) |\, \Sigma_{r=1}^{M} \beta_r g_r(x)| < \Sigma_{x \notin D} \sigma(x) \omega(x) \Sigma_{r=1}^{M} \beta_r g_r(x)\}$,

where $\beta = (\beta_1, \dots, \beta_M)$. It is the set for which the (equivalent) conditions (5.6), (5.7), (5.8) fail to hold. That is, B is precisely the set of direction in which F is strictly decreasing when starting at the origin. Also note that B is an open convex cone; (naturally, $0 \notin B$).

The set B depends only on the associated sign pattern. A sign pattern σ is optimal or not depending on whether $B(\sigma)$ is empty or non-empty, respectively.

It is useful to introduce the following (quasi) partial ordering among sign patterns over X. Namely, let us say that the sign pattern $\sigma = \{\sigma(x); x \in X\}$ is smaller than the sign pattern $\tau = \{\tau(x); x \in X\}$ (and we write $\sigma \prec \tau$) if either $\sigma = \tau$ or else τ can be obtained from σ by replacing one or more elements $\sigma(x) = 0$ by either -1 or $+1$. Such an ordering is clearly transitive.

If $\sigma \prec \tau$ then the lower sign pattern σ has a larger set D of zero locations. Moreover, condition (5.8) for τ is easily seen to imply the analogous condition for σ. Therefore,

(5.10) $\qquad$ if $\quad \sigma \prec \tau \quad$ then $B(\sigma) \prec B(\tau)$.

THEOREM 3. *If a sign pattern* $\sigma(x); x \in X$ *is optimal then it remains optimal when one or more non-zero elements* ($-$ *or* $+1$) *are replaced by* 0; (*'the more zeros the better'*).

Proof. What is asserted is that $\sigma \prec \tau$, together with the optimality of τ, (that is, $B(\tau)$ is empty), implies the optimality of σ, (that is, $B(\sigma)$ is empty). And this is evident from (5.10).

Note that, relative to the partial ordering on hand, the optimal sign patterns form a lower set while the non-optimal patterns form the (complementary) upper set. In order to be able to recognize a non-optimal pattern, it would be sufficient to have a list of all *minimal* non-optimal patterns σ. For each such non-optimal σ, the set $B(\sigma)$ is non-empty and it would be useful to list not only σ itself but also one or more members β of the associated set $B(\sigma)$. Namely, we know from (5.10) that β also belongs to each set $B(\tau)$ with $\sigma \prec \tau$ and supplies a method for improving any matrix whose ± 1 pattern contains that of σ (ignoring zeros).

Example. Consider the problem of minimizing $\Sigma_{i=1}^{3} \Sigma_{j=1}^{3} |z_{ij}^{(o)} - \alpha_i - \beta_j|$. It is not hard to show that a set of residuals $\mathbf{Z} = (z_{ij})$ which is invariant under median polish (a so-called EMP) is non-optimal if and only if $\sigma = (\sigma_{ij} = \operatorname{sgn} z_{ij})$ shows a subpattern of the type $\sigma_{11} = +1$; $\sigma_{22} = +1$; $\sigma_{33} = -1$. More precisely, Z is non-optimal if and only if, for some permutation (j_1, j_2, j_3) of $(1, 2, 3)$ the values $\eta_i = \sigma_{i, j_i}$ $(i = 1, 2, 3)$ are all non-zero but not all of the same sign. If for instance $\sigma_{13} = -1$; $\sigma_{21} = +1$; $\sigma_{32} = -1$ then the norm of $\mathbf{Z}$ can be reduced by adding a small positive number ϵ to the second and third column, and simultaneously substracting ϵ from the second row.

Comments. A typical algorithm for solving (2.1) (with $p = 1$), that is, for minimizing a function F as in (5.1), would first check whether or not the matrix $\mathbf{Z}$ on hand is optimal. If it is not then one locates somehow an element $\beta \in B$ (known to be non-empty), that is, a direction in which F is strictly decreasing when starting at 0. One may as well proceed in that same direction until a minimum of $F(\lambda\beta)$ is reached. This dictates the choice

(5.11) $\lambda^\circ = \text{med}\{z(x)/h(x)\colon \omega(x)|h(x)|)\}$, where $h(x) = \Sigma_{r=1}^{M}\beta_r g_r(x)$

and leads to the new matrix

$$\mathbf{Z}' = \{z'(x) = z(x) - \Sigma_{i=1}^{M}\lambda^\circ\beta_r g_r(x); x \in X\}.$$

It has a strictly smaller norm. One next checks whether $\mathbf{Z}'$ is optimal and so on.

Median polish as described by (4.2), (4.3) (with $p = 1$) is of a similar type except that one only allows motions parallel to one of the m coordinate axes. We will call $\mathbf{Z}$ an end-product of median polish (EMP) if a single motion of that type does not improve the norm, (though two subsequent motions of that type might). In view of (5.11), this is equivalent to

(5.12) $\text{med}\{z(x)/g_r(x)\colon \omega(x)|g_r(x)|\} = 0, \qquad (r = 1, \ldots, M).$

We will call $\mathbf{Z}$ and endproduct of mid-median polish (EMMP) if (5.12) holds with 'median' replaced by 'mid-median'. Condition (5.12) means precisely that the (equivalent) conditions (5.6), (5.7), (5.8) hold on choosing $\beta_r = +1$ or -1 and $\beta_s = 0$, otherwise, $(r = 1, \ldots, M)$. Thus, by (5.6), (5.12) is equivalent to

(5.13) $-\Sigma_{x\in D}2\omega(x)g_r(x)_- \leq u_r \leq \Sigma_{x\in D}2\omega(x)g_r(x)_+, (r = 1, \ldots, M).$

6. Additional Criteria for Optimality.

THEOREM 4. *In order that* $\mathbf{Z} = z(x); x \in X$ *be optimal, it is necessary and sufficient that numbers* $w(x)$ $(x \in X)$ *exist with*

(6.1)
$$w(x) = \sigma(x)\omega(x) \quad \text{if } x \notin D;$$
$$|w(x)| \leq \omega(x) \quad \text{if } x \in D,$$

and further

(6.2) $\Sigma_{x\in X}w(x)g_r(x) = 0$ *for all* $r = 1, \ldots, M$.

An equivalent condition is that the following moment problem has a solution. Namely, there must exist numbers $W(x)$ $(x \in D)$ *with*

(6.3) $0 \leq W(x) \leq 2\omega(x)$ *for each* $x \in D$

and

(6.4) $\Sigma_{x\in D}W(x)g_r(x) = u_r$ *for all* $r = 1, \ldots, M$.

***Here, D is defined by* (5.4) *and* u_r *is defined by* (5.5).**

Proof. That the two conditions are equivalent is seen by letting $W(x) = \omega(x) - w(x)$, for $x \in D$. From (6.1) and the definitions of $\sigma(x)$ and $\zeta(x)$, (6.2) can be written as

$$\Sigma_{x\in X}\sigma(x)\omega(x)g_r(x) + \Sigma_{x\in X}\zeta(x)(\omega(x) - W(x))g_r(x) = 0.$$

In view of (5.5), this is equivalent to (6.4). □

As is obvious and well-known, see Wagner (1959), the minimization problem (2.1) (with $p = 1$) can be regarded as a linear programming problem to

(I) Minimize: $\Sigma_{x\in X}\omega(x)(p(x) + q(x))$,

subject to the conditions

$$z(x) - \Sigma_{r=1}^{M}\beta_r g_r(x) = p(x) - q(x); p(x) \geq 0; q(x) \geq 0, (x \in X).$$

The variables β_r are real-valued.

In order that $\mathbf{Z}$ be optimal, it is necessary and sufficient that the minimum on hand be

equal to $\Sigma_{x\epsilon X}\omega(x)|z(x)|$, (corresponding to $p(x) = z(x)_+$; $q(x) = z(x)_-$ and $\beta_r = 0$ as an optimal solution).

The dual of problem (I) is to

(II) Maximize: $\Sigma_{x\epsilon X}z(x)w(x)$, subject to the conditions (6.2) and $-\omega(x) \leqslant w(x) \leqslant \omega(x)$.

Each problem has feasible solutions. Thus, optimal solutions exist for each, and the minimum in (I) equals the maximum in (II). Let $w(x)$ ($x \in X$) be a feasible solution of (II). Then

$$\Sigma_{x\epsilon X}z(x)w(x) \leqslant \Sigma_{x\epsilon X}|w(x)z(x)| \leqslant \Sigma_{x\epsilon X}\omega(x)|z(x)|.$$

In order that $\mathbf{Z}$ be optimal and, simultaneously, $w(x)$ be optimal for (II), it is necessary and sufficient that the equality signs hold here. This means that $w(x) = \sigma(x)\omega(x)$ each time that $z(x) \neq 0$. Since optimal solutions for (II) always exist, this proves Theorem 4. □

Remark 1. Note that the conditions (6.1), (6.2) depend only on the sign pattern $\sigma = \{\sigma(x); x \in X\}$ and that it becomes weaker (less demanding) on replacing some elements $+1$ or -1 by 0. This yields a second proof of Theorem 3.

Remark 2. An other proof of Theorem 4 would be as follows. Consider the moment problem (6.3), (6.4). It requires the existence of a (nonnegative) measure μ on D satisfying

$$\int g_r(\xi)\, \mu(d\xi) = u_r; \quad \int \delta_\xi^x\, \mu(d\xi) \leqslant 2\omega(x);$$

($r = 1, \ldots, M$; $x \in D$). As is well-known, see Kemperman (1983), since D is finite such a measure exists if and only if $\rho(x) \geqslant 0$ ($x \in D$) and

$$\Sigma_{r=1}^M \beta_r g_r(\xi) \leqslant \Sigma_{x\epsilon D}\rho(x)\delta_\xi^x \, (\xi \in D) \;\text{ imply }\; \Sigma_{r=1}^M \beta_r u_r \leqslant \Sigma_{x\epsilon D}2\rho(x)\omega(x).$$

One might as well choose $\rho(x) = [\beta_1 g_1(x) + \ldots + \beta_M g_m(x)]_+$ and then one obtains exactly condition (5.6) of Lemma 2, which is indeed equivalent to the optimality of $\mathbf{Z}$.

7. Optimality for a General Layout. Let us now apply the above results to the general layout problem as in (2.4). For simplicity, we assume that $\omega(x) = 1$ ($x \in X$), thus, one is interested in minimizing

$$S = \Sigma_{x\epsilon X}|z^{(o)}(x) - \Sigma_{t\epsilon T}\beta_t(\phi_t(x))| \tag{7.1}$$

by a suitable choice of the regression parameters $\beta_t(y)$. Note that the role of the index $r = 1, \ldots, M$ is presently taken over by the pairs (t, y) with $t \in T$ and $y \in Y_t$. The total number M of such pairs is often large. The function $g_r = g_{t,y}$ on X is presently as in (2.5).

Let $\mathbf{Z} = \{z(x); x \in X\}$ be a fixed matrix, usually arising as a sequence of residuals

$$z(x) = z^{(o)}(x) - \Sigma_{t\epsilon T}\beta_t(\phi_t(x)) \qquad (x \in X), \tag{7.2}$$

with $\mathbf{Z}^{(o)} = z^{(o)}(x); x \in X)$ as the original data. We like to know in how far $\mathbf{Z}$ is optimal.

The number of elements x in the layer $L_t(y) = \{x \in X: \phi_t(x) = y\}$ will be denoted as $n_t(y)$. Let further $n_t^+(y)$, $n_t^o(y)$ and $n_t^-(y)$, respectively, denote the number of elements $x \in L_t(y)$ such that $\sigma(x) = \operatorname{sgn} z(x) = +1, 0$ or -1 respectively. Put

$$u_t(y) = n_t^+(y) + n_t^o(y) - n_t^-(y) = n_t(y) - 2n_t^-(y), \tag{7.3}$$

($t \in T$; $y \in Y_t$). As usual, $D = \{x \in X: z(x) = 0\}$. Theorem 4 yields the following two criteria for optimality.

Criterion I. In order that $\mathbf{Z}$ be optimal, it is necessary and sufficient that there exist numbers $w(x)$ ($x \in X$) such that

$$\Sigma_{\phi_t(x)=y}\, w(x) = 0, \quad \text{for all } t \in T; \text{ all } y \in Y_t, \tag{7.4}$$

and

$$w(x) = +1 \quad \text{if } z(x) > 0;$$
(7.5) $$w(x) = -1 \quad \text{if } z(x) < 0;$$
$$-1 \leq w(x) = \leq +1 \quad \text{if } z(x) = 0.$$

Remark. In certain cases, such as in the two-way layout, one may add the condition that $w(x) \in \{-1, 0, +1\}$ when $x \in D$. If moreover $z(x)$ is of the form (7.2) and the original data $z^{(o)}(x)$ are integers then it follows that $\mathbf{Z}$ has an integral L_1-norm as soon as it is optimal; (thus, a residual matrix with non-integral norm cannot be optimal).

Namely, as follows from the proof of Theorem 4, (7.4) and (7.5) imply that the norm of $\mathbf{Z}$ is equal to

$$\Sigma_{x\in X} w(x)z(x) = \Sigma_{x\in X} w(x)z^{(o)}(x).$$

Criterion II. In order that $\mathbf{Z}$ be optimal, it is necessary and sufficient that there exist numbers $W(x)$ $(x \in D)$ such that

(7.6) $$\Sigma\{W(x){:}\ x \in D; \phi_t(x) = y\} = u_t(y) \quad \text{if } t \in T; y \in Y_t$$

and that further $0 \leq W(x) \leq 2$ for each $x \in D$.

One may describe condition II as requiring the existence of a measure μ on D, having at most a mass 2 at each point of D, and such that, for each $t \in T$, the ϕ_t-projection of μ onto Y_t is precisely equal to the signed measure μ_t on Y having a mass $u_t(y)$ at each $y \in Y_t$. Note that the total algebraic mass of μ_t equals

(7.7) $$Q = \Sigma_{y\in Y_t} u_t(y) = N^+ + N^o - N^- = N - 2N^-,$$

which is independent of $t \in T$. Here, N denotes the number of elements in X while N^+, N^o, N^-, respectively, denotes the number of $x \in X$ with $\sigma(x) = \operatorname{sgn} z(x) = +1$, 0 or -1, respectively.

Obviously, the required measure μ can only exist when

(7.8) $$0 \leq u_t(x) \leq 2n_t^o(y), \quad (t \in T; y \in Y_t).$$

This is equivalent to

(7.9) $$n_t^-(y) \leq n_t(y)/2 \quad \text{and} \quad n_t^+(y) \leq n_t(y)/2, (t \in T; y \in Y_t).$$

In fact, (7.9) is precisely the condition the $\mathbf{Z}$ be an EMP. Equivalently, that for4 all $t \in T$ and $y \in Y_t$ the set of $n_t(y)$ numbers $z(x)$ with $x \in L_t(y)$ (each with weight 1) has 0 as a median. Which is exactly the condition which median polish tries to attain.

It is very easy to check condition (7.9). Thus, our main problem is to decide whether $\mathbf{Z}$ is optimal in a situation where (7.9) is true. In particular, the above given marginal measures μ_t are all nonnegative.

Lemma 2 easily yields the following criterion.

Criterion III. In order that $\mathbf{Z}$ be optimal, it is necessary and sufficient that

(7.10) $$\Sigma_{t\in T}\Sigma_{y\in Y_t} u_t(y)\beta_t(y) \leq \Sigma_{x\in D}\, 2[\Sigma_{t\in T}\beta_t(\phi_t(x))]_+$$

holds for each choice of M real numbers $\beta_t(y)$, $(t \in T; y \in Y_t)$.

The set B defined by (5.9) presently takes the form

(7.11) $$B = \{\beta{:}\ \Sigma_{x\in D}|\Sigma_{t\in T}\beta_t(\phi_t(x))| < \Sigma_{x\notin D}\sigma(x)\,\Sigma_{t\in T}\beta_t(\phi_t(x))\}.$$

Here, β stands for the set of M numbers $\beta_t(y)$ $(t \in T; y \in Y_t)$. The matrix $\mathbf{Z}$ is optimal if and only if B is empty. If B is non-empty then each $\beta \in B$ supplies an explicit way of reducing the norm of $\mathbf{Z}$, see (5.11).

8. Optimality for a Two-Way Layout. Here, we only consider the case of a two-way

$m \times n$ layout ($m \geq 2$; $n \geq 2$) with a single observation $z^{(o)}(x) = z_{ij}^{(o)}$ in cell $x = (i, j)$ and weights $\omega(x) = 1$. One wants to minimize

(8.1) $$S = \Sigma_{i,j}|z_{ij}^{(o)} - \alpha_i - \beta_j|$$

by a suitable choice of the $M = m + n$ numbers $\alpha_i = \beta_1(i)$ and $\beta_j = \beta_2(j)$.

Unspecified indices i and j run through $Y_1 = \{1, \dots, m\}$ and $Y_2 = \{1, \dots, n\}$, respectively. Presently, we have $X = Y_1 \times Y_2$ while $T = \{1, 2\}$ and $\phi_1(x) = i$; $\phi_2(x) = j$ when $x = (i, j)$.

Having chosen the numbers α_i and β_j (at a particular stage of the calculation), one is confronted with the problem whether or not the matrix $\mathbf{Z}$ of residuals is optimal, in the sense that its norm cannot be further reduced.

Let $\mathbf{Z} = (z_{ij})$ be a fixed $m \times n$ matrix. Whether or not $\mathbf{Z}$ is optimal depends only on the sign pattern $\sigma = (\sigma_{ij})$, where $\sigma_{ij} = \operatorname{sgn} z_{ij}$. The number of elements $\sigma_{ij} = -1, 0, +1$, respectively, in the i-th row of σ will be denoted as $n_1^-(i)$, $n_1^o(i)$, $n_1^+(i)$, respectively; similarly, $n_2^-(j)$, $n_2^o(j)$, $n_2^+(j)$ for the j-th column of σ.

Definition. The matrix $\mathbf{Z}$ is called an EMP (or EMMP) if 0 is a median (or mid-median, respectively) of each row and each column of $\mathbf{Z}$. Each EMMP is an EMP. In order that $\mathbf{Z}$ be an EMP it is clearly necessary and sufficient that

(8.2) $$max(n_1^-(i), n_1^+(i)) \leq n/2; \quad \max(n_2^-(j), n_2^+(j)) \leq m/2,$$

for all i and j.

An EMMP cannot have exactly one zero in a row unless n is odd. In fact, an EMMP has, for each fixed i, that $n_1^-(i) = n/2$ if and only if $n_1^+(i) = n/2$ and, similarly, for each fixed j, $n_2^+(j) = m/2$ if and only if $n_2^-(j) = m/2$. An EMP with the latter property will be called a weak EMMP.

A necessary condtion for $\mathbf{Z} = (z_{ij})$ to be optimal is that it be an EMP. Though numerical calculations suggest it, we are not asserting that a median polish (mid-median polish) always leads to an EMP (EMMP). For the case where either m or n is odd, it is not difficult to show that mid-median polish does create a convergent sequence $\mathbf{Z}^{(n)} = (z_{ij}^{(n)})$ of matrices whose limit is an EMMP.

Anyway, at the end of a median polish one is typically confronted with a matrix $\mathbf{Z} = (z_{ij})$ of residuals which already is an EMP or EMMP and then the question arises how one can recognize its optimality. And if this EMP is non-optimal (as is often true) then how should one proceed in determining an optimal matrix of residuals?

CONDITION (A, B). Let $A \subset Y_1$ and $B \subset Y_2$. Let $\mathbf{G}$ and $\mathbf{H}$ denote the submatrices of $\mathbf{Z}$ defined by

(8.3) $$\mathbf{G} = (z_{ij}; i \in A, j \notin B); \quad \mathbf{H} = (z_{ij}; i \notin A, j \in B).$$

The number of positive, zero and negative elements in $\mathbf{G}$ will be denoted as $N_{\mathbf{G}}^+$, $N_{\mathbf{G}}^o$ and $N_{\mathbf{G}}^-$, respectively, while $N_{\mathbf{G}}$ denotes the total number of elements. Similarly, $N_{\mathbf{H}}^+$, $N_{\mathbf{H}}^o$ and $N_{\mathbf{H}}^-$ and $N_{\mathbf{H}}$ for $\mathbf{H}$. We will say that $\mathbf{Z}$ satisfies Condition (A, B) if

(8.4) $$N_{\mathbf{G}}^+ + N_{\mathbf{H}}^- \leq (N_{\mathbf{G}} + N_{\mathbf{H}})/2.$$

Note that Condition (A^c, B^c) requires that $N_{\mathbf{G}}^- + N_{\mathbf{H}}^+ \leq (N_{\mathbf{G}} + N_{\mathbf{H}})/2$.

THEOREM 5. *In order that $\mathbf{Z} = (z_{ij})$ be optimal, it is necessary and sufficient that Condition (A, B) holds for each choice of the subset A of Y_1 and subset B of Y_2.*

Proof. The sufficiency follows from Theorem 8 in Section 9. As to the necessity of (8.4), consider the modified matrix

$$\mathbf{Z}' = (z'_{ij} = z_{ij} - \alpha_i - \beta_j),$$

where $\alpha_i = +\lambda$ when $i \in A$; $\alpha_i = 0$ when $i \notin A$, while $\beta_j = -\lambda$ when $j \in B$; $\beta_j = 0$ when $j \notin B$. Here, λ denotes a sufficiently small positive constant. The effect of this transformation is that each element z_{ij} in **G** is decreased by λ, each element in **H** is increased by λ, while the remaining part of **Z** remains unchanged. This causes on the one hand a decrease in norm by $(N_{\mathbf{G}}^{+} + N_{\mathbf{G}}^{-})\lambda$ and on the other hand an increase in norm by

$$(N_{\mathbf{G}}^{-} + \mathrm{N}_{\mathbf{G}}^{o} + N_{\mathbf{H}}^{+} + \mathrm{N}_{\mathbf{H}}^{o})\lambda = (N_{\mathbf{G}} + N_{\mathbf{H}})\lambda - (N_{\mathbf{G}}^{+} + N_{\mathbf{H}}^{-})\lambda.$$

Hence, unless (8.4) holds the matrix $\mathbf{Z}'$ would have a strictly smaller norm than **Z** and **Z** would not be optimal. □

Remark 1. In order that **Z** be an EMP it is necessary and sufficient that condition (A, B) holds with one of the sets A, B empty and the other consisting either of a single element or else all but a single element. This in turn is equivalent to Condition (A, B) for the case where one of the two sets A, B is either empty or full. Thus, if **Z** is already known to be an EMP then one only needs to verify (8.4) for the case where neither **G** nor **H** is empty.

Remark 2. Theorem 5 suggests the following algorithm for determining an optimal set of residuals. After n steps one has a matrix $\mathbf{Z}^{(n)}$. If it satisfies all conditions (A, B) then it is optimal. If not then the above proof indicates how to arrive at a new matrix $\mathbf{Z}^{(n+1)}$ having a strictly smaller norm. It is best to choose $\lambda = \lambda_n$ in an optimal way, namely, as a median of the $N_{\mathbf{G}} + N_{\mathbf{H}}$ elements g_{ij} and $-h_{ij}$ in **G** and $-$**H**. If the original data $z_{ij}^{(o)}$ are all integers then one can attain that, for all n, also $z_{ij}^{(n)}$ and λ_n and thus $z_{ij}^{(n+1)}$ are integers. But then the norm of the matrix decreases at each step by a positive integer, hence, the process must stop after finitely many steps.

Example. Armstrong, Elam and Hultz (1977) developed a quite different algorithm. Details were given for the following 4×5 matrix

$$\mathbf{Z}^{(o)} = \begin{pmatrix} 350 & 492 & 232 & 220 & 360 \\ 392 & 428 & 253 & 241 & 385 \\ 400 & 498 & 273 & 260 & 401 \\ 320 & 390 & 264 & 240 & 300 \end{pmatrix}.$$

Their method led to the following matrix of residuals

$$\mathbf{Z} = \begin{pmatrix} 0 & 72 & 0 & 0 & 0 \\ 21 & -13 & 0 & 0 & 4 \\ 9 & 37 & 0 & -1 & 0 \\ 0 & 0 & 62 & 50 & -30 \end{pmatrix}.$$

which has norm 299 and was claimed to be optimal. Note that **Z** is an EMP but not an EMMP. Actually, **Z** does not satisfy Condition (A, B) with $A = 1, 2, 3$ and $B = 3, 4, 5$. For, then $N_{\mathbf{G}} = 6$; $N_{\mathbf{G}}^{+} = 4$; $N_{\mathbf{H}} = 3$; $N_{\mathbf{H}}^{-} = 1$ so that (8.4) is violated. This allows an improvment as usual; it is best to choose $\lambda = 9$. In this way, substracting 9 from the first 3 rows and adding 9 to the last 3 columns, one arrives at the matrix

$$\mathbf{Z}' = \begin{pmatrix} -9 & +63 & 0 & 0 & 0 \\ +12 & -22 & 0 & 0 & +4 \\ 0 & +28 & 0 & -1 & 0 \\ 0 & 0 & +71 & +59 & -21 \end{pmatrix}.$$

which has norm 290. Using any of several criteria in Sections 8 and 9, it is easily seen that $\mathbf{Z}'$ is optimal. For instance, substracting 14 from each element in the second column, one obtains a matrix $\mathbf{Z}''$ which has the same norm is is an EMMP and, thus, optimal by Theorem 6 below. This implies that also $\mathbf{Z}'$ is optimal.

Definition. The pair (m, n) of integers ≥ 2 is safe for median polish if each EMP of dimension (m, n) is optimal. Similarly, (m, n) is safe for mid-median polish if each EMMP of dimension (m, n) is optimal.

THEOREM 6. *No pair (m, n) is safe for median polish. And further the only pairs which are safe for mid-median polish are the exceptional pairs $(2, n)$; $(3, 4)$; $(4, 4)$; $(4, 5)$; $(4, 6)$ and their reflections such as $(n, 2)$.*

In fact, for these exceptional pairs (m, n) it is even true that every weak EMMP of dimension (m, n) is optimal.

Remark. In particular, the pairs $(3, 3)$; $(3, 5)$; $(6, 6)$; $(4, 8)$ are all unsafe for mid-median polish. A weak EMMP may be described as a matrix whose sign pattern (having only elements -1, 0, $+1$) is exactly an EMMP. Thus, its sign pattern indicates an EMMP but the matrix may not have the property that 0 is exactly the mid-median of each row and each column. But indeed we know from Section 5 that the sign pattern alone already determines optimality or nonoptimality.

Proof. In order to prove the stated 'unsafety', it suffices to construct an EMP or EMMP which violates one of the conditions (8.4) and, hence, is not optimal. Consider an $m \times n$ matrix $\mathbf{Z}$ which after a suitable permutation of rows and columns takes the form

$$\mathbf{Z} = (z_{ij}) = \begin{pmatrix} \mathbf{G} & \mathbf{K} \\ \mathbf{L} & \mathbf{H} \end{pmatrix} \tag{8.5}$$

Here, $\mathbf{G}$ and $\mathbf{H}$ are of dimension $m_1 \times n_1$ and $m_2 \times n_2$, respectively, $(m_1 + m_2 = m;\ n_1 + n_2 = n)$, while $\mathbf{K}$ and $\mathbf{L}$ are *zero* matrices of dimension $m_1 \times n_2$ and $m_2 \times n_1$, respectively. From Theorem 5, the matrix $\mathbf{Z}$ is non-optimal as soon as (8.4) is false.

In fact, we will choose all the elements of $\mathbf{G}$ as (strictly) positive. Let further the elements of $\mathbf{H}$ be either negative or 0, in an alternating (checkerboard type) fashion, starting with a negative element in the left upper corner of $\mathbf{H}$. In this situation one has $N_{\mathbf{G}}^{+} = N_{\mathbf{G}}$ and $N_{\mathbf{H}}^{-} \geq N_{\mathbf{H}}/2$, hence, the difference between the left and right hand sides of (8.4) is at least $N_{\mathbf{G}}/2 = m_1 n_1/2$. Thus, $\mathbf{Z}$ is non-optimal as soon as m_1 and n_1 are positive. Moreover, $\mathbf{Z}$ is easily seen to be an EMP provided

$$1 \leq m_1 \leq m_2; \quad 1 \leq n_1 \leq n_2. \tag{8.6}$$

That is, $1 \leq m_1 \leq m/2$ and $1 \leq n_1 \leq n/2$. Since such a choice of m_1 and n_1 is always possible, this proves the first assertion of Theorem 6. By the way, the leeway $m_1 n_1/2$ above indicates that there are usually many mild modifications of $\mathbf{Z}$ which are also non-optimal EMP's.

The matrix $\mathbf{Z}$ in (8.5) is even an EMMP provided

$$m_1 < m_2; \quad m_1 \geq 2 \text{ if } m_2 \text{ odd}; \quad n_1 < n_2; \quad n_1 \geq 2 \text{ if } n_2 \text{ odd}, \tag{8.7}$$

and further all non-zero elements z_{ij} equal -1 or 1.
For, then 0 is a mid-median of each row and each column; (if m_2 is odd then some columns of $\mathbf{H}$ have an excess of negative elements; hence, if $m_1 = 1$ then that column would have 0 as a median but not as a mid-median; similarly if n_2 is odd). Therefore, a sufficient cond-

tion for (m, n) to be 'unsafe' for mid-median polish is that (8.6) can be strengthened to (8.7).

Choosing $m_1 = n_1 = 2$, this is true if both $m \geqslant 5$ and $n \geqslant 5$. Letting $m_1 = n_1 = 1$, it also holds whem m and n are odd, $m \geqslant 3$ and $n \geqslant 3$. Letting $m_1 = 1$ and $n_1 = 2$, this approach also covers the pairs $(3, n)$ with $n \geqslant 5$. Similarly for the pairs $(m, 3)$ with $m \geqslant 5$.

It only remains to consider the pairs $(4, n)$ with $n \geq 7$, (the pairs $(n, 4)$ having the same character). Since one can enlarge the matrix $\mathbf{Z}$ by adding pairs of columns of the type $\left(\begin{smallmatrix} +1 & -1 & +1 & -1 \\ -1 & +1 & -1 & +1 \end{smallmatrix}\right)^T$, ($T$ for transpose), one easily sees that it suffices to construct for the (m, n) pairs (4, 7) and (4, 8) an EMMP of the type (8.5) with $m_1 = 2$ and $n_2 = 4$ and such that (8.4) is false; thus, we do not require that $\mathbf{L}$ and $\mathbf{K}$ are zero matrices. Examples with $m = 4, n = 7$ are:

$$\left(\begin{array}{ccc|cccc} + & + & 0 & + & 0 & 0 & 0 \\ 0 & 0 & + & + & 0 & 0 & 0 \\ \hline 0 & 0 & 0 & - & - & - & 0 \\ 0 & 0 & 0 & - & 0 & 0 & - \end{array}\right) \qquad \left(\begin{array}{ccc|cccc} + & + & - & + & - & 0 & 0 \\ - & - & + & + & + & 0 & 0 \\ \hline + & 0 & 0 & - & - & - & + \\ - & 0 & 0 & - & + & + & - \end{array}\right)$$

where $+$ and $-$ may, for instance be interpreted as $+1$ and -1, respectively. Examples with $m = 4$ and $n = 8$ are:

$$\left(\begin{array}{cccc|cccc} + & + & + & 0 & 0 & 0 & 0 & 0 \\ + & 0 & 0 & + & 0 & 0 & 0 & 0 \\ \hline - & 0 & 0 & 0 & - & - & 0 & 0 \\ - & 0 & 0 & 0 & 0 & 0 & - & - \end{array}\right) \qquad \left(\begin{array}{cccc|cccc} + & + & + & - & - & 0 & 0 & 0 \\ + & - & - & + & + & 0 & 0 & 0 \\ \hline - & 0 & 0 & 0 & - & - & + & + \\ - & 0 & 0 & 0 & + & + & - & - \end{array}\right).$$

Finally, let $\mathbf{Z}$ be a weak EMMP of dimension (m, n). It only remains to show that $\mathbf{Z}$ is optimal for the special dimensions $(2, n)$; (3, 4); (4, 4); (4, 5) and (4, 6). This will be done by verifying that in these cases $\mathbf{Z}$ satisfies all conditions (A, B), see Theorem 5. Rearranging rows and columns, one may assume that the associated matrices $\mathbf{G}$ and $\mathbf{H}$ as in (8.3) are located as in (8.5). In view of Remark 1 following Theorem 5, one may assume that the m_t and n_t $(t = 1, 2)$ are positive integers. For brevity, let $p = N^+_{\mathbf{G}}$ and $q = N^-_{\mathbf{H}}$. It must be shown that

$$p + q \leqslant (N_{\mathbf{G}} + N_{\mathbf{H}})/2 = (m_1 n_1 + m_2 n_2)/2, \tag{8.8}$$

whenever $\mathbf{Z}$ is a weak EMMP having one of the above dimensions. Thus, $\mathbf{Z}$ is an EMP such that each row containing exactly $n/2$ positive (negative) elements has the property that all the other $n/2$ elements in that row are negative (positive); equivalently such a row contains no zeros. Similarly for columns.

The case $(2, n)$ is particularly easy. Here $m_1 = m_2 = 1$ while $N_{\mathbf{G}} = n_1$ and $N_{\mathbf{H}} = n_2$ with $n_1 + n_2 = n$. It must be shown that $p + q \leqslant n/2$, with p as the number of positive elements among the z_{1j} with $1 \leqslant j \leqslant n_1$ and q as the number of negative elements among the z_{2j} with $n_1 < j \leq n$. But z_{2j} is negative if and only if z_{1j} is positive. Hence, $p + q$ equals the number of positive elements in the first row and therefore $p + q \leq n/2$.

The case (4, 4) can be reduced to the case (6, 4), while the case (3, 4) can be reduced to the case (5, 4). Namely, if $\mathbf{Z}$ is a weak EMMP of dimension $(m, 4)$ then adding two rows of the type $\left(\begin{smallmatrix} + & - & + & - \\ - & + & - & + \end{smallmatrix}\right)$, one obtains a weak EMMP of dimension $(m + 2, 4)$. And if this new matrix $\mathbf{Z}'$ is optimal then so is the original matrix $\mathbf{Z}$. For, it is easily seen that condition (8.8) for $\mathbf{Z}$ follows from the analogous condition for $\mathbf{Z}'$; (the implication would also be an easy consequence of Criterion I of Section 7). It remains to show (8.8) for the cases (4, 5); (4, 6).

Let us do the case where $\mathbf{Z}$ is of dimension (4, 6). Consider for instance the situation that $m_1 = 2$; $m_2 = 2$; $n_1 = 3$ and $n_2 = 3$. It must be shown that $p + q \leqslant 6$. Suppose z_{11},

z_{12}, z_{13} are all positive, thus $p \geqslant 3$ and z_{14}, z_{15}, z_{16} must be negative, therefore, $q \leqslant 3$. If a column has two negative elements the other two must be positive. Hence, at least q of the elements z_{24}, z_{25}, z_{26} are positive, thus, at most $3 - q$ of the elements z_{21}, z_{22}, z_{23} are positive, showing that $p \leqslant 6 - q$.

A similar reasoning applies when z_{21}, z_{22}, z_{23} are all positive. Thus, one may assume that at most two of z_{11}, z_{12}, z_{13} are positive and at most two of z_{21}, z_{22}, z_{23} are positive, hence, $p \leqslant 4$. Similarly, one may assume that at most two of z_{34}, z_{35}, z_{36} are negative and at most two of z_{44}, z_{45}, z_{46} are negative, hence $q \leqslant 4$. One is ready if $p \leqslant 3$ and $q \leqslant 3$.

Suppose instead that for example $p = 4$. Typically, interchanging the first two rows or first three columns if necessary, the elements z_{11}, z_{12}, z_{21} are positive and further one of z_{22} or z_{23}. This forces z_{31} and z_{41} to be negative. If z_{22} were positive then also z_{32} and z_{42} would be negative and then $q \leqslant 2$. Thus, suppose instead that z_{23} is positive.

One is ready if each row of **H** contains at most one negative element, (for, then $q \leqslant 2$). If not then typically z_{44} and z_{45} are negative. This forces z_{42}, z_{43} and z_{46} to be positive which in turn forces z_{32} and z_{33} to be negative which in turn forces z_{34}, z_{35}, z_{36} to be positive and therefore that $q = 2$.

The above takes care of the (actually most difficult) case that **G** (and thus **H**) has dimension (2, 3). The other cases follow by a quite similar reasoning. Analogously for the case where the weak EMMP **Z** is of dimension (4, 5). We omit the details. This completes the proof of Theorem 6. □

9. Optimality for the General Two-Way Layout. Here, we study the case of a two-way $m \times n$ layout with weights $\omega(x) = 1$, this time allowing for several values z_{ijk} ($k = 1, \ldots, k_{ij}$) in cell (i, j) where $k_{ij} = 0$ is possible. This more general case becomes important in applications where one has a large number of observations and one likes to keep m and n relatively small so as to simplify the calculations. Unspecified indices i, j and k will run through Y_1, Y_2 and $\{1, \ldots, k_{ij}\}$, respectively.

Our problem is to minimize the sum (2.2). Thus, one needs to determine whether a given matrix of residuals

$$\mathbf{Z} = (z_{ijk} = z^{(O)}_{ijk} - \alpha_i - \beta_j)$$

is optimal. A necessary condition for **Z** to be optimal is that it be an EMP. Equivalently, that for each of the $M = m + n$ layers $L_1(i)$ and $L_2(j)$ the associated set of numbers z_{ijk} (with i fixed or j fixed) has 0 as a median. This is equivalent to

(9.1) $$\max(n_1^-(i), n_1^+(i)) \leq n_1(i)/2; \quad \max(n_2^-(j), n_2^+(j)) \leq n_2(j)/2.$$

Here,

(9.2) $$n_1(i) = \Sigma_{j=1}^{n} k_{ij}; \quad n_2(j) = \Sigma_{i=1}^{m} k_{ij}$$

is the number of elements in $L_1(i)$ and $L_2(j)$, respectively. Further,

(9.3) $$n_1^-(i) = \Sigma_{j=1}^{n} k_{ij}^-$$

is the number of negative elements $z(x) = z_{ijk}$ with $x = (i, j, k)$ in layer $L_1(i)$, thus i is fixed. Similarly for $n_1^o(i)$, $n^+_1(i)$ and $n_2^-(j)$, $n_2^o(j)$, $n^+_2(j)$. Moreover, k_{ij}^-, k_{ij}^o and k^+_{ij}, respectively, denote the number of negative, zero and positive elements z_{ijk}, respectively, located in cell (i, j).

Analogously to (7.3) we define

(9.4) $$u_1(i) = n_1(i) - 2n_1^-(i); \quad u_2(j) = n_2(j) - 2n_2^-(j).$$

The EMP property (9.1) is equivalent to

(9.5) $$0 \leq u_1(i) \leq 2n_1^o(i) = \Sigma_{j=1}^n 2k_{ij}^o;\ 0 \leq u_2(j) \leq 2n_2^o(j) = \Sigma_{i=1}^m 2k_{ij}^o.$$

THEOREM 7. *In order that* $\mathbf{Z} = (z_{ijk})$ *be optimal, it is necessary and sufficient that there exist numbers* W_{ij} *satisfying*

(9.6) $$0 \leq W_{ij} \leq 2k_{ij}^o;$$

and

(9.7) $$\Sigma_{j=1}^n W_{ij} = u_1(i); \quad \Sigma_{i=1}^m W_{ij} = u_2(j),$$

for all $i \in Y_1$ *and* $j \in Y_2$.

Proof. This result is a direct consequence of Criterion II in Section 7, applied to the set X of triplets $x = (i, j, k)$ while $z(x) = z_{ijk}$. Further put $W_{ij} = \Sigma_k W(i, j, k)$, where $W(x) = 0$ when $z(x) \neq 0$. □

Remark 1. Since k_{ij}^o and $u_t(y)$ are all integers one may even require that the W_{ij} are integers. Namely, the moment problem (9.6), (9.7) corresponds to the usual transportation problem which has a totally unimodular matrix, see Garfinkel and Nemhauser (1972) p. 73 and Hu (1969) p. 123. If all $n_1(i)$, $n_2(j)$ and thus the $u_t(y)$ are even then one can even attain that W_{ij} are even. This additional information may simplify the problem of deciding whether $\mathbf{Z}$ is optimal.

In view of the Remark following (7.5), we may conclude that an optimal matrix $\mathbf{Z}$ of residues necessarily has an integral norm provided all the original data $z_{ijk}^{(o)}$ were integers.

Remark 2. If one allows not only additive adjustments of the form $\alpha_i + \beta_j$ but also one or more additive adjustments of the form $\gamma_r g_r(i, j)$ (with the g_r as given functions and the γ_r as free constants) then optimality of $\mathbf{Z} = (z_{ijk})$ is equivalent to the existence of numbers W_{ij} satisfying (9.6), (9.7) and, moreover, the additional 'moment' conditions

(9.8) $$\Sigma_{i,j} W_{ij} g_r(i, j) = \Sigma_{i,j}(k_{ij} - 2k_{ij}^-) g_r(i, j).$$

CONDITION (A, B). Let A and B be subsets of $Y_1 = \{1, \ldots, m\}$ and $Y_2 = \{1, \ldots, n\}$, respectively. Given the matrix $\mathbf{Z}$, consider the associated arrays

(9.9) $$\mathbf{G} = (z_{ijk}; i \in A, j \notin B); \quad \mathbf{H} = (z_{ijk}; i \notin A, j \in B).$$

We will say that $\mathbf{Z}$ satisfies Conditions (A, B) if

(9.10) $$N_{\mathbf{G}}^+ + N_{\mathbf{H}}^- \leq (N_{\mathbf{G}} + N_{\mathbf{H}})/2.$$

Here,

$$N_{\mathbf{G}} = \Sigma_{i \in A} \Sigma_{j \notin B} k_{ij}; \quad N_{\mathbf{G}}^+ = \Sigma_{i \in A} \Sigma_{j \notin B} k^+_{ij}$$

denote the number of elements in $\mathbf{G}$ and the number of positive elements in $\mathbf{G}$, respectively. Similarly for $N_{\mathbf{H}}$ and $N_{\mathbf{H}}^-$.

THEOREM 8. *In order that* $\mathbf{Z} = (z_{ijk})$ *be optimal, it is necessary and sufficient that* $\mathbf{Z}$ *satisfies Condition* (A, B) *for each choice of the subsets* A *of* Y_1 *and* B *of* Y_2.

THEOREM 9. *In order that* $\mathbf{Z} = (z_{ijk})$ *be optimal, it is necessary and sufficient that the inequality*

(9.11) $$\Sigma_{i \in A}\, u_1(i) \leq \Sigma_{j=1}^n \min[u_2(j), \Sigma_{i \in A} 2k_{ij}^o]$$

holds for each subset A *of* Y_1.

Moreover, if (9.11) fails for a given subset A *of* Y_1 *then Condition* (A, B) *fails for the associated pair defined by*

(9.12) $$B = \{j \in Y_2 \colon\ u_2(j) < \Sigma_{i \in A} 2k_{ij}^o\}.$$

And in that case the matrix $\mathbf{Z}$ *admits an easy improvement.*

Remark. What is meant here is the improvement

$$\mathbf{Z}' = (z'_{ijk} = z_{ijk} - \alpha_i - \beta_j)$$

with the α_i and β_j as in the proof of Theorem 5. Namely, choose $\alpha_i = +\lambda$ when $i \in A$; $\beta_j = -\lambda$ when $j \in B$ and $\alpha_i = 0$, $\beta_j = 0$, otherwise. The best choice for λ is a median of the $N_{\mathbf{G}} + N_{\mathbf{H}}$ numbers z_{ijk} in $\mathbf{G}$ and $-z_{ijk}$ in $-\mathbf{H}$. Note that this choice of λ depends on the full matrix $\mathbf{Z}$, not only on the associated sign pattern or the numbers k^+_{ij} and k^-_{ij}.

Proof of Theorems 8 and 9. For $t \in T = \{1, 2\}$, let μ_t be the (possibly signed) measure on Y_t having a mass $u_t(y)$ at $y \in Y_t$. Let further $q(\ .\ )$ denote the (nonnegative) measure on $Y_1 \times Y_2$ having a mass $2k^o_{ij}$ at the point (i, j). The criterion for optimality stated in Theorem 7 requires precisely that there exists a (nonnegative) measure μ on $Y_1 \times Y_2$ having marginals μ_1 and μ_2 and such that $\mu(E) \leq q(E)$ for every subset E of $Y_1 \times Y_2$. A necessary condition for the existence of μ is that

(9.13) $$\mu_1(A) \leq q(A \times B^c) + \mu_2(B),$$

for every subset A of Y_1 and every subset B of Y_2. After all, $A \times Y_2 \subset (A \times B^c) \cup (Y_1 \times B)$ and $\mu(A \times Y_2) = \mu_1(A)$; $\mu(Y_1 \times B) = \mu_2(B)$; (taking A empty, this requires that $\mu_2 \geq 0$; similarly $\mu_1 \geq 0$ since $\mu_1(Y_1) = \mu_2(Y_2)$, see (7.7)).

As was shown by Dall'Aglio (1961) and Kellerer (1961), condition (9.13) is also sufficient for the existence of μ, hence, it is equivalent to the optimality of $\mathbf{Z}$. See Strassen (1965) p. 423 for generalizations and further references. The sufficiency of (9.13) is also an immediate consequence of the Ford-Fulkerson max-flow-min-cut theorem, see Ford and Fulkerson (1962), Berge (1970) and Jacobs (1978) p. 539.

For each fixed pair A and B, (9.13) is equivalent to Condition (A, B), proving Theorem 8. After all, using (9.2), (9.3), (9.4), the inequality (9.13) can be written as

$$\Sigma_{i \in A}\Sigma_j(k^+_{ij} + k^o_{ij} - k^-_{ij}) \leq \Sigma_{i \in A}\Sigma_{j \notin B} 2k^o_{ij} + \Sigma_i\Sigma_{j \in B}(k^+_{ij} + k^o_{ij} - k^-_{ij}).$$

Equivalently,

$$0 \leq \Sigma_{i \in A}\Sigma_{j \notin B}(-k^+_{ij} + k^o_{ij} + k^-_{ij}) + \Sigma_{i \notin A}\Sigma_{j \in B}(k^+_{ij} + k^o_{ij} - k^-_{ij}).$$

In view of (9.9) this is equivalent to (9.10).

Given the subset A of Y_1, one might as well choose the subset B of Y_2 so as to make the right hand side of (9.13) as small as possible. For $j \in Y_2$, putting j in B yields a contribution $\mu_2(\{j\}) = u_2(j)$; putting j in B^c yields a contribution $q(A \times \{j\}) = \Sigma_{i \in A} 2k^o_{ij}$. Thus, the best choice for B would be as in (9.12), in which case (9.13) reduces to (9.11). Ths proves the first part of Theorem 9.

If (9.11) fails for a set A then (9.13) fails for the pair A, B with B as in (9.12), which in turn means precisely that Condition (A, B) fails. This in turn allows us to improve the matrix $\mathbf{Z}$ as explained in the above Remark. □

ALGORITHM. A good algorithm for minimizing the sum (2.2) would be to apply the Theorems 7 and 9, at each stage of the calculation, to the matrix $\mathbf{Z} = (z_{ijk})$ of residues on hand. One first tries to construct the set of numbers W_{ij} as in Theorem 7 by using the standard max-flow-min-cut algorithm. If this does not succeed then $\mathbf{Z}$ is optimal and we are ready.

If this attempt does not succeed then the calculation automatically leads to a 'cut' of small capacity which in turn corresponds to the failure of a well-defined Condition (A, B). Using the latter knowledge, one next improves the matrix $\mathbf{Z}$ of residuals as explained in the Remark following Theorem 9. Afterwards, one tests the optimality of the new matrix $\mathbf{Z}'$ of residuals by trying to construct the desired numbers W_{ij}. And so on. Provided the original data $z^{(o)}_{ijk}$ are all integers, one can arrange the calculation so that also all subsequent residual matrices $\mathbf{Z}$ are integral in which case the norm decreases each time by a positive integer. Hence, the calculation will then lead in finitely many steps to an optimal matrix of residuals.

In more detail, in trying to construct (W_{ij}), one considers a *directed capacitated network* with vertex set $V = \{a\} \cup \{b\} \cup Y_1 \cup Y_2$ and with a as the only source, b as the only sink. One has the following directed edges (x, y) and associated capacities $k(x, y)$.

(i) The edges (a, i) with $i \in Y_1$ and capacity $u_1(i)$.

(ii) The edges (i, j) with $i \in Y_1, j \in Y_2$ and capacity $2k^o_{ij}$.

(iii) The edges (j, b) with $j \in Y_2$ and capacity $u_2(j)$. Note that the $u_1(i)$, $u_2(j)$ and k^o_{ij} are integers and that $\Sigma_i u_1(i) = \Sigma_j u_2(j) = Q$ (say), see (7.7).

One proceeds with determing an admissible flow f in this network (with $f(x, y)$ as the flow along the directed edge (x, y)) which maximizes the total flow from a to b. Admissibility means here that $0 \leq f(x, y) \leq k(x, y)$.

Relative to a given admissible flow f, an *unsaturated path* from the vertex a to the vertex x is defined as a sequence $x_o = a, x_1, \ldots, x_{n-1}, x_n = x$ of distinct vertices such that the flow from a to x along that path can be increased. More precisely, this requires that, for $i = 1, \ldots, n$, either (x_{i-1}, x_i) is an edge of the network and $f(x_{i-1}, x_i)$ is smaller than the capacity $k(x_{i-1}, x)$; or (x_i, x_{i-1}) is an edge of the network and $f(x_i, x_{i-1})$ is positive.

Let V_f denote the set of all vertices x such that some unsaturated path leads from a to x. During the construction of V_f, one marks each new member of V_f with a single label pointing to a previously constructed vertex in V_f (from which it 'originated') so as to allow for backtracking. As soon as V_f is found to contain the sink b, one obtains through backtracking an unsaturated path from a to b. One proceeds to increase the flow along that path in an obvious and maximal way. This new flow is again integer valued, provided one starts with an integer valued flow (such as the zero flow). After finitely many steps, no further increase of the total flow from a to b is possible and one has reached an integer flow f with the property that $b \notin V_f$. Let

$$F = \Sigma_{i=1}^m f(a, i) = \Sigma_{j=1}^n f(j, b)$$

be the resulting total flow from a to b. There are the following possibilities.

(I) $F = Q$. In this case, the set of edge flows $W_{ij} = f(i, j)$ $(i \in Y_1; j \in Y_2)$ satisfies (9.6) and (9.7), consequently, the present residual matrix $\mathbf{Z}$ is optimal.

(II) $F < Q$. The $\mathbf{Z}$ is not optimal. In fact, Condition (A, B) fails with

(9.14) $$A = V_f \cap Y_1 \quad \text{and} B = V_f \cap Y_2,$$

allowing us to improve the matrix $\mathbf{Z}$.

Proof. Let E denote the set of edges (x, y) with $x \in V_f$ and $y \notin V_f$. The sum of all the corresponding capacities $k(x, y)$ is called the capacity of E. The set E is known to define a 'cut' whose capacity is equal to the maximal flow F on hand and, thus, is smaller than Q.

In fact, E consists of the edges (a, i) with $i \in A^c$, further the edges (j, b) with $j \in B$ and finally the edges (i, j) with $i \in A$ and $j \in B^c$. As is easily seen, the capacity of this set E being smaller than Q means exactly that (9.13) is false and, hence, that Condition (A, B) fails. □

REFERENCES

ABDELMALEK, N. N. (1974). On the discrete L_1 approximation and L_1 solutions of overdetermined linear equations, *J. Approx. Theory 11* 38–53.

AMEMIYA, I. and ANDO, T. (1965). Convergence of random products of contractions in Hilbert space, *Acta Sci. Math.* (Szeged) *26* 239–244.

ANSCOMBE, F. (1981). *Computing in Statistical Science through APL*. Springer-Verlag, N.Y.

ARMSTRONG, R. D., ELAM, J. J. and HULTZ, J. W. (1977). Obtaining least absolute value estimates for a two-way classification model, *Commun. Statist. Simula. Computa. B6(4)* 365–381.

ARMSTRONG, R. D., FROME, E. L. and KUNG, D. S. (1979). A revised simplex algorithm for the absolute deviation curve fitting problem, *Commun. Statist. Simula. Computa. B8(2)* 175–190.

BARRODALE, I. and ROBERTS, F. D. K. (1973). An improved algorithm for discrete ℓ_1 approximation, *SIAM J. Numer. Anal. 10* 839–848.

BARRODALE, I. and ROBERTS, F. D. K. (1978). An efficient algorithm for discrete ℓ_1 approximations with linear constraints, *SIAM J. Numer. Anal. 15* 603–611.

BARTELS, R. H., CONN, A. R. and SINCLAIR, J. W. (1978). Minimization techniques for piecewise differentiable functions: The ℓ_1 solution of an overdetermined linear system, *SIAM J. Numer. Anal. 15* 224–241.

BERGE, C. (1973). *Graphs and Hypergraphs*, North-Holland, Amsterdam.

FORD, Jr., L. K. and FULKERSON, D. K. (1962). *Flows in Networks*, Princeton University Press, Princeton.

FORSYTHE, A. B. (1972). Robust estimation of straight line regression coefficients by minimizing p-th power deviations, *Technometrics 14* 159–166.

GABRIEL, K. R. (1983). An alternative computation of median fits in two-way tables, *Tech. Report* No. 83/01, Department of Statistics, University of Rochester, Rochester, N.Y.

GARFINKEL, R. S. and NEMHAUSER, G. L. (1972). *Integer Programming*, Wiley, New York.

GENTLE, J. E. (1977). Least absolute value estimation: an introduction, *Commun. Statist. Simula. Computa. B6(4)* 313–328.

HU, T. C. (1969). *Integer Programming and Network Flows*, Addison-Wesley, Reading, MA.

JACOBS, K. (1978). *Measure and Integral*, Academic Press, New York.

KELLERER, H. (1961). Funktionen auf Produkträumen mit vorgegebenen Marginalfunktionen, *Math. Ann. 144* 323–344.

KEMPERMAN, J. H. B. (1983). On the role of duality in the theory of moments, pp. 63–92 in FIACCO, A. V. and KORTANEK, K. O., eds., *Semi-Infinite Programming and Applications*, Lecture Notes in Economics and Mathematical Systems, vol. 215, Springer-Verlag, New York.

KENNEDY, Jr., W. J. and GENTLE, J. E. (1980). *Statistical Computing*, Marcel Dekker, New York.

MONEY, A. H., AFFLECK-GRAVES, J. F. and BARR, G. D. I. (1982). The linear regression model: L_p norm estimation and the choice of p, *Commun. Statist. Simula. Computa. 11(1)* 89–109.

MOSTELLER, F. and TUKEY, J. W. (1977). *Data Analysis and Regression*, Addison-Wesley, Reading, MA.

SMITH, K. T., SOLMON, D. C. and WAGNER, S. L. (1977). Practical and mathematical aspects of the problem of reconstructing objects from radiographs, *Bull. Amer. Math. Soc. 83* 1227–1270.

STRASSEN, V. (1965). The existence of probability measures with given marginals, *Annals Math. Statist. 36* 423–438.

TUKEY, J. W. (1977). *Exploratory Data Analysis*, Addison-Wesley, Reading, MA.

WAGNER, H. M. (1959). Linear programming techniques for regression analysis, *J. Amer. Statist. Assoc. 54* 206–212.

Inequalities in Statistics and Probability
IMS Lecture Notes–Monograph Series Vol. 5 (1984), 104–108

MARKOV'S INEQUALITY FOR RANDOM VARIABLES TAKING VALUES IN A LINEAR TOPOLOGICAL SPACE

BY ALBERT W. MARSHALL
University of British Columbia

Let X be a random variable taking values in the linear topological space $\mathcal{X}$ and let $C \subset \mathcal{X}$ be the closed convex cone which generates the preordering $\precsim$. For an appropriate definition of EX and for $c \in C$, a sharp upper found for $P[X \succsim \varepsilon]$ is obtained in terms of EX. Similarly, a lower bound for $P[X \prec\!\!\!\not\ \varepsilon]$ is obtained which is sharp in certain special cases.

1. Introduction. If a random variable X satisfies

$$P[X \geqslant 0] = 1, \quad EX = \mu, \tag{1.1}$$

and if $\varepsilon > 0$, then according to Markov's inequality,

$$P[X \geqslant \varepsilon] \leqslant \min\{\mu/\varepsilon, 1\}. \tag{1.2}$$

Moreover there is a distribution for X satisfying (1.1) for which (1.2) holds with equality. Thus (1.2) is "sharp" in the sense that the bound cannot be improved without information in addition to (1.1) about the distribution of X.

This paper is concerned with inequalities similar to (1.2) which hold for random variables that need not be real-valued, but take values in a real or complex linear topological space $\mathcal{X}$. To obtain such extensions, two preliminaries are required: First, meaning has to be given to inequalities "$a \geqslant b$" for a,b in $\mathcal{X}$. Second, meaning must be given to the notion of an expectation.

For random variables taking values in the finite dimensional space $\mathcal{R}^n$, the expected value is naturally taken to be the vector of expected values. More generally, the expected value can be defined, e.g., as a Pettis integral: see Perlman (1974) for a similar use of this integral and for the references contained therein. In this paper, it is assumed only that when it exists, $EX = \int X dP \in \mathcal{X}$ and the following properites are satisfied:

$$\int (X + Y) dP = \int X dP + \int Y dP, \tag{1.3}$$

$$\text{If } A \subset \mathcal{X} \text{ is closed and convex, } P[X \in A] = 1 \text{ implies } \int X dP \in A, \tag{1.4}$$

$$\text{For all events } E \text{ and } c \in \mathcal{X},\ \int_E c\, dP = cP(E). \tag{1.5}$$

The expression $a \geqslant b$ can be rewritten as $a - b \in [0,\infty)$ and $a > b$ can be rewritten as $a - b \in (0,\infty)$. Since $[0,\infty)$ is a closed convex cone with interior $(0,\infty)$, it is natural and standard when replacing $(-\infty,\infty)$ by a linear topological space $\mathcal{X}$ to replace $[0,\infty)$ by a closed convex cone $\mathcal{C} \subset \mathcal{X}$. For $x,y \in \mathcal{X}$, write

$$x \precsim y \quad \text{if } y - x \in \mathcal{C}, \tag{1.6}$$

$$x \prec y \quad \text{if } y - x \in \mathcal{C}^0, \tag{1.7}$$

where $\mathcal{C}^0$ is the interior of $\mathcal{C}$. Defined in this way, $\precsim$ is a preordering of $\mathcal{X}$, i.e.,

$$x \precsim y \quad \text{for all } x \in \mathcal{X} \tag{1.8}$$

$$x \precsim y \text{ and } y \precsim z \text{ implies } x \precsim z,\ x,y,z \in \mathcal{X}. \tag{1.9}$$

AMS 1980 subject classification. Primary 60E15.

Key words and phrases: Markov's inequality, order preserving functions.

Moreover, $\precsim$ satisfies

(1.10) $\quad x \precsim y$ implies $x+z \precsim y+z$ for all $x,y,z \in \mathcal{X}$,

(1.11) $\quad x \precsim y$ implies $\lambda x \precsim \lambda y$ for all $\lambda \geqslant 0$, $x,y \in \mathcal{X}$.

Of course (1.2) is equivalent to

(1.2′) $$P[X<\varepsilon] \geqslant 1-\min\{\mu/\varepsilon, 1\},$$

but such an equivalence does not hold when $\leqslant$ is replaced by a partial order $\precsim$. In Section 2 below, upper bounds are obtained for $P[X \succsim \varepsilon]$ and in Section 3, lower bounds for $P[X \prec \varepsilon]$ (upper bounds for $P[X \not\prec \varepsilon]$) are obtained.

For purposes of this paper, certain families $\mathcal{F}$ of real-valued functions defined on $\mathcal{X}$ play a key role. Some conditions that may be imposed on $\mathcal{F}$ are the following:

(1.12) $\quad x \precsim y$ if and only if $f(x) \leqslant f(y)$ for all $f \in \mathcal{F}$,

(1.12′) $\quad x \prec y$ if and only if $f(x) < f(y)$ for all $f \in \mathcal{F}$,

(1.13) $\quad f \in \mathcal{F}$ implies $f(x) \geq 0$ for all $x \in C$,

(1.14) $\quad f \in \mathcal{F}$ implies $f(ax) \geqslant af(x)$ for all $a \in [0,1]$, $x \in C$.

In what follows, infima or minima taken over empty sets are to be regarded as ∞.

2. Upper Bounds for $P[X \succsim \varepsilon]$.

2.1 Proposition. *Let $C \subset \mathcal{X}$ be a closed convex cone which determines the ordering $\precsim$ via (1.6). Let X be a random variable such that $P[X \in C] = 1$ and $EX = \mu$ exists. Let $\mathcal{F}$ be a set of functions satisfying (1.12), (1.13), (1.14). If $\varepsilon \in C$, then*

(2.1) $$P[X \succsim \varepsilon] \leqslant \min\{1, \inf_{\{f: f \in \mathcal{F}, f(\varepsilon)>0\}} f(\mu)/f(\varepsilon)\}$$

Proof. By using (1.3)–(1.6) and (1.10) if follows that

$$\mu = \int X dp = \int_{\{X \succsim \varepsilon\}} X dP + \int_{\{X \not\succsim \varepsilon\}} X dP \succsim \int_{\{X \succsim \varepsilon\}} X dP \succsim \int_{\{X \succsim \varepsilon\}} \varepsilon dP = \varepsilon P[X \succsim \varepsilon].$$

But this implies that

$$f(\mu) \geqslant f(\varepsilon P[X \succsim \varepsilon]) \geqslant P[X \succsim \varepsilon] f(\varepsilon) \quad \text{for all } f \in \mathcal{F},$$

i.e.,

$$P[X \succsim \varepsilon] \leqslant f(\mu)/f(\varepsilon) \quad \text{for all } f \in \mathcal{F} \text{ such that } f(\varepsilon) > 0. \qquad \square$$

2.2 Proposition. If (1.14) holds with equality for all $f \in \mathcal{F}$, then for each μ, $\varepsilon \in C$, equality is attainable in (2.1).

Proof. Suppose first that upper bound p of (2.1) is 1 and let Y be a random variable such that $P[Y = \mu] = 1$. By (1.4), $EY = \mu$ so that Y satisfies the conditions of Proposition 2.1. By (1.12) and (1.13) it follows that $\mu \succsim \varepsilon$, that is $P[Y \succsim \varepsilon] = 1$, so equality holds in (2.1).

Next, suppose that $p < 1$ and that

$$P[Y = \varepsilon] = p, \quad P[Y = \alpha] = 1-p$$

where $\alpha = (\mu - \varepsilon p)/(1-p)$. Because $p < 1$ it follows from (1.12) and (1.13) that $\mu \not\succsim \varepsilon$ so $\mu - \varepsilon p \not\succsim \varepsilon - \varepsilon p$ or $(1-p)\alpha \not\succsim (1-p)\varepsilon$. Thus $\alpha \not\succsim \varepsilon$, so for this distribution,

$$P[Y \succsim \varepsilon] = P[Y = \varepsilon] = p.$$

To show that $P[Y \in C] = 1$, it is necessary to show only that $\alpha \in C$, since $\varepsilon \in C$ by assumption. Since $f(\mu)/f(\varepsilon) \geqslant p$ for all $f \in \mathcal{F}$ such that $f(\varepsilon) > 0$, it follows that $f(\mu) \geqslant pf(\varepsilon) = f(p\varepsilon)$ for all $f \in \mathcal{F}$ hence $\mu \succsim p\varepsilon$, that is, $\alpha \in C$.

From (1.3) and (1.5), it follows that $EY = p\varepsilon + (1-p)(\mu - \varepsilon p)/(1-p) = \mu$. Consequently Y satisfies the condition of Proposition 2.1 and equality is achieved in (2.1). □

2.3 Example. Suppose $\mathcal{X} = \mathcal{R}^n$ and $\mathcal{C} = \{\mathbf{x} = (x_1, \ldots, x_n): x_i \geq 0, i = 1, \ldots, n\} = \mathcal{R}^n_+$ is the nonnegative orthant. Let $\mathcal{F}$ consist of the coordinate functions $f_1, \ldots, f_n$, where $f_i(x) = x_i$. If $\varepsilon \in \mathcal{R}^n_+$ and $\varepsilon \neq 0$, then

$$\inf_{\{f: f \in \mathcal{F}, f(x) > 0\}} f(\mu)/f(\varepsilon) = \min_{\{i: \varepsilon_i > 0\}} EX_i/\varepsilon_i$$

so that if $\varepsilon_i \geq 0, i = 1, \ldots, n$,

$$P[X_i \geq \varepsilon_i, i = 1, \ldots, n] \leq \min_{\{i: \varepsilon_i > 0\}} EX_i/\varepsilon_i \tag{2.2}$$

This inequality follows from (4.1) or (7.1) of Marshall and Olkin (1960). It is also equivalent to Corollary 2.1 of Jensen and Foutz (1981).

2.4 Example. Let $\mathcal{X}$ be the linear space of $n \times n$ Hermitian matrices and let $\mathcal{C}$ be the convex cone of positive semi-definite matrices. Take $\mathcal{F}$ to consist of functions of the form $f_{\mathbf{a}}$ where $\mathbf{a}$ is a unit vector ($\mathbf{aa}^* = 1$) of a complex numbers and $f_{\mathbf{a}}(\mathbf{A}) = \mathbf{aAa}^*$. Suppose that $\mathbf{C}$ is positive definite. If the random matrix $\mathbf{X}$ is positive definite with probability one, $\inf_{f \in \mathcal{F}} f(E\mathbf{X})/f(\mathbf{C}) = \inf_{\mathbf{a}} \mathbf{a}E\mathbf{X}\mathbf{a}^*/\mathbf{aCa}^* = \min_{\mathbf{bb}^*=1} \mathbf{b}\mathbf{C}^{-1/2}E\mathbf{X}\mathbf{C}^{-1/2}\mathbf{b}^* = \lambda_n[\mathbf{C}^{-1/2}(E\mathbf{X})\mathbf{C}^{-1/2}]$, the minimum characteristic root of $\mathbf{C}^{-1/2}(E\mathbf{X})\mathbf{C}^{-1/2}$. Thus

$$P[\mathbf{X} \succsim \mathbf{C}] \leq \lambda_n[\mathbf{C}^{-1/2}(E\mathbf{X})\mathbf{C}^{-1/2}]. \tag{2.3}$$

This result is given in Corollary 3.3 of Jensen and Foutz (1981).

2.5 Example. Let $\mathcal{X} = \mathcal{R}^n$ and suppose that $\precsim_w$ is the ordering of weak submajorization (see Marshall and Olkin, 1979, p. 10). Restricted to $\mathcal{D} = \{\mathbf{x}: x_1 \geq \ldots \geq x_n\}$, this ordering is generated by the convex cone $\mathcal{C} = \{\mathbf{x}: \Sigma_{i=1}^k x_i \geq 0, k = 1, \ldots, n\}$. Replace the random vector $\mathbf{X} = (X_1, \ldots, X_n)$ by $\mathbf{X}_\downarrow = (X_{[1]}, \ldots, X_{[n]})$ where $X_{[1]} \geq \ldots \geq X_{[n]}$ are obtained by ordering $X_1, \ldots, X_n$. Let $\mathcal{F}$ consist of the functions $f_k(\mathbf{x}) = \Sigma_{i=1}^k x_{[i]}, k = 1, \ldots, n$. If $\varepsilon \in \mathcal{C}$,

$$\min_{\{f \in \mathcal{F}, f(\varepsilon) > 0\}} f(E\mathbf{X})/f(\varepsilon) = \min_{\{k: \Sigma_{i=1}^k \varepsilon_{[i]} > 0\}} = \Sigma_{i=1}^k EX_{[i]}/\Sigma_{i=1}^k \varepsilon_{[i]},$$

so that

$$P[\mathbf{X} \succsim_w \varepsilon] = P[\mathbf{X}_\downarrow \succsim_w \varepsilon] \leq \min_{\{k: \Sigma_{i=1}^k \varepsilon_{[i]} > 0\}} = \Sigma_{i=1}^k EX_{[i]}/\Sigma_{i=1}^k \varepsilon_{[i]}. \tag{2.4}$$

The bound of this inequality is in terms of $E\mathbf{X}_\downarrow$, not of $E\mathbf{X}$. Because $E\mathbf{X}$ is majorized by $E\mathbf{X}_\downarrow$ (Marshall and Olkin (1979), p. 348), it is not possible to replace $E(X_{[i]})$ by the i-th largest component of $E\mathbf{X}$ in the above bound.

3. Upper Bounds for $P\{X \not\prec \varepsilon\}$. In general, $X \succsim \varepsilon$ implies $X \not\prec \varepsilon$ but not conversely, so it is to be expected that a sharp upper bound for $P[X \not\prec \varepsilon]$ will be larger than the corresponding bound for $P[X \succsim \varepsilon]$ found in Section 2.

The following proposition is less satisfactory than Proposition 2.1 because it is little more than Markov's inequality (1.2) and requires additional steps to yield a bound in terms of EX.

3.1 PROPOSITION. Let $\mathcal{C} \subset \mathcal{X}$ be a closed convex cone and let X be a random variable such that $P[X \in \mathcal{C}] = 1$ and that $EX = \mu$ exists. Let $\mathcal{F}$ be a set of functions satisfying (1.12′) and (1.13). If $\varepsilon \in \mathcal{C}^0$ then

$$P[X \not\prec \varepsilon] \leq \min\{1, E \sup_{f \in \mathcal{F}} f(X)/f(\varepsilon)\}. \tag{3.1}$$

Remark. Because $\varepsilon \succ 0$, it follows from (1.12′) and (1.13) that

$$f(\varepsilon) > f(0) \geqslant 0 \quad \text{for all} f \in \mathcal{F}.$$

Proof. From (1.12′), (1.13), and Markov's inequality (1.2) it follows that

$$P[X \not\prec \varepsilon] = P[f(X) \geqslant f(\varepsilon) \ \text{ for some} f \in \mathcal{F}] \leqslant P[\sup_{f\in\mathcal{F}} f(X)/f(\varepsilon) \geqslant 1]$$
$$\leqslant E \sup_{f\in\mathcal{F}} f(X)/f(\varepsilon). \qquad \square$$

The following examples show that (3.1) sometimes leads to sharp bounds in terms of EX.

3.2 Example. Suppose $\mathcal{X} = \mathcal{R}^n$, $\mathcal{C} = \{\mathbf{x} = (x_1, \ldots, x_n): x_i \geqslant 0, i = 1, \ldots, n\} = \mathcal{R}^n_+$. Let $\varepsilon = (\varepsilon_1, \ldots, \varepsilon_n)$ where each $\varepsilon_i > 0$ and let $\mathcal{F}$ consist of the coordinate functions $f_1, \ldots, f_n$ where $f_i(\mathbf{x}) = x_i$. If $\mathbf{X}$ is an $\mathcal{X}$-valued random variable such that $E\mathbf{X} = \mu$ exists, then

$$P[X_i \geqslant \varepsilon_i \ \text{ for some } i = 1, \ldots, n] \leqslant \min\{1, \textstyle\sum_{i=1}^n \mu_i/\varepsilon_i\}. \tag{3.2}$$

Proof. Since $\sup_{f\in\mathcal{F}} f(x)/f(\varepsilon) \leqslant \sum_{f\in\mathcal{F}} f(x)/f(\varepsilon)$ and since $Ef(X) = f(EX)$ for all $f \in \mathcal{F}$, (3.2) follows from (3.1). $\square$

In spite of its apparent crudeness, inequality (3.2) is sharp. To see this, suppose first that the upper bound is less than one and let $\mathbf{e}_i$ be the vector with i-th coordinate 1 and all other coordinates 0. Let $\mathbf{Y}$ be a random vector such that

$$P[\mathbf{Y} = \varepsilon_i \mathbf{e}_i] = \mu_i/\varepsilon_i, \; i = 1, \ldots, n$$
$$P[\mathbf{Y} = \mathbf{0}] = 1 - \Sigma \mu_i/\varepsilon_i.$$

Then $E\mathbf{Y} = \mu$ and equality is attained in (3.2).

Next, suppose the upper bound of (3.2) is one and let $s = \sum_{i=1}^n \mu_i/\varepsilon_i$. Let $\mathbf{Y}$ be a random vector such that

$$P[\mathbf{Y} = s\varepsilon_i \mathbf{e}_i] = \mu_i/s\varepsilon_i.$$

Since $s \geqslant 1$, $P[Y_i \geqslant \varepsilon_i$ *for some* $i = 1, \ldots, n] = 1$.

3.3 Example. Suppose $\mathcal{X}$ consists of $n \times n$ Hermitian matrices and $\mathcal{C}$ consists of the positive semi-definite Hermitian matrices. If $P[\mathbf{X} \in \mathcal{C}] = 1$, $E\mathbf{X} = \mu$ exists and $\mathbf{C}$ is positive definite, then

$$P[\mathbf{X} \not\prec \mathbf{C}] \leqslant \min\{1, \operatorname{tr} \mathbf{C}^{-1/2} \mu \mathbf{C}^{-1/2}\}. \tag{3.3}$$

To obtain (3.3) from (3.1), take $\mathcal{F}$ as in Example 2.4. Denote the largest eigenvalue of an Hermitian matrix $\mathbf{H}$ by $\lambda_1(\mathbf{H})$. Then

$$E \sup_{\mathbf{a}} \mathbf{a X a}^*/\mathbf{a C a}^* = E \sup_{\{\mathbf{a}:\mathbf{aa}^*=1\}} \mathbf{a C}^{-1/2} \mathbf{X} C^{-1/2} \mathbf{a}^* = E\lambda_1(\mathbf{C}^{-1/2} \mathbf{X} C^{-1/2})$$
$$\leqslant E \operatorname{tr} \mathbf{C}^{-1/2} \mathbf{X} \mathbf{C}^{-1/2} = \operatorname{tr} \mathbf{C}^{-1/2} (E\mathbf{X}) \mathbf{C}^{-1/2}.$$

Thus (3.3) follows from (3.1).

To see that (3.3) is sharp, suppose without loss of generality that $\mathbf{C} = \mathbf{I}$; otherwise replace $\mathbf{X}$ by $\mathbf{C}^{-1/2}\mathbf{X}\mathbf{C}^{-1/2}$. Write μ in the form $\mu = \Gamma \mathbf{D} \Gamma^*$ where $\mathbf{D} = \operatorname{diag}(d_1, \ldots, d_n)$ is diagonal and Γ is unitary. Suppose the bound is less than one and let $\mathbf{E}_i = \operatorname{diag} \mathbf{e}_i$ where $\mathbf{e}_i$ is defined in 3.2. If

$$P[\mathbf{Y} = \Gamma \mathbf{E}_i \Gamma^*] = d_i, \quad i = 1, \ldots, n$$
$$P[\mathbf{Y} = \mathbf{0}] = 1 - \Sigma_1^n d_i,$$

then $E\mathbf{Y} = \Sigma\, d_i \Gamma \mathbf{E}_i \Gamma^* = \Gamma(\Sigma d_i \mathbf{E}_i)\Gamma^* = \Gamma \mathbf{D} \Gamma^* = \mu$. Moreover $P[\mathbf{X} \prec \mathbf{I}] = P[\mathbf{X} = \mathbf{0}] = 1 - \operatorname{tr} \mu$ so equality holds in (3.3).

In case the bound of (3.3) is one, the above example can be modified to show that equality is attainable using ideas similar to those used for Example 3.2.

3.4 Example. Let $\mathcal{X} = \mathcal{R}^n$ and supposed that $\precsim_w$ is the ordering of weak submajorization, as in Example 2.5. With $\mathcal{C}$ and $\mathcal{F}$ as in Example 2.5, it follows from (3.1) that

$$P[\mathbf{X} \precsim_w \varepsilon] \leq \Sigma_{k=1}^n [\Sigma_{j=1}^k \mu_{[j]} / \Sigma_{j=1}^n \varepsilon_{[j]}]. \tag{3.4}$$

Acknowledgement. The author would like to thank the editors for their comments on the earlier version of this paper.

REFERENCES

JENSEN, D. R. and FOUTZ, R. V. (1981). Markov Inequalities on partially ordered spaces. *J. Multivariate Anal.* *11* 250–259.

MARSHALL, A. W. and OLKIN, I. (1960). Multivariate Chebyshev Inequalities. *Ann. Math. Statist.* *31* 1001–1014.

MARSHALL, A. W. OLKIN, I. (1979). *Inequalities: Theory of Majorization and Its Applications.* New York: Academic Press.

PERLMAN, M. D. (1974). Jensen's inequality for a convex vector-valued function on an infinite-dimensional space. *J. Multivariate Anal.* *4* 52–65.

Inequalities in Statistics and Probability
IMS Lecture Notes–Monograph Series Vol. 5 (1984), 109–111

PROBABILITY MEASURES ON THE CIRCLE AND THE ISOPERIMETRIC INEQUALITY

BY RICHARD A. VITALE
Claremont Graduate School

The theory of planar convex sets invokes measures of a certain class. Accordingly, the isoperimetric inequality can be translated into quadratic inequalities for probability measures on the unit circle.

Various tools from convex geometry have been put to highly effective use in probability theory. One example is in the application by Anderson (1955) of the Brunn-Minkowski inequality to the probability content of symmetric, convex sets (see Tong (1980, chapter 4) for this and related results). A second instance is the resolution by Egorychev and Falikman of the van der Waerden conjecture by means of mixed discriminants, which arose originally in the study of mixed volumes (see Lagarias (1982)). The latter have recently been applied to combinatorial questions (Stanley (1981)).

That convex geometry and probability should be linked is not too surprising since each has a strong concern with the notion of positivity. What geometers have appreciated for some time, and what perhaps awaits systematic exploitation by probabilists, is that this link can be made concrete. This goes by the historical name of Minkowski's problem (Busemann (1958, pp. 60–67)). Roughly speaking, each compact, convex set in $\mathcal{R}^n$ can be identified with a bounded, positive measure on the unit sphere in that space. From the probabilist's point of view, a wide class of probability measures on the unit sphere *can be realized as compact, convex sets*. Existence of atoms, modes of convergence, and even statistical procedures have natural geometric analogs.

The author will treat some of these questions elsewhere. Here the flavor of the connection will be given by deriving two inequalities for probability measures μ on the unit circle $C = [0,2\pi)$.

INEQUALITY I.

$$\int_C\int_C g(\theta-\lambda)\mu(d\theta)\mu(d\lambda) \leq (2\pi)^{-1} \tag{I}$$

where $g(\lambda) = [2(\pi-\lambda)\sin\lambda - \cos\lambda]/4\pi$ on $[0,2\pi)$ and is extended 2π periodically. Equality holds iff μ admits the representation

$$\mu(d\lambda) = [(2\pi)^{-1} + c_1\cos\lambda + c_2\sin\lambda]d\lambda$$

for constants c_1, c_2.

INEQUALITY II.

$$\int_C\int_C |\sin(\theta-\lambda)|\mu(d\theta)\mu(d\lambda) \leq 2/\pi \tag{II}$$

with equality iff $1/2[\mu(d\lambda) + \mu(d(\lambda+\pi))] = (2\pi)^{-1}\, d\lambda$.

AMS 1980 subject classifications. Primary 60E15. Secondary 52A10..
Key words and phrases: Brünn-Minkowski inequality, isoperimetric inequality, mixed volumes.

Geometrically I and II are versions of the isoperimetric inequality. Once a certain amount of geometric machinery is in place, they fall out immediately. We sketch the arguments.

The point of departure is the notion of the *support function* of a compact, convex subset K of the plane,

$$s_K(\theta) = \max_{x \in K} \langle e_\theta, x \rangle,\ e_\theta = (\cos\theta, \sin\theta).$$

It is possible to show that the class of support functions coincides with functions of the form

$$s(\theta) = a\cos\theta + b\sin\theta + \int_C g(\theta - \lambda)R(d\lambda)$$

where a and b are constants, g is as described above, and R is a positive, bounded measure satisfying

$$(*) \qquad \int_C \sin\lambda\, R(d\lambda) = 0 = \int_C \cos\lambda\, R(d\lambda)$$

(Blaschke (1949, p. 116), Grenander (1976, p. 198), Vitale (1974, 1979)). The trigonometric term amounts to a location parameter so that the final term exhibits the correspondence between sets K and measures of the specified type. If K is a singleton, then R is the zero measure. More generally, the total mass assigned by R is the perimeter of K, $\mathrm{per}(K) = \int_C R(d\lambda)$, and the area of K is quadratic in R

$$\mathrm{area}(K) = 1/2 \int_C \int_C g(\theta - \lambda)R(d\theta)R(d\lambda).$$

Accordingly, the isoperimetric inequality reads

$$(\mathrm{ISO}) \qquad 4\pi \cdot 1/2 \int_C \int_C g(\theta - \lambda)R(d\theta)R(d\lambda) \leq [\int_C R(d\lambda)]^2.$$

with equality iff R is a constant multiple of Lebesgue measure.

Note that (ISO) can be asserted only for measures which annihilate sin and cos. Inequalities I and II represent two ways of approaching this constraint.

For (I), begin by orthogonalizing a given probability measure μ to sin and cos. This probably yields negative values so add back a constant multiple of Lebesgue measure. Thus

$$\hat{\mu}(d\lambda) = \mu(d\lambda) + [\int \pi^{-1}\{1 - \cos(\lambda - \theta)\}\, \mu(d\theta)]d\lambda.$$

(ISO) holds for $\hat{\mu}$. It is direct to show that g annihilates sin and cos, so that the quadratic term can be written with the measure $\mu(d\lambda) + (\pi)^{-1}\, d\lambda$. Noting that $\int \hat{\mu}\, (d\lambda) = 3$ and simplifying expressions yields (I).

For (II), we observe that $|\sin\theta|$ is the support function of the line segment $-1 \leq y \leq 1$. Using a probability measure μ for convex combination leads to a support function $\sigma(\theta) = \int |\sin(\theta - \lambda)|\mu(d\lambda)$. Now $|\sin\theta| = \int g(\theta - \lambda)R(d\lambda)$ where R assigns mass two to $\lambda = 0, \pi$. Then

$$\sigma(\theta) = \int_C \int_C g(\theta - \lambda - \gamma)\mu(d\lambda)R(d\gamma).$$

This is the required representation (the bivariate measure integrating the kernel could in principle be reduced to a univariate one). Writing out (ISO) and simplifying yields (II).

REFERENCES

ANDERSON, T. W. (1955). The integral of a symmetric unimodal function over a symmetric convex set and some probability inequalities. *Proc. Amer. Math. Soc.* *6* 170–176.

BLASCHKE, W. (1949). *Kreis und Kugel*, Chelsea, New York.

GRENANDER, U. (1976). *Pattern Synthesis—Lectures in Pattern Theory, Vol. 1*, Springer-Verlag, New York.

LAGARIAS, J. C. (1982). The van der Waerden conjecture: two Soviet solutions. *Notices Amer. Math. Soc.* *29* 130–133.

STANLEY, R. (1981). Two combinatorial applications of the Alexandrov-Fenchel inequalities. *J. Comb. Th. Ser. A.* *31* 56–65.

TONG, Y. L. (1980). ***Probability Inequalities in Multivariate Distributions,*** Academic, New York.

VITALE, R. A. (1974). Representation of convex, compact subsets of the plane, Report on Pattern Analysis #22, Brown University, Providence, RI.

VITALE, R. A. (1979). Support functions of plane convex sets, Tech. Report, Claremont Graduate School, Claremont, CA.

Inequalities in Statistics and Probability
IMS Lecture Notes–Monograph Series Vol. 5 (1984), 112–114

AN EXPANSION FOR SYMMETRIC STATISTICS AND THE EFRON-STEIN INEQUALITY

BY RICHARD A. VITALE
Claremont Graduate School

The Efron-Stein inequality and a generalization by Bhargava are derived using a tensor-product basis and bounds for covariances of related symmetric statistics.

1. Introduction. Let $S(X_1, \ldots, X_n)$ be a symmetric function of its iid arguments. Its variance can be estimated by the jackknife technique as follows: assuming an augmented iid collection $X_1, \ldots, X_n, X_{n+1}$, form $S_i = S(X_1, \ldots, X_{i-1}, X_{i+1}, \ldots, X_{n+1})$, $i = 1, \ldots, n+1$ and $\bar{S} = (n+1)^{-1}\Sigma_{i=1}^{n+1} S_i$. Then $\mathrm{Var}\, S(X_1, \ldots, X_n)(=\mathrm{Var}\, S_i)$ is estimated by $Q = \Sigma_{i=1}^{n+1} (S_i - \bar{S})^2$. As part of an extensive study, Efron and Stein (1981) showed that Q is necessarily positively biased, an observation that has come to be known as the *Efron-Stein inequality*.

THEOREM 1.

$$\mathrm{Var}\, S(X_1, \ldots, X_n) \leq EQ \tag{1.1}$$

with equality iff S is linear in functions of its individual arguments.

Other proofs and extensions have been given by Bhargava (1980) and Karlin and Rinott (1982), and the inequality has already had interesting applications (Hochbaum and Steele (1982), Steele (1981), Steele (1982)). Our purpose here is to derive the inequality by using an idea exploited for other purposes in Rubin and Vitale (1980): expansion of symmetric statistics in a tensor-product basis. The approach yields attractive, concrete representations and is particularly well-adapted to proving the E-S inequality by first establishing a universal bound on the covariance of related symmetric statistics. It is an alternative to the ANOVA-type expansions used elsewhere.

2. The Efron-Stein Inequality via Covariance Bounds. If $e_0(X_1) \equiv 1$, $e_1(X_1)$, $e_2(X_1), \ldots$ form an orthonormal basis for the square integrable functions of X_1, then products of the type $\Pi_{i=1}^{n} e_{\nu_i}(X_i)$ form an orthonormal basis for the square integrable functions of $\mathbf{X} = (X_1, \ldots, X_n)$. For ease of notation we denote the above product by $e_\nu(\mathbf{X})$, $\nu = (\nu_1, \ldots, \nu_n)$.

THEOREM 2. *For $i \neq j$,*

$$0 \leq \mathrm{Cov}(S_i, S_j) \leq ((n-1)/n)\,\mathrm{Var}\, S_1 \tag{2.1}$$

with equality above iff S_1 is linear in functions of its individual arguments.

Proof. Without loss of generality, assume that the S_i (which are identically distributed) have zero mean. Accordingly, we consider ES_1S_{n+1} as a surrogate for $\mathrm{Cov}(S_i, S_j)$, $i \neq j$. Using the basis given above and symmetry considerations yields

AMS 1980 subject classifications. Primary 60E15, secondary 62H20, 62J10.
Key words and phrases: ANOVA-type expansion, covariance bounds, Efron-Stein inequality, symmetric statistic, tensor-product basis, U-statistic.

$$S_1 = S(X_2, \ldots, X_n, X_{n+1}) = \Sigma\, c_\nu e_\nu(\mathbf{X}), \text{ where } \mathbf{X} = (X_2, \ldots, X_n, X_{n+1}),$$

and

$$S_{n+1} = S(X_1, \ldots, X_n) = S(X_2, \ldots, X_n, X_1) = \Sigma c_\nu e_\nu(\mathbf{X}'), \text{ where } \mathbf{X}' = (X_2, \ldots, X_n, X_1).$$

Then

$$ES_1 S_{n+1} = E\Sigma c_\nu e_\nu(\mathbf{X})\, \Sigma\, c_\mu e_\mu(\mathbf{X}') = \Sigma\, c_\nu c_\mu E e_\nu(\mathbf{X}) \mathbf{e}_\mu(\mathbf{X}').$$

The expectation of $e_\nu(\mathbf{X})e_\mu(\mathbf{X}')$ is zero unless $\nu = \mu$ with $\nu_n = \mu_n = 0$, in which case it is unity. Thus $ES_1S_{n+1} = \Sigma_{\nu_n=0}\, c_\nu^2$, which displays the asserted positive correlation.

For the upper bound, we symmetrize: note that generally for summands $\{\sigma_\nu\}$ which are symmetric in ν

$$\Sigma_{\nu_n=0}\sigma_\nu = n^{-1}\, \Sigma_\nu z_\nu \sigma_\nu$$

where z_ν is the number of zero components in ν. The $\{c_\nu\}$ may be assumed symmetric in ν and hence

$$ES_1 S_{n+1} = n^{-1}\, \Sigma_\nu z_\nu c_\nu^2.$$

Now $z_\nu c_\nu^2 \leq (n-1)c_\nu^2$ for every ν because of the centering of the S_i, which leads to

$$ES_1 S_{n+1} \leq ((n-1)/n)\, \Sigma_\nu c_\nu^2 = ((n-1)/n) \operatorname{Var} S_1.$$

Equality occurs iff $z_\nu = n-1$ for all non-vanishing c_ν. This means that

$S_{n+1} = f(X_1) + \ldots + f(X_n)$ for some f. □

Returning to the Efron-Stein inequality, we note that expanding EQ in (1.1) yields

$$\operatorname{Var} S_1 \leq n \operatorname{Var} S_1 - n \operatorname{Cov}(S_1, S_{n+1})$$

which, upon rearrangement, is the upper inequality in (2.1).

3. A Higher-Order Construction. A natural question to ask is whether a more ample supply of randomness can lead to other estimates and inequalities. Specifically, suppose that S is a symmetric function of n iid. arguments which can now be chosen form $X_1, X_2, \ldots, X_N$ where $n < N$ ($N = n+1$ in the previous section). Proceeding by analogy, for $A = \{\nu_1, \nu_2, \ldots, \nu_n\}$ with distinct $\nu_i \in \{1, 2, \ldots, N\}$, define $S_A = S(X_{\nu_1}, X_{\nu_2}, \ldots, X_{\nu_n})$ and $\bar{S} = \binom{N}{n}^{-1}\Sigma_{|A|=n} S_A$. Then an estimate for Var S_A is $Q = \binom{N-1}{n}^{-1}\Sigma_{|A|=n}(S_A - \bar{S})^2$. This is the set-up studied by Bhargava (1980), who showed that positive bias obtains here as well.

THEOREM 3. *Var $S_A \leq EQ$ with equality iff. S_A is linear in functions of its individual arguments.*

In treating this problem, we establish bounds on covariances as before. These generalize theorem 2 and show that the upper bound is linear in the number of shared arguments (cf. Bhargava (1980, p. 6)).

THEOREM 4. *For $|A \cap A'| = k$, $0 \leq \operatorname{Cov}(S_A, S_{A'}) \leq (k/n) \operatorname{Var} S_A$ with equality above iff S_A is linear in functions of its individual arguments.*

Proof. The argument parallels that of theorem 2; assuming zero mean, we compute $ES'S''$ where

$$S' = S(X_1, \ldots, X_k, Y_{k+1}, \ldots, Y_n), \quad S'' = S(X_1, \ldots, X_k, Z_{k+1}, \ldots, Z_n)$$

(the X,Y,Z variables taken together are iid.). This gives $ES'S'' = \Sigma' c_\nu^2$ where Σ' denotes summation over subscripts ν with vanishing final $n-k$ components. This can be symmetrized to the form

$$ES'S'' = \binom{n}{k}^{-1}\Sigma_\nu \binom{z_\nu}{n-k} c_\nu^2$$

where z_ν is the number of zero components of ν.

This is clearly non-negative and noting that $z_\nu c_\nu^2 \leqslant (n-1)c_\nu^2$ yields the upper bound with the condition for equality. □

Theorem 3 follows directly from the upper bound just given. We merely sketch some important points. In computing EQ, sums of the form $\Sigma_A ES_A S_{A'}$, intervene and calculate out to

$$\Sigma_{k=0}^n \binom{n}{k}\binom{N-n}{n-k}\left[\binom{n}{k}^{-1}\Sigma_\nu \binom{z_\nu}{n-k} c_\nu^2\right],$$

the bracketed quantity being the exact value of the covariance in theorem 4. This leads to

$$EQ = \Sigma_\nu c_\nu^2 \Sigma_{k=0}^n (N/(N-n)) \binom{N-n}{n-k}\binom{N}{n}^{-1}\left[\binom{n}{k} - \binom{z_\nu}{n-k}\right],$$

and a collapse to the lower bound $\Sigma_\nu c_\nu^2 = \operatorname{Var} S_A$.

Acknowledgments. I thank Mike Steele for suggesting the tensor-product basis approach to the Efron-Stein inequality.

REFERENCES

BHARGAVA, R. P. (1980). A property of the jackknife estimation of the variance when more than one observation is omitted. Tech. Report No. 140, Dept. Statist., Stanford Univ.

EFRON, B. and STEIN, C. (1981). The jackknife estimate of variance. *Ann. Statist. 9* 586–596.

HOCHBAUM, D. and STEELE, J. M. (1982). Steinhaus's geometric location problem for random samples in the plane. *Adv. Appl. Prob. 14* 56–67.

KARLIN, S. and RINOTT, Y. (1982). Applications of ANOVA type decompositions for comparisons of conditional variance statistics including jackknife estimates. *Ann. Statist. 10* 485–501.

RUBIN, H. and VITALE, R. A. (1980). Asymptotic distribution of symmetric statistics. *Ann. Statist. 8* 165–170.

STEELE, J. M. (1981). Complete convergence of short paths and Karp's algorithm for the tsp. *Math. Oper. Res. 6* 374–378.

STEELE, J. M. (1982). Optimal triangulation of random samples in the plane. *Ann. Probab. 10* 548–553.

Inequalities in Statistics and Probability
IMS Lecture Notes–Monograph Series Vol. 5 (1984), 115–120

ON CHEBYSHEV'S OTHER INEQUALITY

BY A. M. FINK and Max JODEIT, JR.
Iowa State University and University of Minnesota

We formulate the notion of a best possible inequality. This involves finding the largest class of functions and measures for which an inequality is true. We give two examples of Chebyshev's inequality, e.g. $\int_a^b d\mu \int_a^b fg d\mu \geq \int_a^b f d\mu \int_a^b g d\mu$ for all pairs (f,g) which are increasing if and only if $\int_a^x du \geq 0$, $\int_x^b du \geq 0$ for all x. Other examples include Jensen's inequality.

1. Introduction. Let μ be a probability measure on the real line and f and g increasing functions. Then

$$\int_{\mathbb{R}} fg \, d\mu \geq \int_{\mathbb{R}} f d\mu \int_{\mathbb{R}} g \, d\mu \tag{1.1}$$

says that the random variables f and g are positively correlated. This is Chebyshev's 'other' inequality.

It is common to ask if an inequality is best possible. In most instances this means having the largest (or smallest) constant(s) for which the inequality holds and settling the cases of possible equality. For (1.1) equality holds if one of the functions is a constant or the measure is a point mass.

In this paper we would like to explore a different meaning of 'best possible.' In order to formulate our ideas in the context of inequality (1.1), consider a related version

$$\int_a^b d\mu \int_a^b fg \, d\mu \geq \int_a^b f d\mu \int_a^b g \, d\mu \tag{1.2}$$

where $[a,b]$ is any real interval. It was already observed by Andreief (1883), that (1.2) holds under the hypothesis that

$$[f(x) - f(y)][g(x) - g(y)] \geq 0 \quad \text{for all } (x,y) \in [a,b] \times [a,b], \tag{1.3}$$

and

$$\mu \text{ is a non-negative measure.} \tag{1.4}$$

The condition (1.3) is read "f and g are similarly ordered," see Hardy, Littlewood, and Pólya ((1952), p. 43). (More history of the inequality (1.2) appears in the article by Mitrinović and Vasić (1974).) It is clear that (1.3) is satisfied if both f and g are increasing.

Our viewpoint is that the inequality (1.2) has "two variables," the pairs of functions and the measures. 'Best possible' should mean that:

(A) the inequality (1.2) holds for all similarly ordered pairs if and only if μ is a non-negative measure, and

(B) the inequality (1.2) holds for all non-negative measures if and only if f and g are similarly ordered.

We will show below that both statements are correct. This means that each class, similarly ordered pairs, and non-negative measures, is the largest class for which the inequality can be proved, given that it must hold for all elements in the other class.

Contrast this with the condition

AMS 1980 subject classifications. 26D15, 60E15, 62H20.
Key words and phrases: Chebyshev, inequalities, correlation.

(1.5) f, g are both increasing or both are decreasing, or one is a constant.

The set of pairs satisfying (1.5) is smaller than the set of similarly ordered pairs, so the requirement that (1.2) holds for such pairs is a less restrictive condition on the measure μ which is to satisfy it. In fact, for some signed measures μ inequality (1.2) holds for all pairs satisfying (1.5). We will derive a condtion to replace $\mu \geqslant 0$ for which the appropriate versions of (A) and (B) hold.

In summary, what we are looking for is a class of measures M so that the following statement is true.

"The inequality (1.2) holds for all pairs f, g satisfying (1.5) if and only if $\mu \in M$."

Thus M is the largest class of measures for which we can prove the inequality (1.2) under the conditions (1.5). It is then natural to ask for the largest class of pairs of functions for which inequality (1.2) holds for all measures in M. Our view is that 'best possible' should mean that the conditions are those in (1.5).

In the succeeding sections we formulate this question in general and give several examples of this phenomenon.

2. Function-Measure Duality. Let F be a class of functions and M be a class of measures and $J(f) \geqslant 0$ be an integral inequality.

The classes F and M are said to be in duality with respect to J if (I) $J(f) \geqslant 0$ for all $f \in F$ if and only if the measure is in M, and (II) $J(f) \geqslant 0$ for all measures in M if and only if $f \in F$. This is analogous to the situation, in the theory of locally convex topological linear spaces, leading to weak and weak* topologies on $\mathscr{B}$ and $\mathscr{B}^*$. Each formulation of classes in duality requires specifying the universe of functions or measures. For us here, all functions are to be Borel measurable and all measures are regular Borel (signed) measures.

It is important to notice that if the integrals are linear in the measure, the class M being as large as possible will be a cone. If the inequality is convex in the functions, then F will also be a cone.

As a simple example, the class F_+ of non-negative functions on an interval $[a,b]$ are in duality with the class M_+ of non-negative measures with respect to the inequality

$$\int_a^b f d\mu \geqslant 0. \tag{2.1}$$

The proof is straightforward and is omitted.

The inequality (2.1) is a point of contact with known theory. If C is a cone of functions in a Banach space X, then the measures $M \subset X^*$ for which (2.1) holds is called the conjugate cone C^*, see Kelly and Namioka (1963). If then we look at the class of functions in X for which (2.1) holds for all measures in C^*, then this class might be called $^*(C^*)$. We are interested in cones C for which $^*(C^*) = C$.

As a second example, let $(FI)_+$ be the class of non-negative increasing functions on $[a,b]$ and M_0 be the class of measures μ such that $\int_x^b d\mu \geqslant 0$ for all $x \in [a,b]$. Then $(FI)_+$ and M_0 are in duality with respect to the inequality (2.1). The proof is omitted.

3. Chebyshev's Inequality. We return to the inequality (1.2) which is our main motivation. Observe that if μ is replaced by $-\mu$ the inequality is unchanged. For the two theorems on this inequality we will assume that μ is somewhere positive.

THEOREM 1. *Let SO be the pairs of functions which are similarly ordered and M_+ the set of non-negative measures on $[a,b]$. Then SO and M_+ are in duality with respect to the* *inequality (1.2). Equality holds for a pair in SO and a measure in M_+ if and only if one of the functions is constant a.e. μ.*

Proof. The inequality

$$\tfrac{1}{2}\int_a^b\int_a^b [f(x)-f(y)][g(x)-g(y)]d\mu(x)d\mu(y) \geqslant 0 \tag{3.1}$$

is obvious under the assumptions that the pair (f,g) are similarly ordered (see (1.3)) and $\mu \geqslant 0$. This establishes the sufficiency in both (I) and (II), since if the expression in (3.1) is expanded one gets (1.2). For the necessity, let $\mu = \delta_x + \delta_y$, $x \neq y$. Then the inequality (1.2) is exactly (1.3). Finally, to show that $\mu \geqslant 0$ if (1.2) holds for all similarly ordered pairs, take $f = g = X_{[c,d]}$, the indicator function of an interval $[c,d]$. Then (1.2) is $\mu[a,b]\mu[c,d] \geqslant (\mu[c,d])^2 \geqslant 0$. Thus μ is always the same sign. Since μ is somewhere positive, it is positive everywhere.

To establish the cases of equality, observe that if A and B are any sets of μ positive measure then (3.1) with equality implies that $[f(x)-f(y)][g(x)-g(y)] = 0$, $x \in A$, $y \in B$. Assume that f is not a constant a.e. μ. If further one assumes that $f^{-1}\{\alpha\}$ is always a set of μ measure zero, let A and B be disjoint positive measure sets. Then $f(x)-f(y) \neq 0$ a.e. $x \in A$, $y \in B$ and thus $g(x) \equiv g(y)$ a.e. $x \in A$, $y \in B$. This implies that g is a constant. If there is an α so that $f^{-1}(\alpha) = A$ has positive measure, then take B to be the complement of A and the above argument gives g a constant. □

For our second example on Chebyshev's inequality we consider the class SM (for similarly monotone), those pairs of functions satisfying (1.5). The corresponding measures EP (for end-positive) are those measures μ such that

$$\int_a^x d\mu \geqslant 0 \quad \text{and} \int_x^b d\mu \geqslant 0 \quad \text{for all } x \in [a,b]$$

and $\int_a^b d\mu \neq 0$. This class will reappear in a different example.

THEOREM 2. *The pair SM and EP are in duality with respect to inequality (1.3). Equality holds if and only if (when f,g are right continuous)*

(+) sup supp $R\,dg \leqslant$ inf supp $L\,df$, *and* sup supp $R\,df \leqslant$ inf. supp $L dg$.

If either pair of supports meet, their common point is a set of measure zero for at least one of the measures.

Here sup $\phi \equiv a$, *inf* $\phi \equiv b$, $R(x) = \int_x^b d\mu$ *and* $L(x) = \int_a^x d\mu$.

Proof. We will show the sufficiency of (I) and (II) by writing the inequality in the form (3.1). Assume first that f is increasing and right continuous. Then there is a non-negative measure μ_1 so that for $x < y$,

$$f(y)-f(x) = \mu_1[a,y] - \mu_1[a,x] = \mu_1(x,y] = \int \chi_{(x,y]}(t)d\mu_1(t).$$

Then (3.1) can be written as (μ_2 the measure for g.)

$$\tfrac{1}{2}\int_a^b\int_a^b (f(y)-f(x))(g(y)-g(x))d\mu(x)d\mu(y)$$
$$= \int_a^b\int_a^b (y-x)_+^0 \int_a^b \chi_{(x,y]}(t)d\mu_1(t) \int_a^b \chi_{(x,y]}(s)d\mu_2(s)d\mu(x)d\mu(y)$$
(integrand is 0 on the diagonal)
$$= \int_a^b\int_a^b\int_a^b\int_a^b (y-x)_+^0\, \chi_{(x,y]}(t)\, \chi_{(x,y]}(s)d\mu(x)d\mu(y)d\mu_1(t)d\mu_2(s).$$

The part of the integrand involving x is expressible as

$$\chi(y \wedge t \wedge s > x)\chi(t \vee s \leqslant y).$$

where χ (P(variables)) denotes the indicator functions of the set of variables for which P(variables) is true. This gives

$$\int_a^b\int_a^b\int_a^b \chi(t \vee s \leqslant y)L(y \wedge t \wedge s)d\mu(y)d\mu_1(t)d\mu_2(s)$$
$$= \int_a^b\int_a^b\int_a^b L(t \wedge s)X(t \vee s \leqslant y)d\mu(y)d\mu_1(t)d\mu_2(s)$$
$$= \int_a^b\int_a^b L(t \wedge s)R(t \vee s)d\mu_1(t)d\mu_2(s),$$

where $L(x) = \mu([a,x))$, $R(x) = \mu([x,b])$. Thus (3.1) is equivalent to

$$\int_a^b\int_a^b L(t \wedge s)\mathrm{R}(t \vee s)d\mu_1(t)d\mu_2(s) \geqslant 0. \tag{3.2}$$

This is clearly non-negative. Since any increasing function is the pointwise limit of right continuous functions, the sufficiency is shown.

Conversely, if (3.1) holds for all $(f,g) \in$ SM, then as above, $\int_a^b\int_a^b L(t \wedge s)R(t \vee s)d\mu_1(t)d\mu_2(s) \geqslant 0$ for all non-negative measures μ_1, and μ_2. Thus $L(x_1)R(x_2) \geqslant 0$ if $x_1 \leqslant x_2$. Since $L(x) + R(x) \equiv \mu(I)$ we have $L(x)R(x) \geqslant 0$. If $\mu(I) \neq 0$, $L(x)$ and $R(x)$ have the same sign as $\mu(I)$ unless one of them is 0. The case $\mu(I) = 0$ cannot occur unles $\mu = 0$, for then $0 \geqslant \int f\, d\mu \int g\, d\mu$ for all (f,g) in SM; thus taking $f = g = X_{[a,x)}$ and $f = g = Z_{[a,x]}$ lead to $\mu((x_1,x_2)) = 0$ if $x_1 < x_2$, so $\mu = 0$.

Next we need to show that if (1.2) holds for all end-positive measures μ, the pair (f,g) is in SM.

It may help the exposition to let a, b, c, etc. denote the values taken by f at x, y, z, etc. respectively, and A, B, C and so on denote the corresponding values of g. We need two observations.

First, non-negative measures are end-positive, so f and g are similarly ordered (Theorem 1); i.e., $(a-b)(A-B) \geqslant 0$ for all pairs x,y in I.

If $x < y < z$ the measure $\delta_x - \delta_y + \delta_z$ is end-positive, so by (1.3)

$$aA - bB + cC \geqslant (a-b+c)(A-B+C).$$

This is equivalent to

$$0 \geqslant (a-b)(C-B) + (c-b)(A-B).$$

Neither product can be positive—the other would be negative, all four differences would be non-zero, and similar ordering would imply that both products have the same sign (sgn $(C - B) = \operatorname{sgn}(c - b)$, etc.). Thus we have the second observation: for all triples $x < y < z$, $(a-b)(C-B) \leqslant 0$, $(c-b)(A-B) \leqslant 0$. Note that x and z are "separated."

It is enough to show that $(a - b)(c - b) \leqslant 0$, for all triples $x < y < z$, unless g is constant (i.e. $f(y)$ is between $f(x)$ and $f(z)$ if $x < y < z$). Suppose not. Then for some triple $x_0 < y_0 < z_0$, $(a_0 - \mathrm{b}_0)(c_0 - \mathrm{b}_0) > 0$.

The key argument is this: if $a_0 < b_0$ then by the second observation $(a_0 - \mathrm{b}_0)(C_0 - \mathrm{B}_0) \leqslant 0$ if $z > y_0$, so $C \geqslant B_0$. Similarly, $A \geqslant B_0$ if $x \leqslant y_0$, so $g(y_0)$ is a global minimum for g. Moreover, $A_0 = B_0 = C_0$. If not, say $A_0 > B_0$. Then $(A_0 - B_0)(a_0 - b_0) < 0$, which contradicts similar ordering. In case $a_0 > b_0$ we get that $g(y_0)$ is a global maximum for g and $A_0 = B_0 = C_0$.

Next (still supporting $a_0 < b_0$) we show $A = A_0$, $x \leqslant x_0$, and $C = C_0$, $z \geqslant z_0$. For, if $A > B_0$ for some $x < x_0$, similar ordering gives $a \geqslant b_0 > a_0$. Apply the "$a_0 > b_0$" case of the key argument to a, a_0, b_0 (the triple being $x < x_0 < y_0$). It gives $A = A_0 = B_0$, which contradicts $A > B_0$. Similarly, $C = C_0$ if $z > z_0$.

If g were not constant there would exist y, $x_0 < y < z_0$, such that $B > A_0 = B_0 = C_0$. We may suppose $x_0 < y < y_0$. Now apply the key argument, with the roles of f and g interchanged, to conclude that $a_0 = b = b_0$, which contradicts $a_0 < b_0$.

Since (f,g) and μ satisfy (1.2) if and only if $(-f,-g)$ and μ do, f is monotone unless g is constant. It follows, using similar ordering, that $(f,g) \in$ SM. □

Finally, we discuss the case of equality. If f or g is constant, equality holds, and so do the conditions (+). We shall assume that neither f nor g is constant. We may assume, too, that each is non-decreasing and non-negative (a constant added to f or g adds equal quantities to both sides).

If $\int d\mu = 0$ then we know $\mu = 0$ (for then $0 \geq (\int f d\mu)^2$, where $f = y = \chi_{[a,x)}$, so $L(x) \equiv 0$). Thus we assume $\int d\mu \neq 0$, and we may assume $\int \delta\mu > 0$.

Note that, if $\mu \geq 0$, the case of equality is covered in Theorem 1.

A case of interest, which we largely ignore, is that in which $\int f\, g\, d\mu = 0$ and, say, $\int f\, d\mu = 0$. We shall only consider this under the foregoing assumptions, and the further assumption that f,g are right-continuous. Thus $f(x) = \nu_1[a,x]$, $g(x) = \nu_2[a,x]$, where ν_1, ν_2 are non-negative. Then

$$\int_a^b f d\mu = \int_a^b \int_a^x d\mu_1(t) d\mu(x) = \int_a^b R(t) d\nu_1(t) = 0,$$

so $R d\nu_1 = 0$, and the first condition (+) holds. Also,

$$\begin{aligned}\int_a^b fg\, d\mu &= \int_a^b \int_a^x d\nu_1(t) \int_a^x d\nu_2(s) d\mu(x) = \int\int R(t \vee s) d\nu_1(t) d\nu_2(s) \\ &= \int\int_{t \leq s} R(s) d\nu_1(t) d\nu_2(s) = \int_a^b f(s) R(s) d\nu_2(s) = 0,\end{aligned}$$

so $fR d\nu_2 = 0$. This implies inf supp $f \geq$ sup supp $R d\nu_2$, which gives the second condition in (+). A moment's reflection gives that the third part of (+) also holds. We now assume that none of the integrals in the equality is zero.

Let us verify that (+) is equivalent to equality, if f,g are right-continuous, non-increasing, non-negative, not constant, $\int d\mu > 0$, and none of the integrals in the equality is zero. With $f(x) = \nu_1[a,x]$, $g(x) = \nu_2[a,x]$, we have, as in the sufficiency argument, that

$$\begin{aligned} 0 &= \int\int L(s \wedge t) R(s \vee t) \delta\nu_1(t) d\nu_2(s) \text{ (the limits on the integrals are } a+0 \text{ and } b) \\ &= \int\int_{s \geq t} L(t) R(s) d\nu_1(t) d\nu_2(s) + \int\int_{s<t} L(s) R(t) d\nu_1(t) d\nu_2(s) \\ &= \int_{a+0}^b \int_{a+0}^s L(t) d\nu_1(t) \cdot R(s) d\nu_2(s) + \int_{a+0}^b \int_{a+0}^{t-0} L(s) d\nu_2(s) \cdot R(t) d\nu_1(t), \end{aligned}$$

so both terms are zero. Since the roles of ν_1 and ν_2 can be reversed, the last term is still zero if t–0 is replaced by t. Now (e.g.) $\int_{a+0}^s L(t) d\nu_1(t) \geq 0$ is non-decreasing and $R d\nu_2 \neq 0$, so $L d\nu_{3-i} = 0$ in $[a, c_i)$, where $c_i = \sup(\operatorname{supp} R d\nu_i)$, $i = 1,2$. This gives the first two parts of (+). The last part of (+) follows because if (e.g.) $R(c_2)\nu_2\{c_2\} > 0$, then $0 = \int_{a+0}^{c_2} L(t) d\nu_1(t) \geq L(c_2)\nu_1\{c_2\}$. Finally, if (+) holds the last iterated integrals are both 0, so equality holds.

If f and g are not right-continuous, we can write $f = f_0 + j$, $g = g_0 + k$, where f_0, g_0 are right-continuous and j,k are left-continuous jump functions. Then equality holds for each pair (f_0,g_0), (f_0,k), (j,g_0), (j,k). The conditions (+) are changed by replacing $R(t)$, $L(t)$ by $R(t+0)$, $L(t+0)$ when they appear with dj or dk.

We remark that if R and L are both positive in (a,b), and continuous, equality can only happen if one of f,g is constant.

4. Further Results and Problems. In the foregoing, it was essential to have a suitable integral representation for the functions in the class F. Appropriate manipulations then permitted reduction to the fundamental inequality (2.1). We state two more examples of such results here, with problems we hope are of interest.

We let M_p denote the class of functions f with the representation

$$f(x) = \int_0^\infty (x-t)_+^p d\nu(t), \quad 0 \leq x \leq T \leq \infty,$$

where $x_+ = \max(0,x)$, $(x-t)_+^0$ means $\chi_{[t,\infty)}(x)$, ν is a non-negative measure, and $p \geq 0$.

THEOREM 3. *If p is an integer, M_p is in duality with the set M_p^* of signed measures μ which satisfy*

$$\int_0^T (x-t)_+^p \, d\mu(x) \geq 0, \quad 0 \leq t \leq T,$$

with respect to the inequality (2.1).

The proof uses the basic spline of Curry and Schoenberg (see Schoenberg (1973), p. 3), and a representation theorem of Bernstein (1926).

Problem: If $p > 0$ is not an integer, prove this.

The difficulty is in proving the representation.

Remark. This result, with $p = 1$, can be used to prove Theorem 108, page 89, in Hardy, Littlewood and Pólya (1952).

The second result concerns Jensen's inequality:

$$\varphi(\int f d\mu / \int d\mu) \leq \int \varphi(f) d\mu / \int d\mu.$$

With appropriate modification of "best possible" in our sense (here we have three "variables"), this inequality is best possible for convex φ, Borel measurable functions, and non-negative measures.

If we restrict f to be monotone, we have:

THEOREM 4. *The convex functions are in duality with the end-positive measures with respect to Jensen's inequality, when F is the class of monotone functions.*

The conditions for equality are too lengthy to be stated here.

Elsewhere we have shown that

$$\varphi(\int_0^1 f d\sigma / \int_0^1 d\sigma) \leq [(\int_0^1 f d\sigma)^p / \int_0^1 f^p \, d\sigma] \int_0^1 w(f) d\sigma / \int_0^1 d\sigma$$

holds for all f monotone and $\varphi \in M_p$ if and only if σ is end-positive, with equality for $\varphi(x) = x^p$.

Problem. Is this "best possible" in the sense of this paper?

REFERENCES

ANDRĔIEF, C. (1883), Note sur une relation entre les intégrales définies des produits des fonctions, *Mém. Soc. Ci. Bordeaux 3 2*, 1–14.

BERNSTEIN, S. (1926). *Lecons sur les propriétes extrémales et la meilleure approximation des fonctions analytiques d'une variable réelle*, Paris.

FINK, A. M. (1980), Pointwise inequalities and continuation of solutions of an n[th] order equation, *Tohoku Math. Journal 32*, 169–175.

FINK, A. M. and Jodeit, Max (1976), A generalization of the arithmetic-geometric means inequality, *Proc. Amer. Math. Soc. 61*, 255–261.

HARDY, LITTLEWOOD, PÓLYA (1952), *Inequalities*, 2nd ed. Cambridge University press, London and New York.

KELLEY, J., NAMIOKA (1963), *Linear Topological Spaces*, D. Van Nostrand, Princeton, New Jersey.

MITRINOVIĆ. D. S. and VASIĆ, P. M. (1974), History, Variations and Generalizations of the Cebysev Inequality and the Question of Some Priorities, *Publ. Elek. Fak. Univ. Beograd. No. 461–497*, 1–30.

RIESZ, F. and NAGY, B. (1955), *Functional Analysis*, Frederick Ungar Publ., New York.

SCHOENBERG, I. J. (1973), Cardinal Spline Interpolation, *Regional Conf. Ser. in Applied Math. 12* SIAM, Philadelphia, Pa.

Inequalities in Statistics and Probability
IMS Lecture Notes–Monograph Series Vol. 5 (1984), 121–126

REMARKS AND OPEN PROBLEMS IN THE AREA OF THE FKG INEQUALITY

By KUMAR JOAG-DEV,[1] L. A. SHEPP and RICHARD A. VITALE
University of Illinois, Bell Laboratories and Claremont Graduate School

The FKG inequality is an effective device when the requisite assumptions can be verified. Sometimes these have to be approached circuitously. This is discussed with reference to past uses and suggestions for work on the range of applicability. New areas of potential application are also presented.

1. Sufficiency and Necessity of the Conditions for the FKG Inequality. The FKG inequality in its original form (Fortuin, Ginibre and Kasteleyn (1971)) states that if (a) Γ is a distributive lattice i.e. order isomorphic to an algebra of subsets of a set, (b) f and g are increasing on Γ, (c) μ is a positive function on Γ with

$$\mu(x)\mu(y) \leqslant \mu(x \wedge y)\mu(x \vee y) \quad \text{for all } x,y, \tag{1.1}$$

then

$$\Sigma f(x)\mu(x)\Sigma g(y)\mu(y) \leqslant \Sigma f(x)g(x)\mu(x)\Sigma\mu(y). \tag{1.2}$$

A simple example of how the FKG inequality can be used in a combinatorial setting is the following. Suppose A, A_i are fixed subsets of $N = \{1, \ldots, n\}$ and k, k_i are given integers, $i = 1, \ldots, r$. Choose a subset of S of N at random by choosing each element to be in S independently with probability p, fixed. Let $\bar{A}_i = |A_i \cap S|$. Then

$$P[\bar{A} \geqslant k | \bar{A}_i \geqslant k_i, i \leqslant r] = a_r \geqslant P[\bar{A} \geqslant k] = a_0.$$

To prove this let Γ be the set of all subsets S of N ordered by inclusion, and let $f(S) = \chi(\bar{A}_i \geq k_i, i \leq r)$, $g(S) = \chi(\bar{A} \geq k)$, and $\mu(S) = 1$. It is easy to verify that (a)–(c), (1.1) hold and this gives the result. The result may not seem surprising until it is realized that a_r is not always increasing in r. Indeed with $n = 2$, $A = \{1\}$, $A_1 = \{1,2\}$, $A_2 = \{2\}$ with $p = 1/2$ gives a counterexample since $a_0 = 1/2 < a_1 = 2/3 > a_2 = 1/2$. This class of problems was posed by Frank Hwang and will be further developed elsewhere.

We will see that FKG is often hard to apply even when one feels it should apply. This may also be illustrated by Hwang's example: It can be shown by a direct argument that

$$P[\bar{A}_i \geqslant k_i, i \leqslant r | \bar{A} \geqslant k] \geqslant P[\bar{A}_i \geqslant k_i, i \leqslant r | \bar{A} = k],$$

But Shepp does not see just now how to give an FKG proof. The obvious choice $g(S) = \chi(\bar{A} \leqslant k)$, $\mu(S) = \chi(\bar{A} \geqslant k)$, and f as before yields the desired conclusion but (1.1) fails. Is there a reordering of Γ to make an FKG proof?

FKG themselves point out that (1.1) is not necessary and one could assume the alternate condition

$$2\mu(I)\mu(O) \geqslant \Sigma'\mu(x)\mu(y) \tag{1.1'}$$

---

[1] Research sponsored by the Air Force Office of Scientific Research, Air Force Systems Command under Grant No. 82–K–0007. The United States Government is authorized to reproduce and distribute reprints for government purposes not withstanding any copyright notation thereon.

AMS 1980 subject classifications. Primary 60E15, Secondary 62H20, 06D99

Key words and phrases: Association, distributive lattice, FKG inequality, positive quadrant dependence, total ordering.

which may hold without (1.1), where I and O are the extreme elements of Γ and the sum is over all other elements.

Shepp (1982) shows by an example the slack in (a)–(c), (1.1), where by merely redefining $<$ in the lattice one gets to satisfy (a)–(c), (1.1) and hence obtain (1.2) where the "natural" ordering fails to satisfy (a)–(c), (1.1). Note (1.2) does not depend on "$<$". Ahlswede and Daykin (1978) and others (see survey on FKG by Graham (1983)) give more general versions, also not necessary.

The question then arises as to whether it is possible to find conditions which are necessary or at least closer to being necessary. In this regard, it may be interesting to note that for a distributive lattice of length 2, (1.1) and (1.1′) are equivalent and necessary. In this case, there are two possible structures:

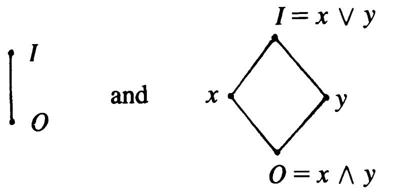


FIGURE 1. Structures for a 2-length lattice.

The first is easily seen. For the second, take two increasing functions f and g with $f(O) = f(y) = g(O) = g(x) = O$ and $f(x) = f(I) = g(y) = g(I) = I$. Then f and g are positively correlated $\Leftrightarrow$ (1.1) $\Leftrightarrow$ (1.1′).

Neither (1.1) nor (1.1′) is necessary for lattices of length greater than 2, but it is interesting to see heuristically why they are sufficient. The clue lies in the assumption of a *distributive* lattice. Among other things, this means that certain sublattice structures do not occur and, in fact, that locally the lattice looks like the pictured length 2 cases. Thus (1.1) and the distributivity assumption are paired to ensure that things work locally.

As mentioned, this approach is generally too strong. One point of departure for an alternate approach is to hold the distributivity assumption in abeyance. This allows previously forbidden sublattice structures, which we can view as lattices in their own right:

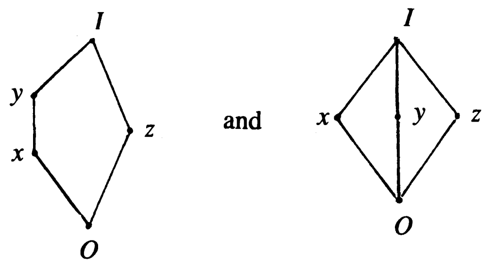


FIGURE 2. Structures for a (nondistributive) lattice.

Consider, for the first structure, increasing f and g with $f(O) = f(z) = g(O) = g(x) = g(y) = O$ and $f(x) = f(y) = f(I) = g(z) = g(I) = 1$. If lattice elements are equiprobable, then $\mathrm{Cov}(f,g) = -{}^1/_5$! A similar example can be given for the second structure. This shows that, for nondistributive lattices, increasing functions are not always positively correlated—at least for equiprobable lattice elements. It would be interesting to see if any general result is possible for the nondistributive case. Subsequent to setting down these remarks, Kemperman's paper (1977, Theorem 7) wich treats the necessity of distributivity was noticed.

The FKG inequality is a powerful device, but it may not be straightforward to use. In fact, as we said, Shepp (1982) treats a problem in which a "natural" ordering fails to satisfy the requirements but another ordering does work. This raises the question of whether there is a systematic way to use the FKG inequality.

One way of formalizing this is as follows:

(Existence) Let a finite set Ω, f, g: $\Omega \to R$, and a probability measure μ on Ω be given. It is desired to find a distributive lattice structure on Ω such that the FKG hypotheses ((a)–(c), (1.1)) hold. When can this be done? Is $\mathrm{Cov}(f,g) \geqslant 0$ close to a condition?

A second question is how long would it take to find a compatible lattice structure.

(Multiplicity) Given the previous set-up (Ω, f, g, μ): what fraction of distributive lattice structures on Ω satisfy the FKG requirements ((a)–(c), (1.1))?

Of course in both of these questions, compatibility of the lattice structure with μ and with the pair (f,g) can be considered separately.

We next discuss two problems to which the FKG inequality seems likely to be able to contribute some insight.

2. A Possible Application to a Partition Problem.

Problem 1. Let a_i, $i = 1, \ldots, m$ be a random sample drawn (without replacement) from $1, 2, \ldots, m+n$ and suppose the remaining numbers are denoted by b_j, $j = 1, \ldots, n$. The sum $S = \Sigma a_i$ has two interpretations.

(i) Let $X_1, \ldots, X_m$; $Y_1, \ldots, Y_n$ be two independent random samples drawn from a single population with a continuous distribution function so that the probability of one or more ties among the observations is zero. Let $R_1, \ldots, R_m$ be the ranks of the X observations among the $m+n$ observations. Then ΣR_i has the same distribution as S, in fact, R_i could be identified with a_i. In this case S is known as the Wilcoxon statistic (with the "null distribution").

(ii) Suppose a_i's are arranged so that $a_1 < a_2 < \ldots < a_m$. Then $a_i - i$ represents the number of b_j's smaller than a_i and it is easy to verify that

$$T = S - m(m+1)/2 = \Sigma_1^m (a_i - i)$$

has a distribution symmetric about $mn/2$ and assumes values, $0, 1, \ldots, mn$.

Further, for an integer k, $0 \leq k \leq mn$,

$$\binom{m+n}{m} P[T = K] = (k;m,n).$$

where $(k;m,n)$ is the number of partitions of k into m parts, with each part $\leqslant n$. In other words, $(k;m,n)$ represents the total number of distinct sequences of non-negative integers $\leqslant n$ and of length m such that each sequence is nondecreasing and the sum of the integers in the sequence is k.

There has been a wealth of literature on the theory of partitions. For the most recent source, see Andrews (1976, Section 3.2).

The problem we are concerned with here is the unimodality of the distribution of T, that is

$$(k;m,n) - (k-1;m,n) \geqslant 0, \quad 0 \leqslant k \leqslant mn/2,$$

which is directly connected with exceedances of a_i over b_j's. Dynkin (1950) proved the unimodality of T; however, the proof is based on the representation theory of Lie algebras. Recently, Hughes (1977) and Stanley (1980, 1981) discuss several problems regarding unimodality of sequences arising from Lie algebras. So far, no direct combinatorial or probabilistic proof is available. One possible approach is to compare the conditional probabilities of T, governed by the order structure of a_i and b_j's while one or more R_i are fixed. Such probability comparisons are similar to those appearing in Shepp (1982). The lattice structure and possible partial orderings are the same as in that paper. Could the FKG inequality again succeed to give a simple proof?

3. A Restriction to Linear Functions.

Problem 2. Let (X_1,X_2) be a pair of real random variables. The conclusion of the FKG inequality for the measure generated by (X_1,X_2) is that

$$\operatorname{cov}[f(X_1,X_2),g(X_1,X_2)] \geq 0, \tag{3.1}$$

for every pair of (co-ordinatewise) nondecreasing functions f,g. This property of positive dependence for (X_1,X_2) was termed as "association" by Esary, Proschan and Walkup (1967). A weaker notion of positive dependence called "positive quadrant dependence" (PQD) is defined by requiring

$$F_{X_1,X_2}(u,v) \geqslant F_{X_1}(u)F_{X_2}(v), \quad \text{for all } (u,v) \tag{3.2}$$

where F with the appropriate subscripts denotes the distribution function. Lehmann (1966) studied PQD and showed that (3.2) is equivalent to

$$\operatorname{cov}(h_1(X_1),h_2(X_2)) \geq 0, \tag{3.3}$$

for every pair of nondecreasing functions h_1,h_2. Although it is easy to show that (3.2) (or (3.3)) does not imply (3.1), if f,g are restricted to linear nondecreasing functions then the implication does hold. This was proved by Shaked (1982). We give a very simple proof. Note that, we want to prove the following:

$$P[X_1>x_1,X_2>x_2] \geqslant P[X_1>x_1]P[X_2>x_2] \quad \text{for all } (x_1,x_2)$$

implies

$$P[\Sigma_1^2 a_iX_i > c,\ \Sigma_1^2 b_iX_i > d] \geq P[\Sigma_1^2 a_iX_i > c]P[\Sigma_1^2 b_iX_i > d].$$

for a_i, b_i nonnegative and arbitrary constants c, d.

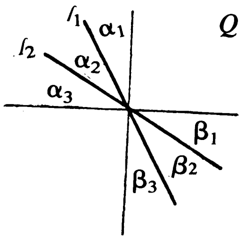


FIGURE 3. Quadrants.

Proof. Let Q, α_i, β_i be the probabilities of the regions as shown, created by interesecting lines l_1, l_2 representing $\Sigma_1^2 a_ix_i = c$ and $\Sigma_1^2 b_ix_i = d$ respectively. We have to show that

$$Q+\alpha_1+\beta_1 \geqslant (Q+\alpha_1+\beta_1+\beta_2)(Q+\alpha_1+\alpha_1+\alpha_2+\beta_1). \tag{3.4}$$

given that

$$Q \geqslant (Q+\Sigma_1^3\alpha_i)(Q+\Sigma_1^3\beta_i). \tag{3.5}$$

However, from (3.5) it is easy to check that

$$Q+\alpha_1+\beta_1 \geqslant (Q+\alpha_1+\Sigma_1^3\beta_i)(Q+\beta_1+\Sigma_1^3\alpha_i)$$

which implies (3.4).

Remark. Notice that in the above proof the quadrant could very easily be replaced by a region defined by an intersection of half planes other than $X_1 > x_1$ and $X_2 > x_2$, the only requirement being that it is *contained* in the intersection of $\Sigma a_i X_i > c$ and $\Sigma b_i X_i > d$.

We will set up an analogy for the measures on lattices. Let Γ be a lattice. Suppose T_1, T_2 are total ordering relations. Consider a partial ordering P induced by T_1, T_2 as follows:

Definition. $x \geq_P y$ *if* $x \geq_{T_1} y$ *and* $x \geq_{T_2} y$.

Given a measure μ on Γ one may define "marginal distribution functions" F_1, F_2 by $F_i(x) = \mu\{y; y \leq_{T_i} x\}$ and the PQD analog would be: for every $x \in \Gamma$,

$$\mu\{y{:}y \leq_P x\} \geq \mu\{y{:}y \leq_{T_1} x\} \cdot \mu\{y{:}y \leq_{T_2} x\}. \tag{3.6}$$

In view of the remark above one may ask the following: Suppose (T_1^*, T_2^*) is another pair of linear ordering on Γ such that the induced partial order P^* is weaker than P above, that is

$$y \leq_P x \Rightarrow y \leq_{P^*} x.$$

Under what conditions would (3.6) be sufficient for the validity of an analogous inequality involving P^* and (T_1^*, T_2^*)?

A related multivariate question is the following. Supose $X_1, \ldots, X_k$ are such that for arbitrary nonnegative constants a_i, b_i and an arbitrary proper subset A of $\{1,2, \ldots k\}$, $\Sigma_{i \in A} a_i X_i$ and $\Sigma_{i \in \bar{A}} b_i X_i$ are PQD, where $\bar{A}$ is the complement of A. This property may be called "disjoint positive linear dependence" (DPLD).

QUESTION. Does DPLD $\Rightarrow$ PLD? Here PLD means $\Sigma_1^k a_i X_i$, $\Sigma_1^k b_i X_i$ are PQD. (This problem is related to some concepts discussed in Joag-Dev (1983)).

It is interesting to see that $k = 3$ is the most crucial while $k = 2$ has already been proved. To see that the case $k = 3$ yields the general result, consider

$$Y_1 = X_1, Y_2 = \Sigma_{i \in A} a_i X_i, Y_3 = \Sigma_{i \in B} b_i X_i,$$

where A, B are *disjoint* and do not containt 1. The triplet Y_1, Y_2, Y_3 is DPLD. If PLD, it will show that the linear combinations containing *one* common variable would be PQD. Using the same technique successively, the cardinality of $A \cap B$ can be increased to k.

To see the relation between these covariance inequalities and FKG, suppose that L is a product lattice of two components L_1 and L_2 with partial ordering P_1 and P_2 respectively. Suppose μ defined on L satisfies FKG condition with respect to the partial ordering induced by (P_1, P_2). Then it follows that for every pair of nondecreasing functions (f,g) defined on L_1 and L_2 respectively,

$$\mathrm{Cov}[f,g] \geq 0. \tag{3.7}$$

However, it is well known that the validity of (3.7) for every pair of nondecreasing functions does not imply FKG inequality. Suppose now we restrict the nondecreasing functions to those which are *linear*, then the above converse seems to be plausible. In fact, it reduces to having DPLD and PLD conditions equivalent.

REFERENCES

AHLSWEDE, R. and DAYKIN, D. E. (1978). An inequality for the weights of two families of sets, their unions and intersections. *Z. Wahrscheinlichkeitstheorie und Verw. Gebiete 43* 183–185.

ANDREWS, G. (1976). *Theory of Partitions. Encyclopedia of Mathematics and Its Applications.* Addison-Wesley.

DYNKIN, E. B. (1950). Some systems of weights of linear representations of semisimple lie groups. *Dokl. Akad. Nauk. SSSR (N.S.) 71* 221–224.

ESARY, J., PROSCHAN, F. and WALKUP, W. (1967). Associated random variables with applications. *Ann. Math. Statist. 38* 1466–1474.

FORTUIN, C. M., GINIBRE, J, and KASTELEYN, P. N. (1971). Correlation inequalities on some partially ordered sets. *Comm. Math Phys. 22* 89–103.

GRAHAM, R. L. (1983). Applications of the FKG inequality and its relatives. To appear in *Mathematical Programming Bonn 1982 — The State of the Art,* ed. by Bachem, Grötschel, and Korte. Springer-Verlag.

HUGHES, J. W. B. (1977). Lie algebraic proofs of some theorems on partitions. *Number Theory and Algebra.* H. Zassenhaus, ed., Academic Press, New York.

JOAG-DEV, K. (1983) Characterization of independence via uncorrelatedness. *Ann. Probability 11* 1037–1041.

KEMPERMAN, J. H. B. (1977). On the FKG inequality for measures on a partially ordered space. *Indag. Math. 39* 313–331.

SHAKED, M. (1982). A general theory of some positive dependence notions. *J. Multivariate Anal. 12* 199–218.

SHEPP, L. A. (1982). The XYZ conjecture and the FKG inequality. *Ann Probability 10* 824–827.

STANLEY, R. (1980). Unimodal sequences arising from Lie algebras. *Young Day Proceedings,* 127–136. Marcel Dekkar, Inc., N.Y.

STANLEY, R. (1981). Two combinatorial applications of the Aleksandrov-Fenchel inequalities. *J. Comb. Theory 31* 56–65.

Inequalities in Statistics and Probability
IMS Lecture Notes–Monograph Series Vol. 5 (1984), 127–140

ASYMPTOTIC INDEPENDENCE AND LIMIT THEOREMS FOR POSITIVELY AND NEGATIVELY DEPENDENT RANDOM VARIABLES[1]

BY CHARLES M. NEWMAN
University of Arizona

For random variables which are associated or which exhibit certain related types of positive and negative dependence, the independence structure is largely determined by the covariance structure. We survey results of this sort with particular emphasis on limit theorems for partial sums of stationary sequences.

1. Introduction. The purpose of this paper is to survey a number of results concerning the degree to which the independence structure is determined by the covariance structure for families of random variables which exhibit certain types of positive or negative dependence. The original such result is due to Lehmann (1966). We first recall Lehmann's definition of positive and negative quadrant dependent (PQD and NQD) random variables. X_1 and X_2 are said to be PQD if

(1.1) $\quad H_{1,2}(x_1,x_2) \equiv P[X_1>x_1,X_2>x_2] - P[X_1>x_1]\,P[X_2>x_2] \geqslant 0$ for all $x_1,x_2 \in \mathcal{R}$;

They are said to be NQD if X_1 and $(-X_2)$ are PQD.

Note that an equivalent condition to (1.1) is that $\mathrm{Cov}(f(X_1), g(X_2)) \geqslant 0$ for all real increasing (i.e. nondecreasing) f and g (such that $f(X_1)$ and $g(X_2)$ have finite variance). *In the following statement of Lehmann's result and throughout the rest of the paper we will assume, unless otherwise mentioned, that all random variables have finite variance.*

THEOREM 1 *(Lehmann (1966)). If X_1 and X_2 are PQD or NQD, then they are independent if and only if* $\mathrm{Cov}(X_1,X_2) = 0$.

Proof. This theorem is an immediate consequence of the identity (obtained from integration by parts),

(1.2) $$\mathrm{Cov}(X_1,X_2) = \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} H_{1,2}(x_1,x_2)\,dx_1\,dx_2$$

and the pointwise positivity (resp. negativity) of $H_{1,2}$ for PQD (resp. NQD) variables. □

The results which we discuss in this paper concern multivariate generalizations of Theorem 1 of two types. The first type is a direct generalization in which joint uncorrelatedness implies joint independence. The second type is an indirect generalization in which approximate uncorrelatedness implies approximate independence in a sufficiently quantitative sense to lead to useful limit theorems for sums of dependent variables. In Section 2, we review all the results of the first type along with an ergodicity result of the second type; with one exception, these are based on inequalities for *distribution* functions. In Section 3, we review a number of results of the second type, including a triangular array limit theorem and a central limit theorem; these are based on inequalities for *characteristic* functions. In Section 4, we present some recent results which extend the inequalities and limit

---

[1] Research supported in part by the National Science Foundation under grant No. MCS 80–19384.
AMS 1980 subject classifications. Primary, 60F05, 62H05; secondary, 60F15, 62H80.
Key words and phrases: Association, Asymptotic Independence, Central Limit Theorem, Positive Dependence, Negative Dependence.

theorems of Section 3 to, e.g., nonmonotonic functions of associated variables; some of the results of Section 4 are new. Finally, in Section 5, we review briefly some related results and open problems; these concern nonstationary sequences, Berry-Esseen asymptotics, invariance principles, and demimartingales.

We conclude this section by noting that many of the results given below, which are stated for both positively and negatively dependent variables, were originally derived, in the referenced papers, only for the positively dependent case; the derivation for the negative case is usually essentially unchanged. There is however a simple, but striking, distinction between the two cases, which can be seen in Theorems 7, 12 and 17. Namely, as a consequence of the elementary Lemma 8, it follows that for a stationary sequence $Y_1, Y_2, \dots$, the decrease in $|\mathrm{Cov}(Y_1, Y_j)|$ as $j \to \infty$ which must be specifically assumed in the positive case in order to have ergodicity or a central limit theorem is automatically valid in the negative case: *stationary negatively dependent sequences are automatically asymptotically independent.*

2. Distribution Function Inequalities and Applications. A finite family $\{X_1, \dots, X_n\}$ of random variables is said to be *associated* if $\mathrm{Cov}(f(X_1, \dots, X_n), g(X_1, \dots, X_n)) \geq 0$ for any real (coordinatewise) increasing functions f and g on $\mathcal{R}^n$; it is said to be *negatively associated* if for any disjoint $A, B \subset \{1, \dots, n\}$ and any increasing functions f on $\mathcal{R}^A$ and g on $\mathcal{R}^B$, $\mathrm{Cov}(f(X_k; k \in A), g(X_l; l \in B)) \leq 0$. The first definition is due to Esary, Proschan, and Walkup (1967) and the second to Joag-Dev and Proschan (1983). Infinite families are associated (resp. negatively associated) if every finite subfamily is associated (resp. negatively associated). These definitions are two of the many possible multivariate generalizations of Lehmann's PQD and NQD; for further discussion of these and related concepts, see Karlin and Rinott (1980a; 1980b), Shaked (1982a), Block, Savits and Shaked (1982), and the references therein. All of the results discussed in this paper apply to associated and most apply to negatively associated families; many results apply under weaker hypotheses as will be discussed below.

There are two almost independent bodies of literature on the subject of associated random variables. One developed from the work of Esary, Proschan and Walkup (1967) and Sarkar (1969) and is oriented towards reliability theory and statistics; the other developed from the work of Harris (1960) and of Fortuin, Kastelyn and Ginibre (1971) and is oriented towards percolation theory and statistical mechanics. It should be noted that in the latter literature, the term "associated" is usually not used but rather variables are said to satisfy the FKG inequalities. Some people use the term FKG inequalities only when the joint distribution satisfies some version of the lattice-theoretic sufficient condition for being associated which was analyzed by Fortuin, Kastelyn and Ginibre (1971). When the joint distribution has a smooth density $p(x_1, \dots, x_n)$, which is strictly positive on all of $\mathcal{R}^n$, this condition is equivalent to

$$(2.1) \qquad (\partial^2/\partial x_i \partial x_j) \ln p \geq 0 \quad \text{for all } i \neq j \text{ and all } x_1, \dots, x_n,$$

which is further equivalent to the "TP_2 in pairs" condition obtained independently (and previously) in Sarkar (1969).

That condition (2.1) is not necessary for association can be seen by considering a trivariate normal vector whose covariance matrix, although (entrywise) positive, is not the inverse of a matrix with nonpositive off-diagonal entries. Such normal variables can easily be constructed; they are associated by the results of Pitt (1982), but do not satisfy (2.1).

A paper which has a nice proof of the sufficiency of the FKG condition along with references to many of the papers in both bodies of literature is Karlin and Rinott (1980a).

The recognition that association is useful in the study of approximate independence seems to have first occurred in Lebowitz (1972). The central limit theorem reviewed in Section 3 below was largely motivated by Lebowitz' results. Also in Lebowitz' paper is an inequality on distribution functions, which we state as Theorem 2, which gives a very simple proof that uncorrelated implies independent for associated variables. This latter result was apparently first stated as a theorem in Wells (1977), but the proof given there was a complicated one based on generalizations of other theorems of Lebowitz (1972) (see Simon (1973)); it was not noticed until recently that this result is an immediate consequence of Lebowitz' basic distribution function inequality. We define for A and B subsets of $\{1, \ldots, n\}$, and real x_j's,

$$H_{A,B} = P(X_j > x_j; j \in A \cup B) - P(X_k > x_k; k \in A)\, P(X_l > x_l; l \in B); \tag{2.2}$$

note that according to (1.1) $H_{i,j} = H_{\{i\},\{j\}}$.

THEOREM 2 *(Lebowitz (1972)). If the X_j's are associated, then*

$$0 \leqslant H_{A,B} \leqslant \Sigma_{k \in A} \Sigma_{l \in B} H_{k,l}; \tag{2.3a}$$

if the X_j's are negatively associated, then for disjoint A,B

$$0 \geqslant H_{A,B} \geqslant \Sigma_{k \in A} \Sigma_{l \in B} H_{k,l}. \tag{2.3b}$$

Proof. Let ρ_j denote the indicator function of the event $\{X_j > x_j\}$ and define

$$\rho_A = \Pi_{j \in A} \rho_j \quad , \quad S_A = \Sigma_{j \in A} \rho_j. \tag{2.4}$$

It is easy to see that ρ_A, ρ_B, $S_A - \rho_A$, S_A, and $S_B - \rho_B$ are all increasing functions of the X_j's; it follows that for associated X_j's

$$0 \leqslant \mathrm{Cov}(\rho_A, \rho_B) \leqslant \mathrm{Cov}(S_A, \rho_B) \leqslant \mathrm{Cov}(S_A, S_B).$$

This yields (2.3a) since $H_{A,B} = \mathrm{Cov}(\rho_A, \rho_B)$ while the right hand side of (2.3a) equals $\mathrm{Cov}(S_A, S_B)$; the case of negative association is similar. □

COROLLARY 3. *Suppose the X_j's are either associated or negatively associated. It follows that $\{X_k; k \in A\}$ is independent of $\{X_l; l \in B\}$ if and only if* $\mathrm{Cov}(X_k, X_l) = 0$, *for all $k \in A$ and $l \in B$; similarly the X_j's are jointly independent if and only if* $\mathrm{Cov}(X_k, X_l) = 0$ *for all $k \neq l$.*

Proof. This is an immediate consequence of Theorems 1 and 2. □

It is clear that (2.3a) remains valid for disjoint A,B, if the hypothesis of association is weakened to make it analogous to a positive version of negative association; there seems to be no standard term for this weakened version of association. Both parts of Corollary 3 are valid for this weakened association (as well as for negative association); the second part of Corollary 3 can also be shown to be valid under even weaker hypotheses on the dependence of X_j's as we now discuss.

We let $\bar{\rho}_j = 1 - \rho_j =$ the indicator function of the event $\{X_j \leqslant x_j\}$, $\bar{\rho}_A = \Pi_{j \in A} \bar{\rho}_j$, and then following Joag-Dev (1983) we define $\{X_1, \ldots, X_n\}$ to be *strongly positive orthant dependent* (SPOD) if for any disjoint $A,B \subset \{1, \ldots, n\}$ and any real x_j's,

$$\mathrm{Cov}(\rho_A, \rho_B) \geq 0,\ \mathrm{Cov}(\bar{\rho}_A, \bar{\rho}_B) \geq 0,\ \mathrm{Cov}(\rho_A, \bar{\rho}_B) \leq 0, \tag{2.5a}$$

and *strongly negative orthant dependent* (SNOD) if analogously

(2.5b) $$\mathrm{Cov}(\rho_A,\rho_B)\leq 0,\ \mathrm{Cov}(\bar\rho_A,\bar\rho_B)\leq 0,\ \mathrm{Cov}(\rho_A,\bar\rho_B)\geq 0.$$

It is immediate that association (resp. negative association) implies SPOD (resp. SNOD) which in turn implies pairwise PQD (resp. NQD).

The following theorem is due to Joag-Dev, but the proof given here is somewhat different than the original one.

THEOREM 4. *[Joag-Dev (1983)]. Suppose the X_j's are either SPOD or SNOD. It follows that they are jointly independent if and only if* $\mathrm{Cov}(X_k,X_l)=0$ *for all* $k\neq l$.

Proof. The theorem is an immediate consequence of Theorem 1 together with the following distribution function inequality. □

LEMMA 5. *Suppose $X_1,\ldots,X_m$ are SPOD; then*

(2.6a) $$0\leq P[X_j>x_j, j=1,\ldots,m]-\Pi_{j=1}^m P(X_j>x_j)\leq K_m \sum_{\substack{k,l=1\\k<l}}^m H_{k,l}$$

where K_m is a constant depending only on m. If the X_j's are SNOD, then

(2.6b) $$0\geq P[X_j>x_j, j=1,\ldots,m]-\Pi_{j=1}^m P(X_j>x_j)\geq K_m \sum_{\substack{k,l=1\\k<l}}^m H_{k,l}.$$

Proof. We consider the SPOD case; the SNOD case is treated similarly. The quantity of interest in the center of (2.6a) may be rewritten as

$$G_m\equiv \mathrm{E}(\Pi_{j=1}^m\rho_j)-\Pi_{j=1}^m E(\rho_j).$$

Its positivity follows easily from repeated application of (2.5a); we wish to obtain the upper bound of (2.6a). Denoting $\Pi_{j=1}^m\rho_j$ by ρ^m, we have

(2.7) $$G_{m+1}=E(\rho_{m+1})G_m+\mathrm{Cov}(\rho^m,\rho_{m+1})\leq G_m+\mathrm{Cov}(\rho^m,\rho_{m+1}),$$

while for $j>n+1$,

$$\mathrm{Cov}(\rho^{n+1},\rho_j)=\mathrm{Cov}(\rho^n,\rho_{n+1}\rho_j)-E(\rho_j)E(\rho^n\rho_{n+1})+E(\rho_{n+1}\rho_j)E(\rho^n)$$

(2.8) $$=\mathrm{Cov}(\rho^n,\rho_{n+1})+\mathrm{Cov}(\rho^n,\rho_j)+\mathrm{Cov}(\rho^n,\bar\rho_{n+1}\bar\rho_j)+\mathrm{Cov}(\rho_{n+1}\rho_j)\cdot E(\rho^n)$$
$$-E(\rho_j)\mathrm{Cov}(\rho^n,\rho_{n+1})\leq \mathrm{Cov}(\rho^n,\rho_{n+1})+\mathrm{Cov}(\rho^n,\rho_j)+\mathrm{Cov}(\rho_{n+1},\rho_j).$$

The last inequality follows from the fact that SPOD implies

$$\mathrm{Cov}(\rho^n,\bar\rho_{n+1}\bar\rho_j)\leq 0,\ \mathrm{Cov}(\rho_{n+1},\rho_j)\geq 0,\ \mathrm{Cov}(\rho^n,\rho_{n+1})\geq 0,$$

while $E(\rho^n)\leq 1$ and $E(\rho_j)\geq 0$. The right hand inequality of (2.6a) follows from (2.7) and (2.8) by induction. □

Remark. If the X_j's are associated or negatively associated, then it is easily seen that (2.3) implies (2.6) with $K_m=1$. It is not known to the author whether this value of K_m is valid under the weaker hypothesis of SPOD or SNOD. It is also not known to the author whether inequality (2.3) (possibly modified by a factor analogous to K_m) and the *first* result of Corollary 3 are valid when only assuming SPOD or SNOD.

An alternative improvement to the second result of Corollary 3 can be obtained from the characteristic function inequalities discussed in Section 3 below. We define X_j's to be *linearly positive quadrant dependent* (LPQD) if for any disjoint A,B and positive λ_j's, $\Sigma_{k\in A}\lambda_k X_k$ and $\Sigma_{l\in B}\lambda_l X_l$ are PQD; *linearly negative quadrant dependent* (LNQD) is defined in the obvious analogous manner. The next theorem is an immediate corollary of Theorem 10 of the next section. We include it here for comparison with Theorem 4.

THEOREM 6. *Suppose the X_j's are either LPQD or LNQD. It follows that they are jointly independent if and only if* $\mathrm{Cov}(X_k,X_l)=0$ *for all* $k\neq l$.

Remark. As in the previous remark, it is not known to the author whether the first result of Corollary 3 is valid under the weaker hypothesis of LPQD or LNQD.

In order to compare Theorems 4 and 6, we present two examples of M. Shaked (1982b) which show that neither SPOD nor LPQD implies the other. Consider three discrete random variables with joint density $p(x_1,x_2,x_3) \equiv P[X_1 = x_1, X_2 = x_2, X_3 = x_3]$. In the first example, $p(0,1,0) = p(0,2,0) = p(1,0,1) = p(1,1,0) = 1/14$, $p(0,2,1) = p(1,0,0) = 2/14$ and $p(0,0,0) = p(1,2,1) = 3/14$; here $\{X_1,X_2,X_3\}$ is not LPQD since $3/14 = P[X_1>0, X_2+X_3 > 1] < P[X_1>0] \cdot P[X_2 + X_3>1] = (7/14)^2$ while a lengthy verification shows that it is in fact SPOD. In the second example, $p(2,2,1) = p(3,2,1) = p(2,3,1) = p(3,3,1) = p(1,1,2) = p(2,1,2) = p(3,1,2) = p(1,2,2) = p(1,3,2) = 1/17$ and $p(1,1,1) = p(3,3,2) = 4/17$; here $\{X_1,X_2,X_3\}$ is not SPOD since $P[X_1>1,X_2>1,X_3>1] = 4/17 < P[X_1>1,X_2>1] \cdot P[X_3>1] = (8/17) \cdot (9/17)$ while a lengthy verification shows that it is in fact LPQD. For another example showing that LPQD does not imply SPOD with more details, see Joag-Dev (1983).

The next theorem on ergodicity is a consequence of Theorem 2. It is implicitly contained in the work of Lebowitz (1972) and is explicitly mentioned in a remark of Newman (1980) in the more general context of sequences indexed by Z^d. A somewhat simpler proof than the following one can be based on Theorem 16 below.

THEOREM 7 *[Lebowitz (1972)]. Let $X_1,X_2, \dots$ be a strictly stationary sequence which is either associated or negatively associated and let T denote the usual shift transformation, defined so that $T(f(X_{j_1}, \dots, X_{j_m})) = f(X_{j_1+1}, \dots, X_{j_m+1})$. Then T is ergodic (i.e., every T-invariant event in the σ-field generated by the X_j's has probability 0 or 1) if and only if*

$$\lim_{n\to\infty} n^{-1}\Sigma_{j=1}^{n}\mathrm{Cov}(X_1,X_j) = 0. \tag{2.9}$$

In particular, if (2.9) is valid, then for any f such that $f(X_1)$ is L_1,

$$\lim_{n\to\infty} n^{-1}\Sigma_{j=1}^{n} f(X_j) = E(f(X_1)) \quad \text{a.s.}$$

In the negatively associated case, (2.9) is automatically valid.

Proof. The necessity of (2.9) follows from the L_2 ergodic theorem which implies $n^{-1}\Sigma_{j=1}^{n}X_j \to E(X_1)$ in L_2. To prove the sufficiency of (2.9) we note that by standard ergodic theory/Hilbert space arguments, it suffices to find two sets S_1 and S_2 of random variables (measurable with respect to the X_j's) each of whose linear combinations are dense in L_2 and a subsequence n_i, such that for any $W_1 \epsilon S_1$, $W_2 \epsilon S_2$,

$$\lim_{i\to\infty} \mathrm{Cov}(W_1, n^{-1}\Sigma_{j=1}^{n_i} T^j W_2) = 0, \tag{2.10}$$

since that would imply that the eigenvalue 1 of T is simple. For $l = 1,2$, we take $S_l = \{\Pi_{j=1}^{m}\rho_j(x_j) : m = 1,2, \dots ; \text{each } x_j \epsilon D_l\}$ where $\rho_j(x_j)$ is the indicator function of $\{X_j > x_j\}$ and D_l is a dense subset of $\mathcal{R}$ to be chosen. To see that linear combinations of S_l are dense in L_2, note that for $x_j \le x_j'$, $\Pi_{j=1}^{m}[\rho_j(x_j) - \rho_j(x_j')]$ is the indicator function of the rectangle, $\{x_j < X_j \le x_j', \text{ for all } j\}$. Defining $H_j(x_1,x_2) = \mathrm{Cov}(\rho_1(x_1),\rho_j(x_2))$ and $\tilde{H}_n = n^{-1}\,\Sigma_{j=1}^{n}H_j$. we see from Theorem 2 that to obtain (2.10) it suffices to show that $\tilde{H}_{n_i}(x_1,x_2) \to 0$ for $x_1 \epsilon D_1, x_2 \epsilon D_2$. But by (2.9) and identity (1.2) we know that $\int_{-\infty}^{\infty}\int_{-\infty}^{\infty}\tilde{H}_n(x_1,x_2)dx_1dx_2 \to 0$ as $n \to \infty$ and moreover $|\tilde{H}_n| \le 1$; it follows there is a subsequence so that $\tilde{H}_{n_i} \to 0$ pointwise (except on a set of zero Lebesgue measure) in $\mathcal{R}^2$ and thus that D_1 and D_2 exist. The final statement of the theorem is a consequence of the following lemma. □

LEMMA 8. *If $X_1,X_2, \dots$ is a (wide sense) stationary sequence with $\mathrm{Cov}(X_i,X_j) \le 0$ for $i \ne j$, then $\Sigma_{j=2}^{\infty}\mathrm{Cov}(X_1,X_j)$ is absolutely convergent and*

(2.11) $$\sigma^2 \equiv \operatorname{var}(X_1)+2\Sigma_{j=2}^{\infty}\operatorname{Cov}(X_1,X_j)\in[0,\operatorname{Var}(X_1)].$$

Proof. This is a consequence of the negativity of $\operatorname{Cov}(X_1,X_j)$ for $j\geq 2$, which implies that $\sigma_n^2 \equiv \operatorname{Var}(n^{-1/2}\Sigma_{j=1}^{\infty}X_j)$ satisfies

(2.12) $$0 \leq \lim_{n\to\infty}(\operatorname{Var}(X_1)-\sigma_n^2) = -2\,\Sigma_{j=2}^{\infty}\operatorname{Cov}(X_1,X_j)$$

Since $\sigma_n^2 \geq 0$, we must have the right hand side of (2.12) finite and bounded by $\operatorname{Var}(X_1)$ which completes the proof. □

3. Characteristic Function Inequalities and Applications. We begin with a simple extension of Theorem 1 which gives a quantitative estimate of the approximate independence between a pair of variables in terms of the covariance.

PROPOSITION 9 *(Newman (1980)). If X and Y are PQD or NQD, then*

(3.1) $$|E(e^{irX+isY})-E(e^{irX})E(e^{isY})|\leq|rs\operatorname{Cov}(X,Y)|, \textit{for all real } r,s.$$

Proof. Integration by parts yields, analogously to (1.2), the identity,

(3.2) $$\operatorname{Cov}(e^{irX}, e^{isY}) = \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} ire^{irx}ise^{isy}H(x,y)dxdy.$$

where H is defined as in (1.1). The triangle inequality, the pointwise positivity (resp. negativity) of H for PQD (resp. NQD) variables, and equation (1.2) then yield (3.1). □

The next theorem is the main ingredient used to obtain the limit theorems of this section.

THEOREM 10 *(Newman (1980)). Suppose $X_1, \ldots, X_m$ are LPQD or NPQD; then*

(3.3) $$|\phi(r_1,\ldots,r_m) - \Pi_{j=1}^{m}\phi_j(r_j)| \leq \Sigma_{\substack{k,l=1\\k<l}}^{m}|r_kr_l\operatorname{Cov}(X_k,X_l)|$$

where ϕ and ϕ_j are given by

$$\phi = E(\exp[i\Sigma_{j=1}^{m}r_jX_j]),\ \phi_j = E(\exp[i\,r_jX_j]).$$

Proof. (3.3) follows from (3.1) by induction on m. The first step of the induction argument is to choose a nontrivial subset A of $\{1,\ldots,m\}$ so that the r_j's have a common sign in A and a common sign in $\bar{A}$, the complement of A. Defining $\phi_B = E(\exp[i\ \Sigma_{j\in B}r_jX_j])$, we then have the left hand side of (3.3) bounded by

(3.4) $$|\phi-\phi_A\phi_{\bar{A}}| + |\phi_A|\,|\phi_{\bar{A}}-\Pi_{l\in\bar{A}}\phi_l| + |\Pi_{\ell\in\bar{A}}\phi_\ell||\phi_A-\Pi_{k\in A}\phi_k|.$$

The first term of (3.4) is bounded by (3.1) while the other two terms are bounded by the induction hypothesis (and the fact that $|\phi_A|,|\phi_l|\leq 1$) to yield the right hand side of (3.3). □

The next theorem is an immediate corollary of Theorem 10. I appears in Newman, Rinot and Tversky (1982) and independently in Wood (1982). It was used in the latter reference for a general analysis of limit theorems for sums of associated variables and in the former reference for a specific application to a model arising in mathematical psychology. In that model there is a collection of "distances," $\{D_{ij};\ 0\leq i<j\leq n\}$, between objects i and j, which are exchangeable random variables and one is interested in the asymptotic behavior of S_n= number of objects in $\{1,\ldots,n\}$ which have object 0 as their nearest neighbor. S_n can be represented as $\Sigma_{j=1}^{n}Y_{n,j}$ where $Y_{n,j}$ is the indicator function of the event that j has 0 as its nearest neighbor. Each $Y_{n,j}$ is Bernoulli ($p=1/n$) and although they are not independent (for fixed n) they can be shown to be associated. The following theorem can then be used to give a particularly simple proof that S_n converges in distribution to Poisson ($\lambda=1$).

THEOREM 11 *(Newman, Rinott, Tversky (1982); Wood (1982)). Suppose $Y_{n,j}$ and $W_{n,j}$*

($n=1,2, \ldots ; j=1,2, \ldots , M_n$) are triangular arrays such that for each n and j, $Y_{n,j}$ is equidistributed with $W_{n,j}$ and such that for each n, the $Y_{n,j}$'s are LPQD or NPQD while the $W_{n,j}$'s are independent. If in addition,

$$\lim_{n\to\infty} \sum_{\substack{k,l=1\\ k<l}}^{M_n} \mathrm{Cov}(Y_{n,k}, Y_{n,l}) = 0, \tag{3.5}$$

then $\sum_{j=1}^{M_n} Y_{n,j}$ converges in distribution to (some) X if and only if $\sum_{j=1}^{M_n} W_{n,j}$ converges in distribution to (the same) X.

Proof. This is an immediate consequence of Theorem 10 and standard arguments. □

The next theorem was the original application of (and motivation for) the characteristic function inequality (3.3). It was first given in Newman (1980) in the more general context of sequences indexed by Z^d. In this paper we sketch a proof based on Theorem 11; for a more detailed proof (using Theorem 10 directly rather than Theorem 11) see Newman (1980) or Newman and Wright (1981). The theorem itself (or more accurately its Z^d indexed generalization) was applied to Ising model magnetization fluctuations (or the equivalent lattice gas model density fluctuations) in Newman (1980) and to the density fluctuations of infinite clusters in percolation models in Newman and Schulman (1981). In the statement of the theorem, note that for LPQD (resp. NPQD) Y_j's

$$\mathrm{Cov}(Y_1,Y_j) \geqslant 0 \quad (\text{resp. } \mathrm{Cov}(Y_1,Y_j) \leqslant 0) \quad \text{for all } j \geqslant 2.$$

THEOREM 12 *(Newman (1980)). Let $Y_1, Y_2, \ldots$ be a strictly stationary sequence which is LPQD or LNQD. Then*

$$\sigma^2 \equiv \mathrm{Var}(Y_1) + 2\sum_{j=2}^{\infty} \mathrm{Cov}(Y_1,Y_j)$$

always exists and $\sigma^2 \epsilon [\mathrm{Var}(Y_1), \infty]$ (resp. $\sigma^2 \epsilon [0, \mathrm{Var}(Y_1)]$) in the LPQD (resp. LNQD) case. If $\sigma^2 \neq \infty$; i.e. if in the LPQD case we additionally assume that

$$\sum_{j=2}^{\infty} \mathrm{Cov}(Y_1,Y_j) < \infty, \tag{3.6}$$

then

$$\lim_{n\to\infty} n^{-1/2} \sum_{j=1}^{n} (Y_j - EY_j) = \sigma Z \tag{3.7}$$

where Z is standard normal and (3.7) refers to convergence in distribution.

Sketch of Proof. The first part of the theorem follows from the positivity or negativity of (Y_1,Y_j) for $j \geq 2$ and Lemma 8 above. In particular, this, together with the non-negativity of the variance of $\sum_{j=1}^{n} Y_j$, *yields in the LNQD case the bound,* $2\sum_{j=2}^{n} \mathrm{Cov}(Y_1,Y_j) \geq -\mathrm{Var}(Y_1)$. For the rest of the theorem, we define "block variables,"

$$Y_k^m = m^{-1/2} \sum_{j=(k-1)m+1}^{km} (Y_j - EY_j) \quad ; \quad m = 1,2, \ldots ; j = 1,2, \ldots \tag{3.8}$$

By straightforward variance estimates, it can be seen that it suffices to show $\lim_{\ell\to\infty} S_\ell^{m_\ell} = \sigma Z$, where

$$S_\ell^m = (m\ell)^{-1/2} \sum_{j=1}^{m\ell} (Y_j - EY_j) = \ell^{-1/2} \sum_{j=1}^{\ell} Y_j^m, \tag{3.9}$$

and m_ℓ is some nondecreasing sequence of positive integers such that $(m_{\ell+1} - m)/m_\ell \to 0$. If for each m, we define $W_j^m (j=1,2, \ldots)$ to be i.i.d. and equidistributed with Y_1^m, then defining

$$\tilde{S}_\ell^m = \ell^{-1/2} \sum_{j=1}^{\ell} W_j^m,$$

we have (by the standard central limit theorem) that for fixed m, $\lim_{\ell\to\infty} \tilde{S}_\ell^m = \sigma_m Z$ *where* $\sigma_m^2 = \mathrm{Var}(Y_1^m)$. More variance estimates show that $\sigma_m \to \sigma$ and thus that for any sequence m_ℓ growing to ∞ sufficiently slowly, $\lim_{\ell\to\infty} \tilde{S}^{m_\ell}_\ell = \sigma Z$. The desired result follows from Theorem 11 by taking $Y_{\ell,j} = \ell^{-1/2} Y_j^{m_\ell}$ and $W_{\ell,j} = \ell^{-1/2} W_j^m$, providing we show that (3.5) is valid.

But (3.5) is a simple consequence of σ_m converging to σ. □

Remark. Note that $\sigma^2>0$ in the LPQD case (except when the *Y_j's are constant) but* σ^2 can in fact vanish in the LNQD case. A trivial example of the latter phenomenon is obtained by taking $Y_j = Z_j - Z_{j-1}$ where $Z_0, Z_1, \ldots$ are i.i.d. standard normal; these Y_j's are not only LNQD but are negatively associated by a result of Joag-Dev and Proschan (1983).

We present the following theorem of Herrndorf (1983) without proof. It disproves a conjecture of Newman (1980) and Newman and Wright (1981) concerning the weakening of condition (3.6).

THEOREM 13 *(Herrndorf (1983)). There exists, $Y_1, Y_2, \ldots$, a strictly stationary nonconstant associated sequence with $K(R) \equiv \mathrm{Var}(Y_1) + 2\,\Sigma_{j\leqslant R}\mathrm{Cov}(Y_1, Y_j)$ slowly varying as $R\to\infty$ (i.e. $K(\lambda R)/K(R)\to 1$ as $R\to\infty$ for any $\lambda>0$) such that*

$$[nK(n)]^{-1/2}\Sigma_{j=1}^{n}(Y_j-EY_j)$$

does not converge in distribution to a standard normal Z.

4. More Characteristic Function Inequalities and Applications. In this section, we present a number of recent results, one of whose motivations is the desire to extend Theorem 12 to a central limit theorem for sums of $f(Y_j)$'s; some of the results are presented here for the first time. If the Y_j's are associated or negatively associated and f is either increasing or decreasing, then Theorem 12 can be directly applied to the $f(Y_j)$'s. We begin with a number of inequalities which are applicable to more general f's.

For f and f_1 complex functions on $\mathcal{R}^m$, we write $f << f_1$ if $f_1 - Re(e^{i\alpha}f)$ is (coordinatewise) nondecreasing for all real α. Note first that $f_1 = [(f_1 - Re(f)) + (f_1 - Re(-f))]/2$ and hence is automatically nondecreasing and second that $f << f_1$ for real f if and only if $f_1 + f$ and $f_1 - f$ are both nondecreasing. We write $f <<_A f_1$ if $f << f_1$ and both f_1 and f depend only on x_j's with $j\epsilon A$. The next two propositions will be used to obtain useful characteristic function inequalities.

PROPOSITION 14. *If h is real, $h << h_1$, and φ is a complex function on $\mathcal{R}$ such that $|\varphi(t)-\varphi(s)| \leqslant |t-s|$ for all t,s, then $\varphi(h) << h_1$. This applies in particular to $\varphi(h) = \exp(ih)$.*

Proof. We denote by Δg the increment in the function g when one or more of the x_j's is increased. We wish to show that for any real α, $\Delta[h_1 - Re(e^{i\alpha}\varphi(h))]\geqslant 0$. But $|\Delta Re(e^{i\alpha}\varphi(h))| \leqslant |\Delta(e^{i\alpha}\varphi(h))| = |\Delta\varphi(h)| \leqslant |\Delta h|$ because of the properties of φ, while $|\Delta h| \leqslant \Delta h_1$ because $h << h_1$. □

PROPOSITION 15 *(Newman (1983)). Suppose $f <<_A f_1$ and $g <<_B g_1$. Define $<f,g> = \mathrm{Cov}(f(X_1, X_2, \ldots), g(X_1, X_2, \ldots))$ where the X_j's are either associated or negatively associated. In the negatively associated case, assume in addition that A and B are disjoint. Then $|<f,g>| \leqslant |<f_1, g_1>|$ iff and/or g is real; otherwise $|<f,g>| \leqslant 2|<f_1, g_1>|$.*

Proof. First suppose f is real. Since $|<f,g>| = \sup(Re(e^{i\alpha}<f,g>):\alpha\epsilon\mathcal{R})$, it suffices to show that $Re(e^{i\alpha}<f,g>) \leqslant |<f_1, g_1>|$. This follows from the assumption that $h \equiv Re(e^{i\alpha}g) << g_1$ and $f << f_1$ and the identities,

$$|<f_1, g_1>| - <f,h> = <f_1, g_1> - <f,h> = \tfrac{1}{2}[<f_1+f, g_1-h> + <f_1-f, g_1+h>] \geqslant 0$$

for the associated case, and

$$|<f_1, g_1>| - <f,h> = -<f_1, g_1> - <f,h> = \tfrac{1}{2}[<f_1+f, g_1+h> + <f_1-f, g_1-h>] \geqslant 0$$

for the negatively associated case. If g is real the argument is the same and if neither are real, one has

$$|\langle f,g\rangle| = |\langle Ref,h\rangle + i\langle Imf,g\rangle| \le |\langle Ref,g\rangle| + |\langle Imf,g\rangle|$$

so that the desired inequality follows from the real f inequality. □

Remark. Proposition 15 is a generalization of Theorem 2. It is possible that the factor 2 appearing when both f and g are complex could be eliminated by a better proof; that would also eliminate the corresponding factors of 2 in the next theorem.

THEOREM 16. *(Newman (1983)). Suppose that for each j,* $X_j = f_j(Y_1,Y_2, \ldots)$, $\tilde{X}_j = \tilde{f}_j(Y_1,Y_2, \ldots)$ *where the* Y_i*'s are associated or negatively associated. Suppose further that* $f_j \ll_{A_j} \tilde{f}_j$ *for each j and, in the negatively associated case, additionally that the* A_j*'s are disjoint. Then the characteristic functions of the* X_j*'s,* ϕ, ϕ_j, ϕ_C*, defined as in Theorem 10 and its proof, satisfy (for disjoint A,B)*

$$|\phi_{A \cup B} - \phi_A\phi_B| \le 2\,\Sigma_{k \in A}\Sigma_{l \in B}|r_k r_l \mathrm{Cov}(\tilde{X}_k,\tilde{X}_l)| \tag{4.1}$$

and

$$|\phi - \Pi_{j=1}^m \phi_j| \le 2\,\Sigma_{\substack{k,l=1\\k<l}}^m |r_k r_l \mathrm{Cov}(\tilde{X}_k,\tilde{X}_l)|. \tag{4.2}$$

Proof. (4.1) follows from Propositions 14 and 15 since the left hand side of (4.1) is $|\langle f,g\rangle|$ with $f = \exp(i\,\Sigma_{j\in A} r_j X_j)$, $g = \exp(i\,\Sigma_{j\in B} r_j X_j)$ and since $\Sigma r_j f_j \ll \Sigma |r_j| \tilde{f}_j$. (4.2) follows from (4.1) essentially as in the proof of Theorem 10 from Proposition 9. □

There is a natural extension of Theorem 11 which follows from Theorem 16 in the same way as Theorem 11 follows from Theorem 10. To save space, we do not state that extension explicitly but rather go on to an extension of Theorem 12. This latter extension was applied in Newman (1983) to the fluctuations in Ising model energy densities and to the fluctuations of infinite cluster surfaces in percolation models.

THEOREM 17 *(Newman (1983)). Let* $Y_1,Y_2, \ldots$ *be a strictly stationary sequence which is associated (resp. negatively associated). Let* $X_j = f(Y_j,Y_{j+1}, \ldots)$ *and* $\tilde{X}_j = \tilde{f}(Y_j, Y_{j+1}, \ldots)$ *(resp.* $X_j = f(Y_j)$ *and* $\tilde{X}_j = \tilde{f}(Y_j)$*) with* $f \ll \tilde{f}$*; in the associated case, assume in addition that*

$$\Sigma_{j=2}^{\infty}\mathrm{Cov}(\tilde{X}_1,\tilde{X}_j) < \infty. \tag{4.3}$$

Then

$$\lim_{n\to\infty} n^{-1/2}\Sigma_{j=1}^{n}(X_j - EX_j) = \sigma Z, \tag{4.4}$$

where Z is standard normal and

$$\sigma^2 = \mathrm{Var}(X_1) + 2\,\Sigma_{j=2}^{\infty}\mathrm{Cov}(X_1,X_j) \in [0,\infty). \tag{4.5}$$

Proof. This theorem follows from Theorem 16 in the same way as Theorem 12 follows from Theorem 10. □

To investigate Theorem 17 in more detail, we restrict attention to $X_j = f(Y_j)$ even in the associated case. In this context, we define for $y,y' \in \mathcal{R}$

$$H(y,y') = H_1(y,y') + \Sigma_{j=2}^{\infty}[H_j(y,y') + H_j(y',y)], \tag{4.6}$$

where

$$H_j(y,y') = P(Y_1 > y,\ Y_j > y') - P(Y_1 > y)\,P(Y_j > y'). \tag{4.7}$$

PROPOSITION 18. *Let* $Y_1,Y_2, \ldots$ *be a strictly stationary (not necessarily* L_2*) sequence* *which is either associated or negatively associated. Then $H(y,y')$ exists for all y,y' with $0 \leq H_1 \leq H \leq \infty$ in the associated case and $-1 \leq H \leq H_1 \leq 1$ in the negatively associated case. H and H_1 are positive semidefinite in the sense that for any real g such that $g(y)H(y,y')g(y')$ is in $L_1(\mathcal{R}^2)$,*

(4.8) $$\int_{-\infty}^{\infty}\int_{-\infty}^{\infty} g(y)H(y,y')g(y')dy\,dy' \geq 0,$$

and similarly for H_1.

Proof. Denoting by $\rho_j(y)$ the indicator function of $\{Y_j > y\}$, and defining

$$H_{(n)}(y,y') = n^{-1}\mathrm{Cov}(\Sigma_{j=1}^{n}\rho_j(y), \Sigma_{k=1}^{n}\rho_k(y')),$$

we see that $H_{(n)}$ is positive semidefinite. In the associated case, the positivity of the H_j's implies that H exists, that $0 \leq H_1 \leq H \leq \infty$ and that $H = \lim H_{(n)}$ is positive semidefinite. In the negatively associated case, the negativity of the H_j's implies that H exists, that $H \leq H_1$, and that

$$\lim_{n\to\infty}[H_{(1)}(y,y') - H_{(n)}(y,y')] = -2\,\Sigma_{j=2}^{\infty}[H_j(y,y') + H_j(y',y)] \equiv G(y,y').$$

Since $H_{(n)}(y,y) \geq 0$ we have that $G(y,y) \leq H_{(1)}(y,y) = H_1(y,y)$ and thus that $H(y,y) = H_1(y,y) - G(y,y) \geq 0$. By the positive semidefiniteness of $H_{(n)}$, we have that

$$[H(y,y)\cdot H(y',y')] - [H(y,y')]^2 = \lim_{n\to\infty}([H_{(n)}(y,y)\cdot H_{(n)}(y',y')] - [H_{(n)}(y,y')]^2) \geq 0,$$

which implies that for any y,y',

$$|H(y,y')| \leq [H(y,y)\cdot H(y',y')]^{1/2} \leq [H_1(y,y)\cdot H_1(y',y')]^{1/2} \leq 1.$$

The positive semidefiniteness of $H = \lim H_{(n)}$ follows from that of $H_{(n)}$. □

Remark. In the associated case, $H(y,y')$ may equal $+\infty$ for some or all values of y,y'. For example, straightforward estimates show that when the Y_i's are jointly normal, then for any y,y', $\Sigma\, H_j(y,y')$ is absolutely convergent if and only if $\Sigma\,\mathrm{Cov}(Y_1,Y_j)$ is absolutely convergent.

We define D_H to be the set of real functions such that $g(y)H(y,y')g(y')$ is in $L_1(\mathcal{R}^2)$ and similarly for H_1, and we say that a real function f on $\mathcal{R}$ is absolutely continuous if it is the indefinite integral of a locally L_1 function f'. Note that $f' \in D_{H_1}$ if and only if the random variable, $\int_0^{Y_1}|f'(t)|dt$, has finite variance.

THEOREM 19. *Let $Y_1,Y_2,\ldots$ be a strictly stationary (not necessarily L_2) sequence which is either associated or negatively associated and let $X_j = f(Y_j)$ where f is an absolutely continuous function. Define $\bar{X}_j = \bar{f}(Y_j)$ where $\bar{f}(y) = \int_0^y|f'(t)|dt$. In the associated (resp. negatively associated) case, assume in addition that $f' \in D_H$ (resp. $\bar{X}_1$ is L_2); it follows that $\bar{X}_1$ is L_2 (resp. $f' \in D_H$) and that (4.4) is valid with σ^2 given by (4.5) or equivalently by*

(4.9) $$\sigma^2 = \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} f'(y)\,H(y,y')f'(y')\,dy\,dy'.$$

Proof. A straightforward generalization of (1.2) and (3.2) yields

(4.10) $$\mathrm{Cov}(g_1(Y_1),g_j(Y_j)) = \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} g'_1(y)g'_j(y')H_j(y,y')\,dy\,dy',$$

providing g_1, g_j are absoluted continuous and the random variables on the left hand side are L_2. If we take $g_1 = g_j = \bar{f}$ and $j = 1$, then the identity (4.10) shows that $\bar{X}_1$ is L_2 if and only if $f' \in D_{H_1}$. In the associated case the inequalities $H \geq H_1 \geq 0$ show that $f' \in D_H$ implies $f' \in D_{H_1}$. In the negatively associated case the positive semidefiniteness of H, the negativity of $H - H_1$, and the positivity of H_1 imply

$$\int|f'Hf'|\,dy\,dy' \leq \int|f'H_1f'|\,dy\,dy' + \int|f'(H-H_1)f'|dy\,dy'$$
$$= \int|f'|H_1|f'|\,dy\,dy' + \int|f'|(H_1-H)|f'|\,dy\,dy' \leq 2\int|f'|H_1|f'|\,dy\,dy',$$

which shows that $f' \epsilon D_{H_1}$ implies $f' \epsilon D_H$. In either case, we now obtain the desired results by applying Theorem 17 and using (4.10) first with $g_1 = g_j = \bar{f}$ to verify (4.3) and then with $g_1 = g_j = f$ to obtain (4.9). □

Remark. In the associated case, the automatic convergence of $\Sigma\mathrm{Cov}(Y_1,Y_j)$ implies by (2) that $H \epsilon L_1(\mathcal{R}^2)$ and then the hypothesis, $f' \epsilon D_H$, of Theorem 19 will be satisfied for any f with $f' \epsilon L_\infty(\mathcal{R}^1)$. In the associated case, if one assumes the convergence of $\Sigma_j \sup(H(y,y') : y,y' \epsilon \mathcal{R})$, then one would have H bounded on $\mathcal{R}^2$ while in the negatively associated case this is automatically the case; in either of these situations, if $X_j = f(Y_j)$ where f has bounded total variation then an $\bar{f}$ can be chosen to satisfy the hypothesis of Theorem 17. The central limit result of Theorem 17 can then be interpreted in terms of the scaled, centered, empirical distribution function,

$$I_n(y) = n^{-1/2}[(\text{number of } i\epsilon\{1, \ldots, n\} \text{ with } Y_i \leq y) - n\,P(Y_i \leq y)]$$
$$= n^{-1/2}[(\text{number of } i\epsilon\{1, \ldots, n\} \text{ with } Y_i > y) - n\,P(Y_i > y)];$$

the result is that I_n converges to a Gaussian process with mean zero and covariance $H(y,y')$ (at least) in the sense of convergence of finite dimensional distributions.

5. Related Results and Open Problems. We present the following triangular array central limit theorem without proof; for more details, see Cox and Grimmett (1982) where this theorem is proved as a consequence of Theorem 10.

THEOREM 20 *(Cox and Grimmett (1982)). Let* $S_n = \Sigma_{j=1}^{M_n}(Y_{n,j} - E\,Y_{n,j})$ *where for each n, the* $Y_{n,j}$*'s are LPQD. Suppose there exist* $c_1, c_2, c_3 \epsilon (0,\infty)$ *and a sequence* $u_l \to 0$ *so that for all* n,j,l, the following hold:

$$\mathit{(5.1)} \qquad Var(Y_{n,j}) \geq c_1,\; E(|Y_{n,j} - E\,Y_{n,j}|^3) \leq c_2.$$

$$(5.2) \qquad \Sigma_{k=1}^{M_n}\mathrm{Cov}(Y_{n,j}, Y_{n,k}) \leq c_3$$

$$(5.3) \qquad \sum_{\substack{k=1 \\ |k-j|\geq l}}^{M_n}\mathrm{Cov}(Y_{n,j}, Y_{n,k}) \leq u_l \quad ;$$

then

$$(5.4) \qquad \lim_{n\to\infty}(\mathrm{Var}(S_n))^{-1/2}\,S_n = Z.$$

Remark. There should be extensions of this theorem to the LNQD case and to sums of nonmonotonic functions of associated or negatively associated $Y_{n,j}$'s (analogous to the extension of Theorem 12 to Theorems 17 and 19); the details of such extensions have not been worked out.

The next theorem is due to Wood (1982; 1983). As in Theorem 20 the existence of absolute third moments is assumed; as a consequence a Berry-Esseen type result is obtained. In order to use this theorem to obtain uniform rates of convergence in the limit (3.7) of Theorem 12, one must control (usually in an ad hoc manner) the asymptotics of the parameters,

for more details and for the proof of the theorem, see Wood (1982; 1983).

THEOREM 21 *(Wood (1982)). Let* $S_n = Y_1 + \ldots + Y_n - nE(Y_1)$*, where* $Y_1, Y_2, \ldots$ *is* *a strictly stationary LPQD sequence with* $E(|Y_1|^3) < \infty$ *and such that*

(5.6) $$0 < \sigma^2 \equiv \mathrm{Var}(Y_1) + 2\,\Sigma_{j=2}^{\infty}\mathrm{Cov}(Y_1, Y_j) < \infty.$$

Then for $n = m \cdot k$,

(5.7) $$|P[n^{-1/2}S_n \leq z] - P[\sigma Z \leq z]| \leq 16\sigma_k^4 m(\sigma^2 - \sigma_k^2)/9\pi v_k^2 + 3v_k/\sigma_k^3 m^{1/2}.$$

The next result extends Theorem 12 in the associated case to an invariance principle (functional central limit theorem); such extensions in the negatively associated case and/or in the more general limits of Theorems 17 and 19 have not been investigated.

THEOREM 22 *(Newman and Wright (1981)). Let* $S_n = Y_1 + \ldots + Y_n - nEY_1$, *where* $Y_1, Y_2, \ldots$ *is a strictly stationary associated sequence with (5.6) valid. Define the stochastic processes,* $W_n(t)$ *for* $0 \leq t \leq T$ *by*

(5.8) $$W_n(t) = (\sigma^2 n)^{-1/2}[S_m + (nt - m)(Y_{m+1} - E\,Y_{m+1})] \text{ for } m/n \leq t \leq (m+1)/n;$$

then W_n *converges in distribution (on* $C[0,T]$ *to the standard Wiener process.*

Sketch of proof. A slight extension of Theorem 12 shows that the finite dimensional distributions of W_n converge to those of W. It remains to show that the distributions of W_n are tight. This is done in Newman and Wright (1981) as a consequence of the inequality (for $\alpha_2 - \alpha_1 > 1$),

(5.9) $$P[\mathrm{Max}(|S_l|, \ldots, |S_n|) \geq \alpha_2 \sqrt{n}\,\sigma_n] \leq [(\alpha_2 - \alpha_1)^2/((\alpha_2 - \alpha_1)^2 - 1)] \cdot P[|S_n| \geq \alpha_1 n^{1/2}\sigma_n],$$

which is derived by using the association of the Y_j's. □

A version of Theorem 12 for sequences indexed by d-dimensional parameters was already given in Newman (1980). The problem of obtaining a d-dimensional invariance principle for $d > 1$ by deriving appropriate d-dimensional maximal inequalities was solved in Newman and Wright (1982) for $d = 2$ by somewhat ad hoc methods; the problem is still open for $d > 2$. In the process of obtaining results for $d = 2$, the status of maximal and other inequalities for d$=1$ was clarified by realizing that there is a close connection between martingales and sums of associated variables. The following definition is due to Newman and Wright (1982).

Definition. An L_1 *sequence,* $S_0 = 0, S_1, S_2, \ldots$, *is a demimartingale (resp. demisubmartingale) if for* $j = 1, 2, \ldots$,

(5.10) $$E((S_{j+1} - S_j)f(S_1, \ldots, S_j)) \geq 0,$$

for all non decreasing (resp. nonnegative and nondecreasing) f *such that the expectation is defined.*

Note that with the natural choice of σ-fields, $S_0, S_1, \ldots$ would be a martingale (resp. submartingale) if the nondecreasing hypothesis were dropped. Note also that the assumption that the $(S_{j+1} - S_j)$'s are mean zero and associated implies that the sequence S_n is a demimartingale. It was shown in Newman and Wright (1982) that many standard martingale (resp. submartingale) inequalities, including Doob's maximal inequality and upcrossing inequality, remain valid for demimartingales (resp. demisubmartingales). In particular, we note that the inequalities of Corollary 6 of Newman and Wright (1982) are sufficient (without recourse to (5.9)) to yield tightness once convergence of finite dimensional distributions to those of a Wiener process is known. This fact, among others, suggests that both an ordinary and functional central limit theorem should be valid in the demimartingale context as it is in the martingale context (see, e.g. Billingsley (1968)).

CONJECTURE 23. *Let $S_0 = 0$, $S_1, S_2, \ldots$ be an L_2 demimartingale whose difference sequence $Y_1 = S_1 - S_0$, $Y_2 = S_2 - S_1, \ldots$ is strictly stationary and ergodic with (5.6) valid. Then W_n defined by (5.8) converges in distribution to the standard Wiener process; in particular,* $\lim n^{-1/2} S_n = \sigma Z$.

Remark. The status of maximal inequalities for sequences which satisfy (5.10) for all nonincreasing f has not yet been investigated. We do not consider that case further.

As a first step toward proving the above conjecture, we have the following result, presented here for the first time.

THEOREM 24. *Let $T_0 = 0$, $T_1, T_2, \ldots$ be an L_2 demimartingale and let $\mathcal{F}_n$ be the σ field generated by $T_0, T_1, \ldots, T_n$. If the T_j's have uncorrelated increments (i.e. if* $\mathrm{Cov}((T_{j+1} - T_j), (T_{k+1} - T_k)) = 0$ *for all* $0 \leq k < j$*), then the sequence $(T_n, \mathcal{F}_n)$ is a martingale.*

Proof. It follows immediately from the uncorrelated increment hypothesis that for each j,

$$E((T_{j+1} - T_j) T_k) = \mathrm{Cov}(T_{j+1} - T_j, T_k) = 0, \text{ for } k = 1, \ldots, j. \tag{5.11}$$

We have used the fact that $T_{j+1} - T_j$ has zero mean, as can be seen by taking f in (5.10) to be alternately $+1$ and -1. It suffices to show that

$$E((T_{j+1} - T_j) \exp[i \textstyle\sum_{k=1}^{j} r_k T_k]) = 0, \text{ for } r_1, \ldots, r_j \in \mathcal{R}, \tag{5.12}$$

in order to conclude that $E(Y_{j+1} | \mathcal{F}_j) = 0$ and thus that $(T_j, \mathcal{F}_j)$ is a martingale. The next proposition shows that (5.12) is a consequence of (5.11). □

PROPOSITION 25. *Suppose f and f_1 are complex functions on $\mathcal{R}^j$ such that $f << f_1$; then for a demimartingale T_n*

$$|E((T_{j+1} - T_j) f(T_1, \ldots, T_j))| \leq E((T_{j+1} - T_j) f_1(T_1, \ldots, T_j))$$

In particular this is the case for $f(t_1, \ldots, t_j) = \exp\,[i \sum_{k=1}^{j} r_k t_k]$ *and* $f_1(t_1, \ldots, t_j) = \sum_{k=1}^{j} |r_k| t_k$.

Proof. The proposition follows easily from Proposition 14 and a portion of the proof of Proposition 15. □

Remark. To clarify somewhat the distinction between martingales, sums of associated variables, and demimartingales we let $S_0 = 0$, $S_n = Z_1 + \ldots + Z_n (n = 1, 2, \ldots)$ where the Z_j's are jointly normal. The sequence S_n is a martingale (resp. submartingale) if and only if $\mathrm{Cov}(Z_k, Z_l) = 0$ for all $k > l$ and $E\,Z_j = 0$ (resp. $E\,Z_j \geq 0$) for all j. By the results of Pitt (1982), the Z_j's are associated if and only if $\mathrm{Cov}(Z_k, Z_l) \geq 0$ for all $k > l$ while the sequence S_n is a demimartingale (resp. subdemimartingale) if and only if $\sum_{j=1}^{\ell} \mathrm{Cov}(Z_k, Z_j) \geq 0$ for all $k > \ell$ and $EZ_j \geq 0$) for all j.

Acknowledgement. The author wishes to thank the University of Arizona, The National Science Foundation, the Courant Institute of Mathematical Sciences of New York University, the Lady Davis Fellowship Trust, and the Institute of Mathematics and Computer Science of the Hebrew University of Jerusalem for their support and hospitality during his sabbatical.

REFERENCES

BILLINGSLEY, P. (1968). *Convergence of Probability Measures*. Wiley, New York.

BLOCK, H. W., SAVITS, T. H. and SHAKED, M. (1982). Some concepts of negative dependence. *Ann. Probab. 10* 765–772

Cox, J. T. and Grimmett, G. (1982). Central limit theorems for associated random variables and the percolation model. Syracuse Univ. preprint.

Esary, J., Proschan, F. and Walkup, D. (1967). Association of random variables with applications. *Ann. Math. Statist.* *38* 1466–1474.

Fortuin, C. Kastelyn, P. and Ginibre, J. (1971). Correlation inequalities on some partially ordered sets. *Comm. Math. Phys.* *22* 89–103.

Harris, T. E. (1960). A lower bound for the critical probability in a certain percolation process. *Proc. Camb. Phil. Soc.* *59* 13–20.

Herrndorf, N. (1983). An example on the central limit theorem for associated sequences. Universität Köln preprint.

Joag-Dev, K. (1983). Independence via uncorrelatedness under certain dependence structures. *Ann. Probab.* *11* 1037–1041.

Joag-Dev, K. and Proschan, F. (1983). Negative association of random variables with applications. *Ann. Statist.* *11* 286–295.

Karlin, S. and Rinott, Y. (1980a). Classes of orderings of measures and related correlation inequalities. I. Multivariate totally positive distributions. *J. Multivariate Anal.* *10* 467–498.

Karlin, S. and Rinott, Y. (1980b). Classes of orderings of measures and related correlation inequalities. II. Multivariate reverse rule distribution. *J. Multivariate Anal.* *10* 499–516.

Lebowitz, J. (1972). Bounds on the correlations and analyticity properties of ferromagnetic Ising spin systems. *Comm. Math. Phys.* *28* 313–321.

Lehmann, E. L. (1966). Some concepts of dependence. *Ann. Math. Statist.* *37* 1137–1153.

Newman, C. M. (1980). Normal fluctuations and the FKG inequalities. *Comm. Math. Phys.* *74* 119–128.

Newman, C. M. (1983). A general central limit theorem for FKG systems. *Comm. Math. Phys.* *91* 75–80.

Newman, C. M., Rinott, Y. and Tversky, A. (1982). Nearest neighbors and Voronoi regions in certain point processes. (to appear in *Adv. Appl. Probab.*)

Newman, C. M. and Schulman, L. S. (1981). Infinite clusters in percolation models. *J. Statist. Phys.* *26* 613–628.

Newman, C. M. and Wright, A. L. (1981). An invariance principle for certain dependent sequences. *Ann. Probab.* *9* 671–675.

Newman, C. M. and Wright, A. L. (1982). Associated random variables and martingale inequalities. *Z. Wahrscheinlichkeitstheorie und Verw. Gebiete* *59* 361–371.

Pitt, L. (1982). Positively correlated normal variables are associated. *Ann. Probab.* *10* 496–499.

Sarkar, T. K. (1969). Some lower bounds of reliability. *Tech. Report. No. 124,* Department of Operations Research and Statistics, Stanford University, Stanford, CA.

Shaked, M. (1982a). A general theory of some positive dependence notions. *J. Multivariate Anal.* *12* 199–218.

Shaked, M. (1982b). Private communication.

Simon, B. (1973). Correlation inequalities and the mass gap in $P(\phi)_2$. I. Domination by the two point function. *Comm. Math. Phys.* *31* 127–136.

Wells, D. R. (1977). Some moment inequalities and a result on multivariate unimodality. Indiana Univ. Ph.D. Dissertation.

Wood, T. E. (1982). Sequences of associated random variables. Univ. of Virginia Ph.D. Dissertation.

Wood, T. E. (1983). A Berry-Esseen theorem for associated random variables. *Ann. Probab.* *11* 1042–1047.

Inequalities in Statistics and Probability
IMS Lecture Notes–Monograph Series Vol. 5 (1984), 141–149

STOCHASTIC ORDERING OF SPACINGS FROM DEPENDENT RANDOM VARIABLES

BY MOSHE SHAKED[1] and Y. L. TONG[2]
University of Arizona and University of Nebraska

Spacings (that is, the differences between successive order statistics) are useful in various applications in statistics. Many properties of the spacing are known when the spacings are constructed from a collection of independent identically distributed (i.i.d.) random variables. In this paper we study the spacings constructed from not necessarily i.i.d. random variables. We introduce models for which two sets of spacings, constructed from two sets of dependent random variables, can be stochastically ordered. Various examples will be given and applications for goodness-of-fit tests, tests for independence, density estimation and tests for outliers will be discussed.

1. Introduction. Let $\mathbf{X} = (X_1, \ldots, X_n)$ denote an n-dimensional random vector and let

$$X_{(1)} \leq \ldots \leq X_{(n)}$$

be the ordered components (order statistics) of $\mathbf{X}$. The nonnegative random variables

$$U_i = X_{(i+1)} - X_{(i)},\ i = 1, \ldots, n-1$$

are called the *spacings* and have various applications in statistics. For example, certain nonparametric test procedures depend on the maximum spacing or on linear combinations of spacings (see, e.g., Pyke (1965), Weiss (1965), Rao and Sethuraman (1970) and Kirmani and Alam (1974)); certain estimation and test procedures based on order statistics, such as those which depend on the range or midrange, involve spacings (David (1970), Ch. 6); and certain tests for slippage (Karlin and Truax (1960)) and outliers (Barnett and Lewis (1978), Ch. 3) also depend on spacings. For a comprehensive treatment of spacings see Pyke (1965, 1972).

In the literature the problem of spacings has been treated extensively under the assumption that $X_1, \ldots, X_n$ are independent and identically distributed (i.i.d.) random variables. In certain applications which involve a mixture of experiments, a (random) change of scale or a random shift in location may take place; then the random variables $X_1, \ldots, X_n$ are no longer independent. In this paper we study how the degree of dependence affects the distribution of the spacings. In the case when $X_1, \ldots, X_n$ are interchangeable, it follows from our main result that (in the model under consideration) the spacings vector $\mathbf{U} = (U_1, \ldots, U_{n-1})$ becomes stochastically smaller if $X_1, \ldots, X_n$ are more positively dependent (that is, when $X_1, \ldots, X_n$ have more tendency to "hang together").

After stating the model and proving the main result in Section 2, we apply the result to an additive, a multiplicative and a ratio model. In Section 3, after combining results given in Shaked and Tong (1985), we obtain a partial ordering property for the spacings which correspond to a number of important multivariate distributions, such as the multivariate normal, multivariate stable, multivariate beta and the Dirichlet distribution. For all these distributions the corresponding spacings vector $\mathbf{U}$ can be partially ordered through the degree of dependence of the components $X_1, \ldots, X_n$ of $\mathbf{X}$.

---

[1] Research partially supported by the NSF Grant MCS-8200098.
[2] Research partially supported by the NSF Grant MCS-8100775.
AMS 1980 subject classifications. Primary 60E15; secondary 62G30.
Key words and phrases. Stochastic ordering of spacings, dependent random variables.

In Section 4 we give applications and study the monotonicity properties of certain well-known procedures concerning goodness-of-fit tests, tests for independence, density estimation and slippage tests for outliers, which all depend on spacings.

2. The Model. Let $\mathbf{X} = (X_1, \dots, X_n)$ and $\mathbf{Y} = (Y_1, \dots, Y_n)$ denote two n-dimensional random vectors and let

$$X_{(1)} \leq \dots \leq X_{(n)}, Y_{(1)} \leq \dots \leq Y_{(n)}$$

be their ordered components. Define the $(n-1)$-dimensional spacings vectors by

$$\mathbf{U} = (U_1, \dots, U_{n-1}) \text{ where } U_i = X_{(i+1)} - X_{(i)},\ i = 1, \dots, n-1.$$
$$\mathbf{V} = (V_1, \dots, V_{n-1}) \text{ where } V_i = Y_{(i+1)} - Y_{(i)},\ i = 1, \dots, n-1.$$

The stochastic ordering of **U** and **V** will be developed under the following model:

Model A. There exist a random vector $\mathbf{Z} = (Z_1, \dots, Z_n)$, random vectors (of any dimension) $\mathbf{W}_1, \mathbf{W}_2, \mathbf{W}_3, \mathbf{W}_4$ and Borel-measurable functions ϕ, δ_1 and δ_2 such that

$$(X_1, \dots, X_n) \stackrel{d}{=} (Z_1\phi(\mathbf{Z}, \mathbf{W}_1) + \delta_1(\mathbf{Z}, \mathbf{W}_3), \dots, Z_n\phi(\mathbf{Z}, \mathbf{W}_1) + \delta_1(\mathbf{Z}, \mathbf{W}_3))$$
$$(Y_1, \dots, Y_n) \stackrel{d}{=} (Z_1\phi(\mathbf{Z}, \mathbf{W}_2) + \delta_2(\mathbf{Z}, \mathbf{W}_4), \dots, Z_n\phi(\mathbf{Z}, \mathbf{W}_2) + \delta_2(\mathbf{Z}, \mathbf{W}_4)).$$

Moreover, the following conditions are satisfied:

A1. $Z_1, \dots, Z_n$ are i.i.d., $\mathbf{W}_i$ is independent of $\mathbf{Z}$, $i = 1,2$.

A2. $\phi(\mathbf{z},\mathbf{w})$, $\delta_1(\mathbf{z},\mathbf{w})$ and $\delta_2(\mathbf{z},\mathbf{w})$ are permutation symmetric functions of $\mathbf{z} = (z_1, \dots, z_n)$ for every fixed $\mathbf{w}$, and $\phi > 0$ for all $(\mathbf{z},\mathbf{w})$ whenever $\mathbf{z}$ is in the support of $\mathbf{Z}$ and $\mathbf{w}$ is in the support of $\mathbf{W}_1$ or of $\mathbf{W}_2$.

A3. Either $\phi(\mathbf{z},\mathbf{w})$ is nondecreasing (componentwise) in $\mathbf{w}$ for every $\mathbf{z}$ and $\mathbf{W}_1 \stackrel{st}{\geqslant} \mathbf{W}_2$, or $\phi(\mathbf{z},\mathbf{w})$ is nonincreasing in $\mathbf{w}$ for every $\mathbf{z}$ and $\mathbf{W}_1 \stackrel{st}{\leqslant} \mathbf{W}_2$.

THEOREM 1. *Assume that* **X** *and* **Y** *have the representation of Model* A *and that* A1, A2 *and* A3 *are satisfied. Then, for all k and all constants* c_{ij}, $i = 1, \dots, n$, $j = 1, \dots, k$, *such that* $\sum_{i=1}^n c_{ij} = 0, j = 1, \dots, k$,

$$E\psi(|\textstyle\sum_{i=1}^n c_{i1}X_{(i)}|, \dots, |\sum_{i=1}^n c_{ik}X_{(i)}|)$$
$$\geqslant E\psi(|\textstyle\sum_{i=1}^n c_{i1}Y_{(i)}|, \dots, |\sum_{i=1}^n c_{ik}Y_{(i)}|)$$

holds for all ψ *which are componentwise nondecreasing such that the expectations exist. Consequently,*

$$\mathbf{U} \stackrel{\text{st}}{\geqslant} \mathbf{V}. \tag{2.1}$$

Proof. Let $Z_{(1)} \leq \dots \leq Z_{(n)}$ denote the order statistics of $\mathbf{Z} = (Z_1, \dots, Z_n)$. Since ϕ, δ_1, and δ_2 are permutation symmetric in $z_1, \dots, z_n$ for every fixed $\mathbf{w}$, we must have, a.s.,

$$\phi(\mathbf{Z},\mathbf{W}_j) = \phi(Z_{(1)}, \dots, Z_{(n)}, \mathbf{W}_j)$$
$$\delta_j(\mathbf{Z},\mathbf{W}_{2+j}) = \delta_j(Z_{(1)}, \dots, Z_{(n)}, \mathbf{W}_{2+j})$$

for $j = 1,2$. This implies that

$$Z_i\phi(\mathbf{Z},\mathbf{W}_j) + \delta_j(\mathbf{Z},\mathbf{W}_{j+2}) \leqslant Z_{i'}\phi(\mathbf{Z},\mathbf{W}_j) + \delta_j(\mathbf{Z},\mathbf{W}_{j+2})$$

holds if and only if $Z_i \leqslant Z_{i'}$. Consequently,

$$(X_{(1)}, \dots, X_{(n)})$$
$$\stackrel{d}{=} (Z_{(1)}\phi(\mathbf{Z},\mathbf{W}_1) + \delta_1(\mathbf{Z},\mathbf{W}_3), \dots, Z_{(n)}\phi(\mathbf{Z},\mathbf{W}_1) + \delta_1(\mathbf{Z},\mathbf{W}_3))$$
$$= \phi(\mathbf{Z},\mathbf{W}_1)(Z_{(1)}, \dots, Z_{(n)}) + (\delta_1(\mathbf{Z},\mathbf{W}_3), \dots, \delta_1(\mathbf{Z},\mathbf{W}_3)).$$

Hence, for all c_{ij}, $i = 1, \ldots, n, j = 1, \ldots, k$, such that $\sum_{i=1}^{n} c_{ij} = 0$,

$$(\Sigma_{i=1}^{n} c_{i1}X_{(i)} \ldots , \Sigma_{i=1}^{n} c_{ik}X_{(i)}) \stackrel{d}{=} \phi(\mathbf{Z},\mathbf{W}_1)(\Sigma_{i=1}^{n} c_{i1}Z_{(i)}, \ldots , \Sigma_{i=1}^{n} c_{ik}Z_{(i)}).$$

Without loss of generality assume that ϕ is nondecreasing in $\mathbf{w}$ and that $\mathbf{W}_1 \stackrel{st}{\geq} \mathbf{W}_2$. Then

$$\begin{aligned} &E\psi(|\Sigma_{i=1}^{n} c_{i1}X_{(i)}|, \ldots , |\Sigma_{i=1}^{n} c_{ik}X_{(i)}|) \\ &\quad = E\psi(\phi(\mathbf{Z},\mathbf{W}_1)|\Sigma_{i=1}^{n} c_{i1}Z_{(i)}|, \ldots , \phi(\mathbf{Z},\mathbf{W}_1)|\Sigma_{i=1}^{n} c_{ik}Z_{(i)}|) \\ &\quad = E\zeta_1(\mathbf{Z}) \end{aligned}$$

where $\zeta_1(\mathbf{z})$ is the conditional expectation of ψ, over the distribution of $\mathbf{W}_1$, given $\mathbf{Z} = \mathbf{z}$. Let $\zeta_2(\mathbf{z})$ denote the similar conditional expectation of ψ over the distribution of $\mathbf{W}_2$. Since ϕ is a nondecreasing function of $\mathbf{w}$ it follows that $\psi(\phi(\mathbf{z},\mathbf{w})|\Sigma_{i=1}^{n} c_{i1}z_{(i)}|, \ldots , \phi(\mathbf{z},\mathbf{w})|\Sigma_{i=1}^{n} c_{ik}z_{(i)}|)$ is also a nondecreasing function of $\mathbf{w}$ for every fixed $\mathbf{z}$. Thus, $\mathbf{W}_1 \stackrel{st}{\geq} \mathbf{W}_2$ implies that

$$\zeta_1(\mathbf{Z}) \geq \zeta_2(\mathbf{Z}) \text{ a.s.}$$

The proof is now completed by applying the equality

$$E\psi(|\Sigma_{i=1}^{n} c_{i1}Y_{(i)}|, \ldots , |\Sigma_{i=1}^{n} c_{ik}Y_{(i)}|) = E\zeta_2(\mathbf{Z}).$$

The last statement of the theorem follows by defining $k = n-1$, $c_{i+1,i} = 1$, $c_{i,i} = -1$, $c_{i,j} = 0, j \neq i, i+1$ where $i = 1, \ldots, n-1$.

In the following we consider special forms of Model A.

(a) (An additive model). Assume that there exist constants $a \geq 0$, b and d (b and d have the same sign) and independent random variables $Z_1, \ldots , Z_n$ and W such that

$$\begin{aligned} (X_1, \ldots , X_n) &\stackrel{d}{=} (aZ_1 + bW, \ldots , aZ_n + bW) \\ (Y_1, \ldots , Y_n) &\stackrel{d}{=} (cZ_1 + dW, \ldots , cZ_n + dW). \end{aligned} \tag{2.2}$$

Note that without loss of generality we can assume that $b \geq 0$ and $d \geq 0$ because otherwise one can replace W by $-W$.

Letting W_1 and W_2 be degenerate at $a > 0$ and $c > 0$, respectively, letting $W_3 \stackrel{d}{=} W_4 \stackrel{d}{=} W$, setting $\phi(\mathbf{z},w) = w$, $\delta_1(\mathbf{z},w) = bw$ and $\delta_2(\mathbf{z},w) = dw$ and assuming $a \geq c$, it is easy to see that $\mathbf{X}$ and $\mathbf{Y}$ of (2.2) have the representation of Model A and satisfy A1, A2 and A3 provided $Z_1, \ldots , Z_n$ are i.i.d.

In some applications (see Shaked and Tong (1985)) $\mathbf{X}$ and $\mathbf{Y}$ have the same marginals, that is,

$$X_i \stackrel{d}{=} Y_i, \; i = 1, \ldots , n. \tag{2.3}$$

In other applications the following condition, which is weaker than (2.3), holds:

$$\Sigma_{i=1}^{n} EX_i = \Sigma_{i=1}^{n} EY_i \tag{2.4}$$

For example, (2.4) holds if $EZ_1 = \ldots = EZ_n = EW = 0$ or if $EZ_1 = \ldots = EZ_n = EW$ and $a + b = c + d$.

Let $U_i = X_{(i+1)} - X_{(i)}$ and $V_i = Y_{(i+1)} - Y_{(i)}$, $i = 1, \ldots , n-1$. From Theorem 1 it follows that if $\mathbf{X}$ and $\mathbf{Y}$ satisfy (2.2) with $a \geq c$ then

$$(U_1, \ldots , U_{n-1}) \stackrel{st}{\geq} (V_1, \ldots , V_{n-1}). \tag{2.5}$$

Shaked and Tong (1985) considered the condition

$$(EX_{(1)}, \ldots , EX_{(n)}) \succ (EY_{(1)}, \ldots , EY_{(n)}). \tag{2.6}$$

that is,

$$\Sigma_{i=1}^{k} EX_{(i)} \leq \Sigma_{i=1}^{k} EY_{(i)}, \; k = 1, \ldots , n-1$$

and

$$\Sigma_{i=1}^{n} EX_{(i)} = \Sigma_{i=1}^{n} EY_{(i)}.$$

They denoted the relation (2.6) by $\mathbf{X} \underset{D}{\succ} \mathbf{Y}$ and discussed some applications. Following their arguments it follows that if $\mathbf{X}$ and $\mathbf{Y}$ satisfy (2.5) and (2.4) then $\mathbf{X} \underset{D}{\succ} \mathbf{Y}$. Thus

PROPOSITION 1. *If* $\mathbf{X}$ *and* $\mathbf{Y}$ *satisfy* (2.2) *with* $a \geq c$ *and* (2.4), *then* $\mathbf{X} \underset{D}{\succ} \mathbf{Y}$.

The special case where $\mathbf{X}$ and $\mathbf{Y}$ have the representation

$$(2.7) \qquad \begin{aligned} (X_1, \ldots, X_n) &\stackrel{d}{=} (1-\rho)^{1/\alpha}(Z_1, \ldots, Z_n) + \rho^{1/\alpha}(W, \ldots, W), \\ (Y_1, \ldots, Y_n) &\stackrel{d}{=} (1-\eta)^{1/\alpha}(Z_1, \ldots, Z_n) + \eta^{1/\alpha}(W, \ldots, W), \end{aligned}$$

where $Z_1, \ldots, Z_n$ and W are as in (2.2), $0 \leqslant \rho < \eta \leqslant 1$ and $\alpha > 0$, is Model 4.1 in Shaked and Tong (1985).

Note that in this special case we can actually have a stronger statement concerning the distribution of $\mathbf{U}$ and $\mathbf{V}$ corresponding to $\mathbf{X}$ and $\mathbf{Y}$. That is,

$$(2.8) \qquad \begin{aligned} (U_1, \ldots, U_{n-1}) &\stackrel{d}{=} (1-\rho)^{1/\alpha}(Z_{(2)}-Z_{(1)}, \ldots, Z_{(n)}-Z_{(n-1)}) \\ (V_1, \ldots, V_{n-1}) &\stackrel{d}{=} (1-\eta)^{1/\alpha}(Z_{(2)}-Z_{(1)}, \ldots, Z_{(n)}-Z_{(n-1)}). \end{aligned}$$

Thus we have

$$(2.9) \qquad (U_1, \ldots, U_{n-1}) \stackrel{d}{=} \left(\frac{1-\rho}{1-\eta}\right)^{1/\alpha}(V_1, \ldots, V_{n-1}).$$

Consequently

$$(2.10) \qquad \Sigma_{i=1}^{n-1}\lambda_i U_i \stackrel{d}{=} \left(\frac{1-\rho}{1-\eta}\right)^{1/\alpha}\Sigma_{i=1}^{n-1}\lambda_i V_i$$

for all $\lambda_1, \ldots, \lambda_{n-1}$.

(b) (A multiplicative model). Assume $\mathbf{X}$ and $\mathbf{Y}$ have the representation

$$\begin{aligned} (X_1, \ldots, X_n) &\stackrel{d}{=} W_1(Z_1, \ldots, Z_n) + (W_3, \ldots, W_3), \\ (Y_1, \ldots, Y_n) &\stackrel{d}{=} W_2(Z_1, \ldots, Z_n) + (W_4, \ldots, W_4), \end{aligned}$$

where W_1 and $\mathbf{Z}$ are independent, W_2 and $\mathbf{Z}$ are independent and $Z_1, \ldots, Z_n$ are i.i.d. If W_1 and W_2 are nonnegative a.s. and $W_1 \overset{st}{\geqslant} W_2$ then this is a special case of Model A and Theorem 1 applies.

(c) (A ratio model). In certain situations $X_1, \ldots, X_n$ have the representation

$$(X_1, \ldots, X_n) \stackrel{d}{=} (Z_1 /(\Sigma_{i=1}^{n} h(Z_i) + W_1), \ldots, Z_n /(\Sigma_{i=1}^{n} h(Z_i) + W_1)),$$

where $Z_1, \ldots, Z_n$ and W_1 are independent and $Z_1, \ldots, Z_n$ are i.i.d. [When $Z_i \geqslant 0$, $W_1 > 0$ a.s. and h is the identity function, then

$$(X_1, \ldots, X_n) \stackrel{d}{=} (Z_1/(\Sigma_{i=1}^{n} Z_i + W_1), \ldots, Z_n/(\Sigma_{i=1}^{n} Z_i + W_1)).]$$

In this model if

$$(Y_1, \ldots, Y_n) \stackrel{d}{=} (Z_1 /(\Sigma_{i=1}^{n} h(Z_i) + W_2), \ldots, Z_n /(\Sigma_{i=1}^{n} h(Z_i) + W_2)),$$

$h \geqslant 0$ and $W_2 > 0$ a.s., then $W_2 \overset{st}{\geqslant} W_1$ implies $\mathbf{U} \overset{st}{\geqslant} \mathbf{V}$. This follows from Theorem 1 because in Model A one can take

$$\begin{aligned} \phi(\mathbf{z}, w) &= [\Sigma_{i=1}^{n} h(z_i) + w]^{-1} \\ \delta_1(\mathbf{z}, w) &= \delta_2(\mathbf{z}, w) = 0. \end{aligned}$$

We note in passing that it is easy to show that $\mathbf{U} \overset{st}{\geqslant} \mathbf{V}$ also for spacings vectors constructed from $\mathbf{X}$ and $\mathbf{Y}$ which satisfy the following model:

Model B. There exist a random vector $\mathbf{Z} = (Z_1, \ldots, Z_n)$, a random vector $\mathbf{W}$ and Borel-measurable functions ϕ_1, ϕ_2, δ_1 and δ_2 such that

$$(X_1, \ldots, X_n) \stackrel{d}{=} (Z_1\phi_1(\mathbf{Z},\mathbf{W}) + \delta_1(\mathbf{Z},\mathbf{W}), \ldots, Z_n\phi_1(\mathbf{Z},\mathbf{W}) + \delta_1(\mathbf{Z},\mathbf{W})),$$
$$(Y_1, \ldots, Y_n) \stackrel{d}{=} (Z_1\phi_2(\mathbf{Z},\mathbf{W}) + \delta_2(\mathbf{Z},\mathbf{W}), \ldots, Z_n\phi_2(\mathbf{Z},\mathbf{W}) + \delta_2(\mathbf{Z},\mathbf{W})).$$

Moreover, the following conditions are satisfied:

B1. $Z_1, \ldots, Z_n$ are i.i.d. and $\mathbf{W}$ and $\mathbf{Z}$ are independent.

B2. $\phi_1(\mathbf{z},\mathbf{w})$, $\phi_2(\mathbf{z},\mathbf{w})$, $\delta_1(\mathbf{z},\mathbf{w})$, and $\delta_2(\mathbf{z},\mathbf{w})$ are permutations symmetric functions of $z_1, \ldots, z_n$ for every fixed $\mathbf{w}$, and $\phi_1 \geqslant \phi_2 > 0$ over the support of $(\mathbf{Z},\mathbf{W})$.

The main difference between Models A and B is that in Model B we have two functions ϕ_1 and ϕ_2 compared to the single function ϕ of Model A. But, in Model B, ϕ_1 and ϕ_2 are not required to be monotone.

We end this section by showing that Model 4.2 (unlike Model 4.1) of Shaked and Tong (1985), which involves positive dependence by mixture, does not necessarily imply the basic relation (2.1).

A random vector $\mathbf{Y}$ is called positively dependent by mixture (PDM) if there exists a random vector $\mathbf{W}$ such that, given $\mathbf{W} = \mathbf{w}$, $Y_1, \ldots, Y_n$ are conditionally i.i.d. Shaked (1977) and Shaked and Tong (1985) showed that, in some respects, a PDM random vector $\mathbf{Y}$ is more positively dependent than a random vector $\mathbf{X}$ of i.i.d. random variables where $\mathbf{X}$ and $\mathbf{Y}$ have the same marginals. One can expect that the spacings vectors $\mathbf{U}$ and $\mathbf{V}$, corresponding to $\mathbf{X}$ and $\mathbf{Y}$ satisfy $\mathbf{U} \stackrel{st}{\geqslant} \mathbf{V}$. The following example shows that this is not necessarily the case.

Example. Let Y_1 and Y_2 have the joint probabilities

$y_1 \backslash y_2$	1	2	3
1	1/6	0	1/6
2	0	1/3	0
3	1/6	0	1/6

and let X_1 and X_2 be i.i.d. such that $X_i \stackrel{d}{=} Y_i$, $i = 1,2$, that is $P[X_1 = 1] = P[X_1 = 2] = P[X_1 = 3] = 1/3$. Then $P[V_1 \geqslant 2] = P[Y_{(2)} - Y_{(1)} = 2] = 1/3$ whereas $P[U_1 \geqslant 2] = P[X_{(2)} - X_{(1)} = 2] = 2/9$. Hence it is not true that $U_1 \stackrel{st}{\geqslant} V_1$ although (Y_1, Y_2) is PDM.

It follows that if $\mathbf{X}$ and $\mathbf{Y}$ satisfy Model 4.2 of Shaked and Tong (1985) then it is not necessarily true that the corresponding spacings satisfy $\mathbf{U} \stackrel{st}{\geqslant} \mathbf{V}$.

3. Examples. In this section we describe some examples of well-known distributions for which the results of Section 2 apply.

(a) *Exchangeable normal variables.* Let $\mathbf{X}$ be a multivariate normal random vector with means μ_x, variances σ^2 and correlations ρ; let $\mathbf{Y}$ be another multivariate normal random vector with means μ_y, variances σ^2 and correlations η. If $0 \leqslant \rho < \eta \leqslant 1$ then $\mathbf{U} \stackrel{st}{\geqslant} \mathbf{V}$. This follows from (2.7) with $\alpha = 2$ where $Z_1, \ldots, Z_n$ and W are i.i.d. normal random variables with mean 0 and variance σ^2. **Note that adding μ_x to all X_i's and μ_y to all Y_i's does not change the distributions of U and V.**

(b) ***Multivariate Cauchy and stable variables.*** It is shown in Shaked and Tong (1985) that some exchangeable stable random vectors **X** and **Y** have the representation (2.7). Hence Theorem 1 applies.

(c) *Multivariate Dirichlet and beta variables*. Let **X** have the distribution defined by

$$(X_1, \dots, X_n) = (Z_1/(\Sigma_{i=1}^n Z_i + W), \dots, Z_n/(\Sigma_{i=1}^n Z_i + W))$$

where $Z_1, \dots, Z_n$ are i.i.d. gamma random variables (for Dirichlet) or i.i.d. chi-squared random variables (for multivariate beta), W is a gamma random variable or a chi-squared random variable (with the same shape parameter but possibly with different scale parameter) and **Z** and W are independent. In this case, as is shown in Section 2, a partial ordering of the spacings can be obtained via the value of the shape parameter of W.

Note that in this case $X_1, \dots, X_n$ are not positively dependent. Actually they are negatively correlated. The result of Theorem 1 can be interpreted here by saying that the less negatively dependent are $X_1, \dots, X_n$ the smaller stochastically are the corresponding spacings.

4. Applications

4.1. Goodness of fit tests. Let $Z_1, \dots, Z_n$ be random variables, let $Z_{(1)}, \dots, Z_{(n)}$ be the corresponding order statistics and let $U_i = Z_{(i+1)} - Z_{(i)}$, $i = 1, \dots, n-1$, be the corresponding spacings. Statistics like the largest spacing, the kth smallest spacing, partial sums of ordered spacings, etc., have been used in statistical literature to construct tests of goodness of fit and related hypotheses (see, e.g., Rao and Sobel (1980) and references there). Pyke (1965) discussed statistics of the form $\Sigma_{i=1}^{n-1} g(U_i)$ where g is some monotone function. A general form for all the statistics mentioned above is $\Sigma_{i=1}^{n-1} g_i(U_i)$ where the g_i's are all monotone in the same direction (see Weiss (1957)).

In most applications the Z_i's are i.i.d. with a common distribution F, say, and one is interested to test $H_0: F = F_0$ where F_0 is a given distribution (which may or may not depend on some unknown parameters). The hypothesis H_0 is then rejected if $\Sigma_{i=1}^{n-1} g_i(U_i)$ is large; tables of critical values have been prepared for various choices of the g_i's.

In some practical situations it may happen that the assumption of independence of the observations is not valid. For example, a random shift of all the observations combined with a change of scale may transform the Z_i's into dependent random variables with the same (or with different) marginals as the original Z_i's (we denote these dependent random variables then by Y_i's). For example, assume that the Z_i's are i.i.d. normal random variables with mean μ and variance σ^2. Define

$$Y_i = (1-p)^{1/2} Z_i + p^{1/2} Z, \; i = 1, \dots, n.$$

where Z is a random shift, independent of $Z_1, \dots, Z_n$ and having a normal distribution with mean λ and variance σ^2. Then each Y_i is a normal random variable with mean $(1-p)^{1/2}\mu + \rho^{1/2}\lambda$ and variance σ^2, but now the Y_i's are not independent. In most applications $\mu = \lambda = 0$ (the condition $\lambda = 0$ means that, on the average, the random shift is zero), so that the Y_i's have the same common marginal distribution as the original Z_i's, but they are not independent. It follows from Theorem 1, then, that if one uses the test statistic $\Sigma_{i=1}^{n-1} g_i(U_i)$, where the g_i's are nondecreasing, then one has a smaller probability of rejection of H_0 than intended. Actually, Theorem 1 shows that the more dependent the Y_i's are, the smaller is the probability of rejection of H_0. The opposite is true if the g_i's are nonincreasing.

Of course the above analysis is valid whenever the Y_i's are distributed according to any multivariate stable distribution (see Section 3).

Relation (2.10) is particularly useful in this setting. To see this let $\rho = 0$ in (2.7) and

$\eta \in (0,1)$. Then $(X_1, \ldots, X_n) = (Z_1, \ldots, Z_n)$, i.e. $X_1, \ldots, X_n$ are i.i.d. whereas $Y_1, \ldots, Y_n$ are not independent. The corresponding spacings **U** and **V**, constructed from **X** and **Y**, respectively, satisfy (2.8) and hence (2.9) and (2.10). If η is known and if the test statistic is $\Sigma_{i=1}^{n-1} \lambda_i U_i$ then the critical values for testing H_0 (mentioned above) can be obtained from existing tables by multiplying the tabulated values by $(1-\eta)^{1/\alpha}$ [recall that here we take $\rho = 0$].

We remark in passing that comments which are similar to the above apply to any statistic which is a monotone function of the spacings. In particular, they apply to the statistics discussed in del Pino (1979) which are monotone functions of the k-spacings: $Z_{(k+1)}-Z_{(1)}, \ldots, Z_{(n)}-Z_{(n-k)}$.

4.2. Tests for independence. The discussion in Section 4.1 shows that if observations are dependent in the sense Model A instead of being independent then the significance levels of many tests, which use these observations, are different than the desired ones. In particular, if the g_i's are nondecreasing then the probability of rejection decreases as the observations become more positively dependent.

One way to interpret this discussion is to observe that the statistics $\Sigma_{i=1}^{n-1} g_i(U_i)$ actually yield unbiased tests for the hypothesis which claims that the random variables are independent versus the alternative which claims, for example, that they are positive dependent in the sense (2.2) with $a \geq c$ *and* $d \geq b = 0$.

We note in passing that the resulting tests are not necessarily optimal in any sense. We do not try to derive here any optimality property for any test. We remark, however, that one advantage of the above tests is that existing tables of critical values of statistics of the form $\Sigma_{i=1}^{n-1} g_i(U_i)$ and existing results about their asymptotic distributions (see Pyke (1965) and references there) may be applied for testing the hypothesis of independence mentioned above.

4.3. Empirical distributions and quantile function estimates. Let $Z_1, \ldots, Z_n$ be identically distributed random variables with a common distribution F. Let $Z_{(1)} \leq \ldots \leq Z_{(n)}$ be the corresponding order statistics. The empirical distribution function, $\hat{F}$, is (denoting $Z_{(n+1)} = \infty$)

$$\begin{aligned}\hat{F}(z) &= 0 \quad \text{if } z < Z_{(1)}\\ &= i/n \quad \text{if } z \in [Z_{(i)}, Z_{(i+1)}),\ i = 1, \ldots, n,\end{aligned}$$

that is, $\hat{F}$ is constant on intervals whose length are the spacings associated with the Z_i's.

In most applications the Z_i's are independent and then $\hat{F}$ is an estimate of F whose properties are well known. However, in some applications the Z_i's are dependent in the sense of Model A [then we denote them by $Y_1, \ldots, Y_n$] although marginally $Y_i \overset{st}{=} Z_i$. In that case $\hat{F}$ is not necessarily an unbiased estimator of F. By Theorem 1, the more positively dependent the Y_i's are, the shorter (stochastically) the corresponding spacings are and thus, the shorter (stochastically) the range of the support of $\hat{F}$ is. Geometrically, if **X** and **Y** satisfy (2.1) then the graph of the $\hat{F}$ based on the Y_i's will be (stochastically) steeper than the graph of the $\hat{F}$ based on the X_i's.

Thus, various statistics which are functions of $\hat{F}$ can be compared stochastically. This is the case if these statistics are nondecreasing functions of the underlying spacings. For example, various measures of dispersion (such as the range, the interquartile range, the sample variance, etc.) computed from $\hat{F}$ based on the X_i's are stochastically larger than the same computed from $\hat{F}$ based on the Y_i's.

The inverse of F,

$$Q(u) = \inf\{x : F(x) \geq u\}$$

and its density q (if it exists), are called, respectively, the *quantile* function and the *quantile-density* function. Parzen (1979) discusses various estimators $\hat{Q}$ and $\hat{q}$ of Q and q. Graphically, one of the estimators, $\hat{Q}$, is obtained by "inverting" $\hat{F}$ (that is, flipping the graph of $\hat{F}$ around the main diagonal of the bivariate plane). Parzen (1979) also suggests various "sensible" estimates of q which are obtained by by differentiating "smooth" versions of $\hat{Q}$. For example, one estimator of q is given by

$$\hat{q}(u) = n(Z_{(i)}-Z_{(i-1)}) \text{ for } u \in (\frac{i-1}{n}, \frac{i}{n}), i = 2, \ldots, n.$$

The comments about the influence of positive dependence on $\hat{Q}$ and $\hat{q}$ are similar to the ones made above about $\hat{F}$. Various monotone functionals of $\hat{Q}$ are discussed in Parzen (1979). Thus, one can stochastically compare various statistics based on a $\hat{Q}$ which was constructed from X_i's to similar statistics based on a $\hat{Q}$ which was constructed from Y_i's where **X** and **Y** satisfy (2.1).

4.4. Tests for outliers. For $i = 1, \ldots, n$, let Z_i be a normal random variable with mean μ_i and variance σ^2. Consider the null hypothesis $H_0: \mu_1 = \ldots = \mu_n$ and the following possible alternatives which state that one or k of the Z_i's are outliers:

A : $\mu_1 = \ldots = \mu_{i-1} = \mu_{i+1} = \ldots = \mu_n < \mu_i$ for some $i \in 1, \ldots, n$ (one of the Z_i's is an outlier caused by a slippage to the right),

A$'$: $\mu_1 = \ldots = \mu_{i-1} = \mu_{i+1} = \ldots = \mu_n > \mu_i$ for some $i \in 1, \ldots, n$ (one of the Z_i's slipped to the left),

A$''$: $\mu_1 = \ldots = \mu_{i-1} = \mu_{i+1} = \ldots = \mu_n \neq \mu_i$ for some $i \in 1, \ldots, n$ (one of the Z_i's is an outlier),

B_k:$n-k$ μ_i's are equal to an unknown μ and the other μ_i's are larger than μ

(there are k outliers caused by slippages to the right). Similarly B'_k and B''_k can be defined.

Various tests have been proposed for testing these and similar alternatives when the Z_i's are assumed to be independent (Barnett and Lewis (1978, pp. 89–115)). For example, if σ^2 is known then one can test A [respectively, B_k] by rejecting H_0 if

$$T_A \equiv \sigma^{-1}\phi_1(\mathbf{Z}) \equiv \sigma^{-1}(Z_{(n)} - \bar{Z}) > c \text{ for some } c$$

[respectively,

$$T_{B_k} \equiv \sigma^{-1}\phi_k(\mathbf{Z}) \equiv \sigma^{-1}(Z_{(n)} + \ldots + Z_{(n-k+1)} - k\bar{Z}) > c \text{ for some } c];$$

here $\bar{Z} = n^{-1}(Z_1 + \ldots + Z_n)$. Similarly, A' [respectively B'_k] can be tested by rejecting H_0 when

$$T_{A'} \equiv \sigma^{-1}\phi'_1(\mathbf{Z}) \equiv \sigma^{-1}(\bar{Z} - Z_{(1)}) > c \text{ for some } c$$

[respectively,

$$T_{B'_k} \equiv \sigma^{-1}\phi'_k(\mathbf{Z}) \equiv \sigma^{-1}(k\bar{Z} - Z_{(1)} - \ldots - Z_{(k)}) > c \text{ for some } c].$$

The two-sided alternative A'' may be tested by rejecting H_0 when

$$T_{A''} \equiv \sigma^{-1}\psi(\mathbf{Z}) \equiv \sigma^{-1}\max(Z_{(n)} - \bar{Z}, \bar{Z} - Z_{(1)}) > c \text{ for some } c$$

and B''_k may be tested by rejecting H_0 when

$$T_{B''_k} \equiv \sigma^{-2}\tilde{\psi}(\mathbf{Z}) \equiv \sigma^{-2}\Sigma_{i=1}^{n}(Z_i - \bar{Z})^2 > c \text{ for some } c$$

(see Dixon (1950, p. 490)) or by rejecting H_0 when

$$\tilde{T}_{B''_k} \equiv \sigma^{-1}\phi(\mathbf{X}) \equiv \sigma^{-1}(Z_{(n)}-Z_{(1)}) > c \text{ for some } c.$$

If σ^2 is unknown, but an independent estimate S_υ^2 of σ^2 is available, then one can test the various alternatives by replacing σ by S_υ in the above statistics (see details in Barnett and Lewis (1978, pp. 89–115)).

Note that all the above test statistics are nondecreasing functions of the spacings (for example $Z_{(n)}-\bar{Z} = n^{-1}\sum_{i=1}^{n-1}(Z_{(n)}-Z_{(i)}) = n^{-1}\sum_{i=1}^{n-1}\sum_{j=i+1}^{n-1}U_j$). It follows from Theorem 1 that if the observations are not independent but instead that a random shift and a rescaling in the sense Model A have been applied to the Z_i's [denote them then by Y_i's] leaving the marginals unchanged, then the significance level of each of the above tests may be smaller than the desired one.

Of course, the same analysis applies also to Z_i's which have distributions other than normal (see Section 3 for examples).

REFERENCES

BARNETT, V. and LEWIS, T. (1978). *Outliers in Statistical Data,* Wiley, New York.

DAVID, H. A. (1970). *Order Statistics.* Wiley, New York.

DIXON, W. J. (1950). Analysis of extreme values, *Ann. Math. Statist. 21* 488–506.

KARLIN, S. and TRUAX, D. R. (1960). Slippage problems. *Ann. Math. Statist. 31* 296–324.

KIRMANI, S. N. U. A. and ALAM, S. N. (1974). On goodness of fit tests based on spacings. *Sankhya Ser. A 36* 197–203.

PARZEN, E. (1979). Nonparametric statistical data modelling. *J. Amer. Statist. Assoc. 74* 105–121.

DEL PINO, G. E. (1979). On the symptotic distribution of k-spacings with applications to goodness-of-fit tests. *Ann. Statist. 7* 1058–1065.

PYKE, R. (1965). Spacings. *J. Roy. Statist. Soc. Ser. B 27* 395–449.

PYKE, R. (1972). Spacings revisited. *Sixth Berkeley Symp. on Math. Statist. and Prob., Vol. I* (eds. L. M. Le Cam, J. Neyman and E. L. Scott) University of California Press, Berkeley, 417–427.

RAO, J. S. and SETHURAMAN, J. (1970). Pitman efficiencies of tests based on spacings. *In Non-parametrics Techniques in Statistical Inference* (ed. M. L. Puri) Cambridge University Press, Cambridge, 267–273.

RAO, J. S. and SOBEL, M. (1980). Incomplete Dirichlet integrals with applications to ordered uniform spacings. *J. Multivariate Anal. 10* 603–610.

SHAKED, M. (1977). A concept of positive dependence for exchangeable random variables. *Ann. Statist. 5* 505–515.

SHAKED, M. and TONG, Y. L. (1985). Some partial orderings of exchangeable random variables by positive dependence. To appear in *J. Multivariate Anal.*

WEISS, L. (1957). Tests of fit in the presence of nuisance location and scale parameters. *Ann. Math. Statist. 28* 1016–1020.

WEISS, L. (1965). On asymptotic sampling theory for distributions approaching the uniform distribution. *Z. wahrscheinlichkeitstheorie und Gebiate 4* 217–221.

Inequalities in Statistics and Probability
IMS Lecture Notes–Monograph Series Vol. 5 (1984), 150–155

PROBABILISTIC ORDERING OF SCHEFFÉ POLYHEDRA

BY R. BOHRER and H. P. WYNN
University of Illinois and Imperial College

Some inequalities for spherically symmetric distributions are discussed using simple ideas from convex geometry. There are two dual orderings of the size of convex polytopes with respect to "width" in a random direction. One is equivalent to the ordering of content with respect to all spherically symmetric distributions. The other is the stochastic version of the mean width of convex geometry. Dual versions of known results are given and in particular the complete classification of the Platonic solids is listed. Some remarks are made about future developments.

1. Introduction. There is a duality between a problem in statistics of ordering certain regions with respect to probability content and an ordering based on the support function of a convex set. Recent results are discussed in the light of this connection.

The first ordering arises naturally in the statistical theory of multiple comparisons and by now has a considerable literature. Let $\mathcal{C}$ be a class of sets in p-dimensional Euclidean space $\mathcal{E}^p$. Let F be a family of probability measures on $\mathcal{E}^p$ with respect to which every member of $\mathcal{C}$ is measurable. We say that for two members C_1 and C_2 of $\mathcal{C}$, $C_1 > C_2$ if $\mu(C_1) > \mu(C_2)$ for all μ in F. It is usual to specialize $\mathcal{C}$ and F in various ways. Typically C may comprise all convex radially symmetric sets ($x \in C$ implies $-x \in C$) and F may be all unimodal, spherically symmetric distributions or their multivariate normal versions. Much of this material is summarised in Tong (1980).

In this paper we first restrict $\mathcal{C}$ to all closed star-shaped regions: $x \in C$ implies $\lambda x \in C$ for all $0 \leqslant \lambda \leqslant 1$. Thus C contains all points on the ray to the boundary point in any direction. Let F consist of all spherically symmetric distributions: all measures preserved under any rotation about the origin. We refer to the induced ordering as $>_h$. The following simple geometric characterisation comes as Theorem 1 in Bohrer and Wynn (1982). Let s be a random direction in $\mathcal{E}^p$ which may be interpeted as a point distributed with the uniform distribution on the surface of the unit sphere S_{p-1} in $\mathcal{E}^p$. Let $h(C,s)$ be the distance to the boundary of C from the origin in the direction s. Then the result is that, for C_1 and C_2 in $\mathcal{C}$, $C_1 >_h C_2$ if an only if $h(C_1,s) > h(C_2,s)$, where $>$ is stochastic ordering:

$$P[h(C_1,s) \geqslant r] > P[h(C_2,s) \geqslant r] \qquad \text{for all } 0 \leqslant r \leqslant \infty.$$

The proof follows directly from the fact that it is sufficient to prove that the p–1 dimensional area of the intersection with the spherical shell rS_{p-1} of radius r is at least as great for C_1 as for C_2, for all $0 \leqslant r \leqslant \infty$. Then since C_1 and C_2 are star-shaped these intersections are (proportional to) the s-probability of the boundary in the direction s lying outside or on rS_{p-1}.

Measures of the size of convex bodies abound in the field of convex geometry which has had a resurgence in recent years but has been little used in the field of multiple comparisons. The subject arises as a foundation for Minkowski's geometry of numbers and in particular for his theorem on the volume of n-dimensional lattices (see Stewart and Tall (1979) for an elementary treatment). One arm of the subject is loosely called integral geometry

AMS 1980 subject classifications. 62J15.
Key words and phrases: convexity, convex polytopes, multiple comparisons, stochastic geometry.

(see Santalo (1976)). More recently many advances have been made in stochastic geometry and mathematical morphology (Serra (1982), Coleman (1979)) and in describing the combinatorial and geometric properties of the face lattices of convex polyhedra (see Grunbaum (1967), McMullen and Shephard (1971)).

Our second ordering is based, then, on a classical idea from convex geometry. Let $\mathcal{C}$ now be the class of closed convex sets in $\mathcal{E}^p$ containing the origin. They are obviously star-shaped. The support function $H(C,s)$ of C is $\mathcal{C}$ in the direction s is defined to be

$$H(C,s) = \sup\{<y,s>|\, y \in C\}.$$

Note that since the origin is in C, $H(C,s) \geqslant 0$ for C in $\mathcal{C}$. With s random $H(C,s)$ becomes a random variable. Then we define $C_1 >_H C_2$ if and only if $H(C_1,s) > H(C_2,s)$ where again we mean stochastic ordering as defined above.

Both $>_h$ and $>_H$ are orderings of the size of the sets in $\mathcal{C}$. We now show that there is a close connection between them. Let C be a closed convex set containing the origin. The dual set of C which has the same properties is defined as follows

$$C^* = \{y|\, <x,y> \leqslant 1 \text{ for all } x \in C\}.$$

Note that $C^{**} = C$.

THEOREM 1. *Let C_1 and C_2 be closed convex sets containing the origin. Then $C_1 >_h C_2$ if and only if $C_2^* >_H C_1^*$.*

Proof. We can denote a general point $y \in \mathcal{E}^p$ by $y = rs$ where $r \geqslant 0$ and s is a point on the unit sphere. Then for a general closed convex set C containing the origin $h(C,s) = \sup\{r|y = rs \text{ in } C\}$. Then since $C^{**} = C$ this can be written

$$h(C,s) = \sup\{r|\, <x,rs> \leqslant 1 \text{ for all } x \text{ in } C^*\} = r^*, \text{ say},$$

while

$$H(C^*,s) = \sup\{<x,s>|\text{ for all } x \in C^*\}.$$

Clearly if this supremum is achieved at x^* in C then $r^* = <x^*,s>^{-1}$. Thus $H(C^*,s) = h(C,s)^{-1}$, with the value taken as ∞ when $h(C,s) = 0$. This immediately gives the inverse relationship between the orderings expressed in the theorem. □

There are a number of results using the support function as a measure of the size of a convex set C. Most important of these is that based on the so-called mean width $W(C)$:

$$W(C) = E\{H(C,s) + H(C,-s)\},$$

where E denotes expectation with respect to random s. The quantity $W(C)$ is invariant under change or origin and according to a result due originally to Crofton (see Hadwiger (1957) for the general case) W(C) is proportion to the surface area of C. The constant of proportionality depends only on how one defines the measure of area of the unit sphere. Now in our case C contains the origin so that $E\{H(C,s)\} = ½W(C)$. It is clear that $>_H$ implies the ordering of $E(H)$. Thus Theorem 1 combined with the Crofton result gives the following corollary: $C_1 >_h C_2$ implies that the surface area of C_2^* is at least that of C_1^*. We also have that the volume of C_1 is at least that of C_2. These necessary results can be used to obtain counterexamples to or conjectures about the ordering $>_h$. We shall return to this idea later. For a recent related paper on integral geometry see Enns and Ehlers (1980).

2. Scheffé Polyhedra. In the multiple comparison literature many interesting regions are obtained as one or two-sided confidence regions, or their translation to the origin. Thus in some testing or confidence procedures we may construct intervals based on estimates

$\hat{\theta} = (\hat{\theta}_1, \ldots, \hat{\theta}_p)$ for parameters $\theta = (\theta_1, \ldots, \theta_p)$ which take the form $<c_i, \hat{\theta} - \theta> \leq a_i$, $(i=1, \ldots, q)$ where the c_i are vectors of known coefficients. Writing $\mathbf{X} = \hat{\theta} - \theta$ we obtain $<c_i, \mathbf{X}> \leq a_i$, $(i = 1, \ldots, q)$. In many applications the vector $\mathbf{X}$ can be assumed to have a spherically symmetric distribution about the origin. Moreover the standardisation may make all the a_i equal. If this is the case then we term the $\mathbf{X}$-region so defined a Scheffé region (or polyhedron): the polyhedron consisting of the intersection of half spaces each of whose boundary hyperplane is tangent to the same sphere centred at the origin. Without loss of generality in what follows we shall take the sphere to be the unit sphere. Thus define for any collection of distinct points $S = \{s_1, \ldots, s_q\}$ in the unit sphere a Scheffé region to be

$$C(S) = \{<x, s_i> \leq 1, i = 1, \ldots, q\}.$$

For two-sided regions $s_i \in S$ implies $-s_i \in S$ $(i = 1, \ldots, q)$. In this case $C(S)$ is centrally symmetric.

Because we choose to use the unit sphere, the dual of a Scheffé region $C(S)$ based on $S = \{s_1, \ldots, s_q\}$ is just the convex hull of S, $C(S)^* = \text{conv}(S)$, and every s_i is an extreme point of $C(S)^*$. While $C(S)$ circumscribes the unit sphere $C(S)^*$ inscribes it.

From a statistical point of view the main interest is in giving conditions or examples for which $C(S_1) >_h C(S_2)$ holds. We can then make claims about the size of confidence level of the relevant procedure. Sometimes $C(S_1)$ will arise as a non-standard case whereas $C(S_2)$ may have a well known form and the μ content be tabulated. The claim then would be that the test based on S_1 may be conservative so that $\mu\{C(S_2)\}$ provides a lower bound to the confidence level or size. We now interpret some known results in the light of the geometric considerations of the last section.

Let $p = 2$ and order the s_i around the unit circle so that the angles subtended at the origin between adjacent s_i are given by

$$\cos^{-1}<s_i, s_{i+1}> = \theta_i, (i = 1, \ldots, q-1), \cos^{-1}<s_q, s_1> = \theta_q.$$

THEOREM 2. *Let $p = 2$ and $C(S_1)$ and $C(S_2)$ with angles (as above) $\theta = (\theta_1, \ldots, \theta_q)$ and $\phi = (\phi_1, \ldots, \phi_q)$. Then $C(S_1) >_h C(S_2)$ if and only if $\theta \succ \phi$ in the sense of majorization.*

Note that if S_2 has more points s_i than S_1 then we merely extend θ by adjoining the requisite number of zeros. Theorem 2 appears in Marshall and Olkin (1979, Chapter 8, Proposition E7) based on the earlier work of Wynn (1975); see also Bohrer and Wynn (1982). None of this work mentions the duality of the last section. The consequence of Theorem 1 is that $\theta \succ \phi$ is also equivalent to $C(S_2)^* >_H C(S_1)^*$ but the $C(S)^*$ are now the inscribed polygons. The Croften result then shows that the perimeter of S_2 is greater than that of S_1. The results of Marshall and Olkin in the same section (1979, Chapter 8, Proposition E1 to E6) concerning the area and perimeter of the inscribed figure are very close to this.

One of the best known results is that of Das Gupta, Eaton, Olkin, Perlman, Savage and Sobel (1972) and can be rewritten with a more geometric flavour. A p-pyramid is defined to be the convex hull of the union of a convex set K in $\mathcal{E}^{p-1}$ containing the origin and a line segment $[0, x]$. For a discussion of the general case see Grunbaum (1967). Three dimensional visualisation is useful with K being the base of the pyramid and $[0, x]$ the (not necessarily vertical) axis. Thus we define the pyramid as $P = \text{conv}(K \cup [0, x])$. A bipyramid is $P = \text{conv}(K \cup [-x, x])$. The results of Das Gupta, Eaton, Olkin, Perlman, Savage and Sobel (1972) can be restated as follows: For any spherically symmetric measure μ, the μ content of the dual P^* increases as x rotates towards the plane of K while lying in a fixed two dimensional "vertical" plane, when K is radially symmetric.

THEOREM 3. *Let K be a convex radially symmetric set in $\mathcal{E}^{p-1}$. Let a be a fixed unit vector in $\mathcal{E}^{p-1}$ and e_p the unit vector orthogonal to $\mathcal{E}^{p-1}$. Let $c > 0$ be a fixed constant and define a general vector $x = c\{(1-\lambda)^{1/2} e_p + \lambda^{1/2}a\}$. Let $P^*(\lambda)$ be the dual of the bipyramid $P(\lambda) = \mathrm{conv}(K \cup [-x, x])$. Then the ordering $>_h$ is increasing in λ, for $\lambda > 0$, in the sense that $\lambda_1 \geq \lambda_2 > 0$ implies $P^*(\lambda_1) >_h P^*(\lambda_2)$.*

The dual result that $P(\lambda)$ is decreasing in λ with respect to $>_H$ seems to be a new result in convex geometry although the implied decrease in surface area follows from an element argument. The case $p = 2$ follows by elementary geometry. A purely geometric proof in the case $p = 3$ has been given by the authors (Bohrer and Wynn (1983)). The special case when K has $2(p-1)$ vertices so that $P^*(\lambda)$ is p-dimensional parallelogram is the generalisation of a result due to Šidák (1968) for normal distributions.

The paper of Wynn and Bohrer (1982) classifies the platonic solids C_4 tetrahedron, C_6 cube, C_8 octahedron, C_{12} pentagonal dodecahedron and C_{20} icosahedron so that $C(S_i)$ $>_h C(S_j)$ when $i < j$ when all the figures are incident to the same (unit) sphere centred at the origin. The dual result is that $C_{12} >_H C_{20} >_H C_6 >_H C_8 >_H C_4$ when all the solids inscribe the same sphere. This is because (with lazy notation) $C_4 \equiv C_4{}^*$, $C_6 \equiv C_8{}^*$ and $C_{12} = C_{20}{}^*$. The implied ordering for volume and surface area must have been known from antiquity but the statement for the orderings $>_h$ and $>_H$ seem to be new. The work also studied the semi-regular rhomboidal dodecahedron C'_{12}. This arises out of Studentised range test with four means. It is a 12-sided solid each of whose faces is a rhombus with semi-axes in the ratio 1: $\sqrt{2}$. We showed that $C_8 >_h C'_{12} >_h C_{12}$. The dual result for the inscribed solids is that $C_{20} >_H C'^{\star}_{12} >_H C_6$. The solid $C'^{\star}_{12}$ is called the cuboctahedron and is obtain by suitably cutting off the corners of the unit cube (see Coxeter (1948) for a full description of all the solids).

4. Counter-Examples and Conjectures. As mentioned above $E\{H(C,s)\}$ is proportional to the surface area for radially symmetric convex sets. Clearly $E(h(C,s)^2)$ is proportional to the volume. In any case uniform μ giving volume content is spherically symmetric. A useful property of $H(C,s)$ is that it is additive with respect to direct sums. Thus

$$H(C_1 + C_2, s) = H(C_1, s) + H(C_2, s),$$

so that $E\{H(C_1 + C_2, s)\} = E\{H(C_1, s)\} + E\{H(C_2, s)\}$. It is not clear that $>_H$ is preserved under the direct sum operation but one can certainly use the result for expectations to eliminate any reversal of the ordering. That is to say if $C_1 >_H C'_1$ and $C_2 >_H C'_2$ then it is impossible for $(C'_1 + C'_2) >_H (C_1 + C_2)$ to hold strictly. Thus by Theorem 1 if $C'_1 >_h C_1$ and $C'_2 >_h C_2$ it is impossible for $(C_1 + C_2) >_h (C'_1 + C'_2)$ to hold strictly. This is a general indication that direct sums tend to preserve the direction of the ordering. The authors are engaged on a programme to search among regions generated under direct sums from the known results in the last section to establish a wide range of new examples. One interesting case is when C_1 and C'_1 are non centrally symmetric regions for which $C_1 >_h C_1$ and we put $C_1 = -C_2$ and $C'_1 = C'_2$. The direct sums then, statistically, are the regionsob tained from all pairwise contrasts among the defining linear functions of $C_1(C'_1)$. That is to say if $C_1 = C(S)$ where $S = s_1, \ldots, s_p$ then $C_1 + C_2 = C(S^-)$ where $S^- = s_i - s_j \mid i,j = 1, \ldots, q$.

The intuition for higher dimensions from Theorem 2 and the results on the Platonic solids is that in some general sense for a fixed number of $(p-1)$-dimensional faces the Scheffé region which is most regular is a minimal member of the ordering $>_h$. It appears that this is the case for the Platonic solids although the only one for which this is properly established is the cube with the added restriction of radial symmetry for which it follows from Theorem 3 above (Das Gupta, Olkin, Perlman, Savage and Sobel (1972)). The minimal member idea is closely related to minimal packing problems. Indeed it is well known that the p-simplex, p-cube and the p-dimensional generalisation of C'_{12} can be packed into $\mathcal{E}^p$. The use of orderings rather than volumes or other mean-size measures in packing theory may be new and will be the subject of a further paper. Another development which would be valuable would be a characterisation of these minimal regions in terms of their fundamental groups, that is the groups under which they remain invariant, where of course such a group exists. There must surely be a relationship between the structure of the finite subgroups of the full orthogonal group $0(p)$ and the $>_h$ ordering of their corresponding invariant Scheffé polyhedra. It was hoped to give some simple results in this paper but these too must wait for further developments.

Acknowledgements. We are grateful for conversations with Richard Vitale at the Nebraska Conference on Inequalities which partly led to the use of the support function, see his paper in this volume (Vitale (1983)). This is one reason why this paper differs in some degree from the one presented there. We are also grateful for discussion on stochastic geometry with Rodney Coleman.

REFERENCES

BOHRER, R. and WYNN, H. P. (1982). Spherically symmetric probability orderings useful in multiple comparisons. *Ann. Statist. 10* 1253–1260.

BOHRER, R. and WYNN, H. P. (1983). A geometric proof of the DEOPSS inequality for spherically symmetric distributions in 3 dimensions. In: *Festschrift for E. L. Lehmann,* Wadsworth, Belmont, CA.

COLEMAN, R. (1979). An introduction to mathematical stereology. Memoirs 3. Department of Theoretical Statistics, University of Aarhus.

COXETER, H. S. M. (1948). *Regular Polytopes.* Methuen, London.

DAS GUPTA, S., EATON, M., OLKIN, I., PERLMAN, M. D., SAVAGE, L. J., and SOBEL, M. (1972). Inequalities on the probability content of convex regions for elliptically contoured distributions. *Proc. Sixth Berkeley Symp. Math. Statist. Probab. 2* 241–256.

ENNS, E. G. and EHLERS, D. F. (1980). Random paths originating within a convex region and terminating on its surface. *Austal. J. Statist. 22* 60–68.

GRUNBAUM, B. (1967). *Convex Polytopes.* Wiley, New York.

HADWIGER, H. (1952). Uber zwei quadratische Distanzintegrale fur Eikorper. *Arch. Math.* (Basel) *3* 142–144.

MARSHALL, A. W. and OLKIN, I. (1979). *Inequalities: Theory of Majorization and Its Applications.* Academic Press, New York.

MCMULLEN, P. and SHEPHARD, G. C. (1971). *Convex Polytopes and the Upper Bound Conjecture.* London Math. Soc., Lecture Notes 3.

SANTALO, L. A. (1976). *Integral Geometry and Geometric Probability.* Addison-Wesley, Mass.

SERRA, J. (1982). *Image Analysis and Mathematical Morphology.* Academic Press, New York.

SIDÁK, Z. (1968). On multivariate normal probability of rectangles: their dependence on correlations. *Ann. Math. Statist. 39* 1425–1434.

STEWART, I. N. and TALL, D. O. (1979). *Algebraic Number Theory.* Chapman and Hall, London.

TONG, Y. L. (1980). *Probability Inequalities in Multivariate Distributions.* Academic Press, New York.

VITALE, R. A. (1984). Probability measures on the circle and the isoperimetric inequality. In *Inequalities in Statistics nad Probability,* Y. L. Tong, ed., Institutes of Mathematical Statistics, Haywood, CA, 109–111.

WYNN, H. P. (1977). An inequality for certain bivariate probability integrals. *Biometrika 64* 411–414.

Inequalities in Statistics and Probability
IMS Lecture Notes–Monograph Series Vol. 5 (1984), 156–164

REGIONS WHOSE PROBABILITIES INCREASE WITH THE CORRELATION COEFFICIENT AND SLEPIAN'S THEOREM

BY S. W. DHARMADHIKARI and KUMAR JOAG-DEV[1]
Southern Illinois University and University of Illinois

Let **X** have a multivariate normal distribution. Slepian (1962) proved that the upper and lower orthants ($\mathbf{x} \le \mathbf{c}$) and ($\mathbf{x} \ge \mathbf{c}$) have the property that their probabilities are nondecreasing in each ρ_{ij}. This easily implies, in the bivariate case, that if $A = Q_1 \cup Q_3 \cup B$, where Q_1 is an upper quadrant, Q_3 is a lower quadrant, B is a disjoint union of horizontal or vertical infinite strips and the interiors of Q_1, Q_3 and B are disjoint, then $P(A)$ is nondecreasing in ρ. This paper shows that, within a broad class of bivariate regions, sets A of the type described above are the only sets whose probabilities increase with the correlation coefficient when the means and the variances of X_1, X_2 take arbitrary values. Some results are also given for the cases where the means and the variances are restricted in some way.

1. Introduction. Let $\mathbf{X} = (X_1, \ldots, X_n)$ have the multivariate normal distribution with mean vector μ, variance vector σ^2 and correlation matrix (ρ_{ij}). Slepian (1962) proved that certain orthant probabilities are nondecreasing in each ρ_{ij} separately. This result and its generalizations have several applications; see, for example, Slepian (1962), Šidák (1968) and Joag-dev, Perlman and Pitt (1983). It is natural to ask whether there are sets other than orthants whose probabilities are nondecreasing in each ρ_{ij}. In this paper, we deal mainly wth the bivariate case and obtain a result which can be considered as a partial converse to Slepian's result.

Following the number of the quadrants in the plane, we denote by Q_1 an upper quadrant of the type $x_1 \geqslant a_1$, $x_2 \geqslant a_2$. A lower quadrant will be denoted by Q_3. The term infinite horizontal strip will mean a set defined by $-\infty < x_1 < \infty$, $a_2 \leqslant x_2 \leqslant b_2$. An infinite vertical strip is defined similarly. We note that the probability of an infinite horizontal or vertical strip is constant in ρ. Therefore, the following corollary of Slepian's result is immediate. For ease of reference, we state it as a theorem.

SLEPIAN'S THEOREM. *Let $A \subset \mathcal{R}^2$ have the form $A = Q_1 \cup Q_3 \cup B$, where B is a finite disjoint union or horizontal (or vertical) infinite strips and the interiors of Q_1, Q_3 and B are disjoint. Then $P(A)$ is nondecreasing in ρ.*

In section 2, we show that, within a broad class of bivariate regions, the sets A described in Slepian's theorem are the only sets whose probabilities are nondecreasing in ρ, when the means μ_1, μ_2 and the variances σ_1^2, σ_2^2 are allowed to take arbitrary values. Such a result can be considered to be a partial converse to Slepian's theorem. When the means and variances are restricted in some way, it is possible to obtain some additional regions whose probabilities increase with ρ. Some results in this direction are given in Section 3. In Section 4, we discuss an equivalent form of Slepian's result in terms of covariances and show that its generalization based on the concept of association fails.

---

[1] Research of the second author was supported by grant AFOSR–82–0007 from the Air Force Office of Scientific Research. The U.S. Government is authorized to reproduce and distribute reprints for Government purposes notwithstanding any copyright notation thereon.
AMS 1980 subject classification. 62E99, 62F99, 62H00.
Key words and phrases: Orthant probabilities, monotonicity in correlations.

2. A Partial Converse to Slepian's Theorem. In this section we first describe a broad class $\mathcal{D}$ of sets in $\mathcal{R}^2$ and then show that the only sets in $\mathcal{D}$ whose probabilities are nondecreasing in ρ for all values of μ_1, μ_2, σ_1 and σ_2 are the sets A described in Slepian's theorem.

Let $\mathcal{D}$ denote the class of all sets D with the following properties.

(1) D is a subset of $\mathcal{R}^2$ and coincides with the closure of its interior.

(2) The boundary of D consists of a finite number of line segments.

These two properties easily imply the following useful property.

(3) If $\mathbf{a}$ is a boundary point of D which is not a vertex, then there is an $\varepsilon > 0$ such that the intersection of the disc $|\mathbf{x} - \mathbf{A}| < \varepsilon$ with D is a convex set with a nonempty interior.

It is clear that $\mathcal{D}$ is a fairly broad class which includes all closed quadrants and their finite disjoint unions. It is also easy to see that every set considered in Slepian's theorem is in $\mathcal{D}$. We also note that the boundary of a set in $\mathcal{D}$ may contain line segments which are neither horizontal nor vertical.

Theorem 1 below concerns the class $\mathcal{D}$. We suspect that the theorem holds for a wider class of sets such as the class of sets with Jordan boundaries. We also believe that the heart of our proof will carry over to the more general case.

Suppose again that (X_1, X_2) has a bivariate normal distribution with parameters μ_1, μ_2, σ_1^2, σ_2^2 and ρ.

Definition 1. A set $A \subset \mathcal{R}^2$ is called an *S-region* if $P(A)$ is nondecreasing in ρ for all values of $\mu_1, \mu_2, \sigma_1^2, \sigma_2^2$.

We need a Lemma (see Figure 1).

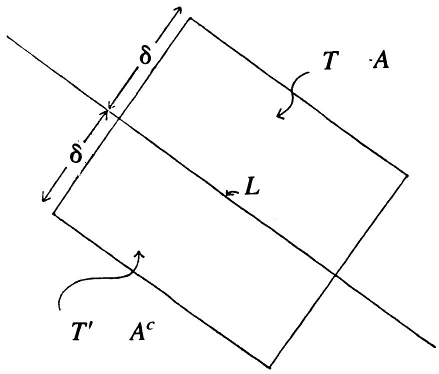


FIGURE 1. Illustration for the proof of Lemma 1.

LEMMA 1. *Let $A \subset \mathcal{R}^2$ be such that there is a line segment L in the boundary of A and a* $\delta > 0$ *such that* (a) *the open rectangle T with L as one side and height* δ *is contained in A;* (b) *the mirror image T'* of T about L is disjoint from A; and (c) *L is neither horizontal nor vertical. Then A is not an S-region.*

Proof. First assume that the slope of L is negative. By changing the origin and scales, if necessary, we may assume that

(i) L is contained in the line $x_1 + x_2 = 0$;

(ii) T is described by the conditions

$$0 < x_1 + x_2 < \delta \text{ and } |x_1 - x_2| < \eta.$$

where 2η is the length of L. The mirror image T' of T about L is then described by

$-\delta < x_1 + x_2 < 0$ and $|x_1 - x_2| < \eta$.

Write $U = X_1 + X_2$ and $V = X_1 - X_2$. Suppose that $\mu_1 = \mu_2 = \sigma_1^2 = \sigma_2^2 = \epsilon$, where $0 < 2\epsilon < \delta$. We use Z to denote a standard normal random variable.

Since $T \subset A$, we have $P(A) \geq P(T) = P[0 < U < \delta] \cdot P[|V| < \eta]$. But

$$P[0 < U < \delta] = P[-(\sqrt{2}\epsilon/\sqrt{1+\rho}) < Z < (\delta - 2\epsilon)/(2\epsilon(1+\rho))^{1/2}]$$
$$\to 1 \quad \text{as } \pi \to -1.$$

and

$$P[|V| < \eta] = P[|Z| < \eta/(2\epsilon(1-\rho))^{1/2}] \to P[|Z| < \eta/2\sqrt{\epsilon}] \quad \text{as } \rho \to 1.$$

Therefore, $\lim_{\epsilon\to 0} \liminf_{\rho\to(-1)} P(A) = 1$. On the other hand T' is disjoint from A. Therefore

$$1 - P(A) \geq P(T') = P[-\delta < U < 0] \cdot P[|V| < \eta].$$

Again

$$P[|V| < \eta] = P[|Z| < \eta/(2\epsilon(1-\rho))^{1/2}] \to 1 \quad \text{as } \rho \to 1.$$

and

$$P[-\delta < U < 0] = P[-(\delta+2\epsilon)/(2\epsilon(1+\rho))^{1/2} < Z < -(2\epsilon/(1+\rho))^{1/2}]$$
$$\to P[-(\delta+2\epsilon/2\sqrt{\epsilon} < Z < -\sqrt{\epsilon}] \quad \text{as } \rho \to 1.$$

Therefore,

$$\lim_{\varepsilon\to 0} \limsup_{\rho\to 1} P(A) \leqslant 1/2.$$

We thus see that, if ε is sufficiently close to zero, then

$$\liminf_{\rho\to(-1)} P(A) > \limsup_{\rho\to 1} P(A).$$

This shows that A is not an S-region. The case where the line segment L has positive slope can be handled similarly, the only change being that the mean vector is taken outside the rectangle T. The lemma is thus proved. □

We are now ready to prove a partial converse to Slepian's theorem. While the proof is somewhat long, it is elementary and is broken down into several simple steps.

THEOREM 1. *Let $D \in \mathcal{D}$ be an S-region. Then D is of the form $Q_1 \cup Q_3 \cup B$, where B is a finite disjoint union of horizontal (or vertical) infinite strips, the interiors of Q_1, Q_3 and B are disjoint and one or more of Q_1, Q_3, B may be empty.*

Proof. Recall that D satisfies the conditions (1), (2) and (3) stated at the beginning of this section.

Step 1. If L is a line segment in the boundary of D which is neither vertical nor horizontal, condition (3) shows that D would satisfy the conditions of Lemma 1 and could not be an S-region. Therefore, every line segment in the boundary of D is either horizontal or vertical.

Step 2. Suppose $\mathbf{a}$ is a vertex of D. Let $\varepsilon > 0$ and let N denote the disc $|\mathbf{x} - \mathbf{a}| < \varepsilon$. The horizontal and vertical lines through $\mathbf{a}$ divide N into four parts which we denote by N_1, N_2, N_3, N_4. We claim that, if ε is sufficiently small, then $D \cap N$ is the union of one or more of the N_i's.

In view of Step 1, we can use condition (2) to choose ε so that no boundary point of D is in the interior of any one of the N_i's. Now suppose, if possible, that there are points

b and **c** in the interior of N_1 such that $\mathbf{b} \in D^c$. Then the line segment [**b**,**c**] must contain a boundary point **d** of D and this point **d** must be in the interior of N_1. But this contradicts the choice of ϵ. Thus, the interior of N_1 must be either completely contained in D or completely outside D. The same argument holds for the other N_i's. Thus, the claim at the beginning of this step is verified.

Step 3. What we have proved so far tells us that, around a vertex, the set D is either a quadrant or the union of two or three quadrants. We can thus classify the vertices conveniently into ten types to be designated as NE, NW, SW, SE, $(\mathrm{NE})^c$, $(\mathrm{NW})^c$, $(\mathrm{SW})^c$, $(\mathrm{SE})^c$, NE $\cup$ SW, NW $\cup$ SE. We illustrate two of these types in Figure 2.

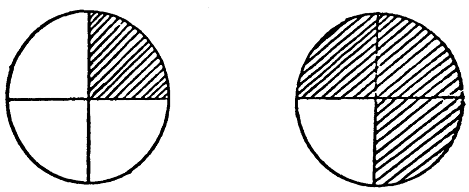

(a) NE vertex (b) $(\mathrm{SW})^c$ vertex

FIGURE 2. Illustration for the proof of Theorem 1 (Step 3).

Calculations similar to those in the proof of Lemma 1 show that the existence of a vertex of the type NW, SE, $(\mathrm{NE})^c$, $(\mathrm{SW})^c$ or NW $\cup$ SE would contradict the fact that D is an S-region. Therefore, a vertex of D must be one of the five types NE, SW, $(\mathrm{NW})^c$, $(\mathrm{SE})^c$ or NE $\cup$ SW. In what follows, we treat a NE $\cup$ SW vertex as both a NE vertex and a SW vertex.

Step 4. Any vertex of D, of one of the acceptable types in Step 3, is defined by two half lines starting at the vertex. We now show that no other vertex of D can be on any one of these defining half lines. Suppose, for instance, that $\mathbf{a} = (a_1, a_2)$ is a NE-type vertex. If there is a vertex on the half line $x_1 > a_1$, $x_2 = a_2$, then the closest such vertex (to **a**) must be either a NW vertex or a $(\mathrm{SW})^c$ vertex, which is impossible; see Figure 3. Thus there cannot be any vertex on the half line $x_1 > a_1$, $x_2 = a_2$. The same argument applies to vertices of the type SW, $(\mathrm{NW})^c$ and $(\mathrm{SE})^c$.

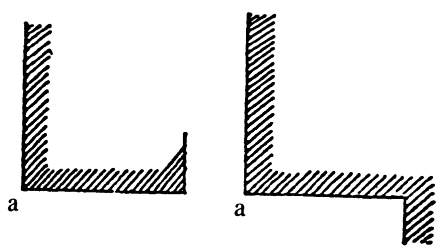


FIGURE 3. Illustration for the proof of Theorem 1 (Step 4).

Step 5. Let $\mathbf{a} = (a_1, a_2)$ be a vertex of the NE type. We show that the entire quadrant $x_1 \geq a_1$, $x_2 \geq a_2$ is contained in D. To see this, let (b_1, b_2) be a point of D^c in the open quadrant $x_1 > a_1, x_2 > a_2$. Since D^c is open, there is a neighborhood of **b** which is contained in D^c. Therefore (see Figure 4), we can start from a point **c** in such a neighborhood and proceed vertically downward to hit the set D at a point **d** which is not a vertex of D. Now, if we proceed horizontally to thge left from **d** we must hit a vertex **e**, which is a SE, $(\mathrm{NE})^c$ or SE $\cup$ NW vertex. Since all these types are impossible, we have reached a contradiction. Thus the entire quadrant determined by a NE type vertex is contained in D. The same result holds for a SW type vertex.

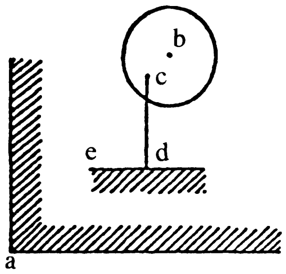


FIGURE 4. Illustration for the proof of Theorem 1 (Step 5).

Step 6. *If* $\mathbf{a} = (a_1, a_2)$ is a (NW)^c type vertex, then one can show by following elementary arguments as above that the entire open quadrant $x_1 < a_1$, $x_2 > a_2$ is outside D. A similar result holds for a (SE)^c type vertex.

Step 7. It follows easily from Steps 4, 5 and 6 that D can have at most one vertex of any given type. For instance, if there are two NE type vertices, then, by Step 5, it cannot happen that one of the quadrants is contained in the other. But then, we are bound to get a vertex on one of the defining half lines which is impossible by Step. 4.

Step 8. As our final step, we show that, if D has a (NW)^c vertex or a (SE)^{Cc} vertex, then D contains an infinite horizontal or vertical strip. We give the proof for a (NW)^c vertex. Let $\mathbf{a} = (a_1, a_2)$ be a (NW)^c vertex. The horizontal and vertical lines through $\mathbf{a}$ divide the plane into four quadrants, which we denote by A_1, A_2, A_3, A_4 in the usual order. By Step 6, we know that the interior of A_2 is completely outside D. If D coincides with $A_1 \cup A_3 \cup A_4$, then D clearly contains an infinite strip. So suppose that there is a point of D^c in the interior of A_1 for some $i = 1,3,4$. Three cases arise.

Case (i). Suppose we can find a point $\mathbf{b}$ of D^c in the interior of A_1; (see Figure 5). Since D^c is open, we may assume that, if we proceed leftward from $\mathbf{b}$, we would hit D at a boundary point $\mathbf{c}$ which is not a vertex. The boundary of D at $\mathbf{c}$ must be vertical. If we proceed downward from $\mathbf{c}$ and reach a vertex $\mathbf{d}$, then $\mathbf{d}$ must be either a NW vertex or a (NE)^c vertex. Since both these types are impossible, there is no vertex on the half line $x_1 = c_1$, $x_2 \leqslant c_2$. Now proceed upward from $\mathbf{c}$. If we do not reach a vertex at all, then D clearly contains a vertical strip. So suppose that we do reach a vertex $\mathbf{e}$. In view of Step 4, $\mathbf{e}$ must be of the (SE)^c type. We now claim that the half line $x_1 = c_1$, $x_2 > e_2$ is in the interior of $\mathbf{d}$. To see this, suppose that we proceed upward from $\mathbf{c}$ to reach a boundary point $\mathbf{f}$ of D. If $\mathbf{f}$ is a vertex, then $\mathbf{f}$ can be of either the (NW)^c type, which is impossible by Step 7, or the (NE)^c type, which is impossible by Step 3. Thus $\mathbf{f}$ is not a vertex. Further, the boundary of D at $\mathbf{f}$ is horizontal. Now, if we go leftward from $\mathbf{f}$, we must reach a vertex,

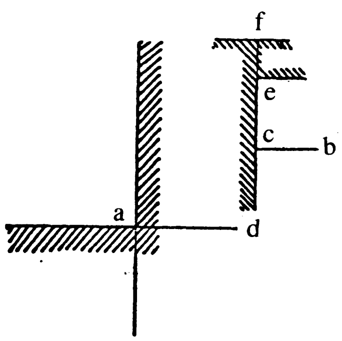


FIGURE 5. Illustration for the proof of Theorem 1 (Step 8).

which is either a SE vertex or a $(\mathrm{NE})^c$ vertex. This contradiction proves the above claim. We now see that, for some $\delta > 0$, the vertical strip $c_1 - \delta < x_1 \leqslant c_1$, $-\infty < x_2 < \infty$ must be contained in D.

Case (ii). If we can find a point $\mathbf{b}$ of D^c in the interior of A_3, then the discussion in Case (i) above easily shows that D contains an infinite horizontal strip.

Case (iii). If neither Case (i) nor Case (ii) arises, then $A_1 \cup A_3 \subset D$ and we can find a point $\mathbf{b}$ of D^c in the interior of A_4. Again, we may assume that, if we proceed leftward from $\mathbf{b}$, we would hit D at a boundary point $\mathbf{c}$ which is not a vertex. From this point on, the proof given in Case (i) applies word for word.

We have thus shown that the existence of a $(\mathrm{NW})^c$ vertex implies that D contains an infinite strip. The same conclusion clearly holds if D has a $(\mathrm{SE})^c$ vertex.

We are now ready to put everything together. If D contains an infinite horizontal (or vertical) strip, then we can remove the finite disjoint union B of all such strips from D to get a set E. Of course, if D does not contain an infinite strip, then $B = \phi$ and $E = D$. *In any case,* E is an S-region, $E \in \mathcal{D}$ and E does not contain an infinite strip. Now observe that:

(a) By Step 8, any vertex of E is a NE or a SW or a NE SW vertex, (b) by Step 7, E has at most two vertices, (c) if E has exactly one vertex, then E has the form Q_1, Q_3 or $Q_1 \cup Q_3$, (d) if E has two vertices, then one must be a NE vertex and the other a SW vertex, in this case, E has the form $Q_1 \cup Q_3$. The theorem is now completely proved. □

5. The Effect of Restrictions on Means or Variances. The results of Section 2 show that within the reasonably broad class of sets $\mathcal{D}$, the subclass of sets whose probabilities increase with ρ is rather narrow. However, it should be noted that the means and variances were completely unrestricted. One may therefore ask whether, under some restrictions on means and variances, one can identify additional sets whose probabilities increase with ρ. In this section we show that, at least in some cases, the answer is in the affirmative. We again assume that (X_1, X_2) has the bivariate normal distribution with parameters μ_1, μ_2, σ_1^2, σ_2^2 and ρ.

Example 1. Suppose that μ_1, μ_2 are fixed. Consider the half space H defined by the inequality $a_1x_1 + a_2x_2 \geqslant k$, where a_1, a_2 have the same sign. Suppose that (μ_1, μ_2) is outside H. That is, $a_1\mu_1 + a_2\mu_2 < k$. Then

$$P(H) = P[Z \geq (k - a_1\mu_1 - a_2\mu_2)/((a_1^2\sigma_1^2 + a_2^2\sigma_2^2 + 2a_1a_2\sigma_1\sigma_2\rho))^{1/2}$$

which is nondecreasing in ρ because $a_1a_2 \geq 0$. The same conclusion holds if $a_1a_2 < 0$ and (μ_1, μ_2) belongs to H. We note that H does belong to the class $\mathcal{D}$.

In the case where σ_1, σ_2 are fixed and μ_1, μ_2 are allowed to vary, we have been unable to find any regions (additional to those already found in Section 2) whose probabilities increase with ρ.

The rest of this section considers the case where $\mu_1, \mu_2, \sigma_1, \sigma_2$ are all fixed. Since a change of origin does not change the variances we assume that $\mu_1 = \mu_2 = 0$. Write $\alpha = (\sigma_1/\sigma_2)$. For the given value of α, we now construct a family of sets whose probabilities increase with ρ. Consider the following conditions on a set S in $\mathcal{R}^2$.

(i) If $(x_1,x_2) \in S$, then the entire line segment joining (x_1,x_2) to $((x_1 + \alpha x_2)/2, (x_1 + \alpha x_2)/2\alpha)$ as in S.

(ii) If $(x_1,x_2) \in S$, then $(x_1 + \alpha c, x_2 + c) \in S$ for every $c > 0$.

(iii) $(x_1,x_2) \in S \Rightarrow x_1 + \alpha x_2 \geq 0$.

Condition (i) is a convexity condition. Condition (ii) says that the variables "hang together" in S. The third condition is not natural but is related to the fact that $\mu_1 = \mu_2 = 0$. We saw in Example 1 that the position of the mean vector is important in determining whether a given set has its probability increasing in ρ.

THEOREM 2. *Suppose $S \subset \mathcal{R}^2$ is a set which satisfies the conditions* (i), (ii), (iii) *stated above. Let $\mu_1 = \mu_2 = 0$ and $(\sigma_1/\sigma_2) = \alpha$. Then $P(S)$ is nondecreasing in* ρ.

Proof. Let $U_1 = (X_1 + \alpha X_2)/(\sqrt{2}\ \sigma_1)$ and $U_2 = (X_1 - \alpha X_2)/(\sqrt{2}\ \sigma_1)$. Then U_1, U_2 are independent with zero means and variances $(1 + \rho)$ and $(1 - \rho)$ respectively. Under this transformation, the set S is converted into a set T such that

(A) If $(u_1,u_2) \in T$, then the entire line segment joining (u_1,u_2) to $(u_1,0)$ is in T.

(B) If $(u_1,u_2) \in T$, then $(u_1 + t, u_2) \in T$, for all $t > 0$.

(C) $(u_1,u_2) \in T \Rightarrow u_1 \geq 0$.

Let Q_ρ denote the distribution of $\mathbf{U} = (U_1,U_2)$. Then $P(S) = Q_\rho(T)$. Now, if D_ρ denotes the 2×2 diagonal matrix whose (1,1) entry is $(1 + \rho)^{-1/2}$ and (2,2) entry is $(1 - \rho)^{-1/2}$, then $Q_\rho(T) = Q_0(D_\rho T)$. But clearly $\rho_1 < \rho_2 \Rightarrow D_{\rho_1}(T) \subset D_{\rho_2}(T)$. Therefore $Q_{\rho_1}(T) \leqslant Q_{\rho_2}(T)$, whenever $\rho_1 \leqslant \rho_2$. The theorem is thus proved. □

The generalization of Theorem 2 to the k-variate equi-correlated case is straightforward and we state it in the theorem below without proof. Again we assume that the means are zero, $\sigma_i^2 = \mathrm{Var}(X_i)$ and $\rho = \mathrm{corr}(X_i,X_j)$, for all $i \neq j$. If $\mathbf{x} = (x_1, \ldots, x_n)$, we write $x^* = (1/n)\Sigma(x_i/\sigma_i)$. We also write $\sigma = (\sigma_1, \ldots, \sigma_n)$.

THEOREM 3. *Let X have the equi-correlated multivariate normal distribution with zero means. Suppose $S \subset \mathcal{R}^n$ be such that*

(1) $\mathbf{x} \in S \Rightarrow$ *the entire line segment joining* $\mathbf{x}$ *and* $(\sigma_1 x^*, \sigma_2 x^*, \ldots, \sigma_n x^*)$ *is in* S.

(2) $\mathbf{x} \in S \Rightarrow (\mathbf{x} + c\sigma) \in S$, *for all* $c \geq 0$.

(3) $\mathbf{x} \in S \Rightarrow \Sigma(x_i/\sigma_i) \geqslant 0$.

Then $P(S)$ is a nondecreasing function of ρ.

Example 2. One may ask whether the class of regions whose probabilities are nondecreasing in ρ is closed under intersections. The answer is in the negative even if attention is restricted to "increasing" sets. To see this, suppose that $\mu_1 = \mu_2 = 0$ and $\sigma_1 = \sigma_2 = 1$. Let $S_1 = \{(x_1,x_2): x_1 \geqslant 0\}$ and $S_2 = \{(x_1,x_2): x_2 \geqslant -(1 + \epsilon)x_1\}$, where $\epsilon > 0$. Then $P(S_1)$ and $P(S_2)$ are both nondecreasing in ρ because of Example 1. Now, if ϵ is close to zero, then $P(S_1 \cap S_2)$ is close to ½ when $\rho = -1$ and close to ⅜ when $\rho = 0$. Thus $P(S_1 \cap S_2)$ is not nondecreasing in ρ.

4. An Equivalent Form of Slepian's Result. According to Yanagimoto and Okamoto (1969), a random vector (X_1,X_2), whose distribution P_ρ depends on a parameter ρ, is said to have larger positive quadrant dependence under ρ_1 than under ρ_2 if

(4.1) $$P_{\rho_1}(X_1 \leqslant x_1, X_2 \leqslant x_2) \geqslant P_{\rho_2}(X_1 \leqslant x_1, X_2 \leqslant x_2) \text{ for all } (x_1,x_2).$$

They showed that (4.1) is equivalent to

$$\mathrm{Cov}[f_1(X_1), f_2(X_2); \rho_1] \geqslant \mathrm{Cov}[f_1(X_1), f_2(X_2); \rho_2] \tag{4.2}$$

for all nondecreasing functions f_1, f_2. Now Slepian's result shows that a bivariate normal vector (X_1, X_2) satisfies (4.1) and consequently it satisfies (4.2), whenever $\rho_1 > \rho_2$. Here the concept of "positive quadrant dependence" has been given an ordering relation which agrees, at least for the bivariate normal case, with the ordering based on the weaker concept of dependence, namely, the correlation coefficient.

Observe that the functions f_1, f_2 have separate arguments. We may ask whether $\mathrm{Cov}[h_1(X_1,X_2), h_2(X_1,X_2)]$ is nondecreasing in ρ if h_1, h_2 are nondecreasing in each argument and (X_1,X_2) is bivariate normal. This question is clearly related to the concept of association introduced by Esary, Proschan and Walkup (1967). That the answer is in the negative is indicated by Example 2. This example gives a set $S_1 \cap S_2 = B$, say, such that $P(B)$ is near $3/8$ when $\rho = 0$ and near $1/2$ when $\rho = -1$. If h denotes the indicator function of B, then h is increasing and $\mathrm{Var}[h(X_1,X_2)]$ is larger at $\rho = -1$ than at $\rho = 0$. Another example is as follows.

Example 3. Consider two quadrants

$$Q_0 = \{(x_1,x_2) : x_1 \geq 0, x_2 \geqslant 0\}$$

and

$$Q_t = \{(x_1,x_2) : x_1 \geqslant 0, x_2 \geqslant t\}, \text{ where } t > 0.$$

Again let (X_1,X_2) have a bivariate normal distribution with zero means, unit variances and correlation coefficient ρ. Denote the density function of (X_1,X_2) by g. Then

$$P(Q_t) = \int_0^\infty \int_t^\infty g(x_1,x_2;\rho)\,dx_2\,dx_1 \tag{4.3}$$

Using (4.3) and the fact that

$$(\partial/\partial\rho)g = (\partial^2/\partial x_1, \partial x_2)g,$$

we get

$$\begin{aligned} &d/d\rho[P(Q_0 \cap Q_1) - P(Q_0)P(Q_t)] \\ &= g(0,t;\rho) - P(Q_0)g(0,t;\rho) - P(Q_0)g(0,0;\rho). \end{aligned} \tag{4.4}$$

When ρ is near 1, $g(0,0;\rho)$ is large, $g(0,t;\rho)$ is small and $P(Q_1)$ is bounded away from zero. Therefore the above derivative (4.4) is negative for ρ near 1. Equivalently, the indicators of Q_0 and Q_1 are nondecreasing functions whose covariance is decreasing in ρ near $\rho = 1$.

Acknowledgment. Thanks are due to the referee who made useful comments on the earlier draft of the paper, pointed out the need for the results of section 2 and provided a simpler proof of Theorem 2.

REFERENCES

ESARY, J. D. PROSCHAN, F. and WALKUP, D. W. (1967). Associated random variables with applications. *Ann. Math. Statist. 38* 1466–1474.

JOAG-DEV, K., PERLMAN, M. and PITT, L. (1983) Association of normal random variables and Slepian's inequality. *Ann. Probab. 11* 451–455.

YANAGIMOTO, T. and OKAMOTO, M. (1969). Partial orderings of permutations and monotonicity of rank correlation statistic. *Ann. Inst. Statist. Math. 21* 489–506.

SIDAK, Z. (1968). On multivariate normal probabilities of rectangles: Their dependence on correlations. *Ann. Math. Statist. 39* 1425–1434.

SLEPIAN, D. (1962). The one-sided barrier problem for Gaussian noise. *Bell System Tech. J. 41* 463–501.

Inequalities in Statistics and Probability
IMS Lecture Notes–Monograph Series Vol. 5 (1984), 165–172

MOMENT INEQUALITIES WITH APPLICATIONS TO REGRESSION AND TIME SERIES MODELS

BY TZE LEUNG LAI[1] and CHING ZONG WEI[2]
Columbia University and the University of Maryland

Herein we review several important moment inequalities in the literature and discuss their applications to strong (almost sure) limit theorems for linear processes and for least squares estimates in multiple regression models.

1. Introduction and Summary. A classical model for random noise in the regression and time series literature is that of *equinormed orthogonal* random variables ϵ_n, i.e.,

$$E(\epsilon_i \epsilon_j) = 0 \quad \text{for } i \neq j, \qquad = \sigma^2 \quad \text{for } i = j. \tag{1.1}$$

Such random variables have the important mean square property that for all constants c_i,

$$E(\Sigma_{i=m}^n c_i\epsilon_i)^2 = \sigma^2 \Sigma_{i=m}^n c_i^2 \qquad \text{for all } n \geq m. \tag{1.2}$$

For example, the so-called Gauss-Markov model in multiple regression theory is of the form

$$z_i = \beta_1 t_{i1} + \ldots + \beta_k t_{ik} + \epsilon_i \,(i = 1, 2, \ldots) \tag{1.3}$$

where t_{ij} are known constants, z_i are observed random variables, $\beta_i, \ldots, \beta_k$ are unknown parameters, and ϵ_i are equinormed orthogonal random variables that represent unobservable random errors. Throughout the sequel we shall let $\mathbf{T}_n$ denote the design matrix $(t_{ij})_{1\leq i\leq n,\ 1\leq j\leq k}$, and let $\mathbf{Z}_n = (z_1, \ldots, z_n)'$. For $n \geq k$, the least squares estimate $\mathbf{b}_n = (b_{n1}, \ldots, b_{nk})'$ of $\beta = (\beta_1, \ldots, \beta_k)'$ based on the design matrix $\mathbf{T}_n$ and the response vector $\mathbf{Z}_n$ is given by

$$\mathbf{b}_n = (\mathbf{T}'_n\mathbf{T}_n)^{-1}\mathbf{T}'_n\mathbf{Z}_n, \tag{1.4}$$

provided that $\mathbf{T}'_n\mathbf{T}_n$ is nonsingular. From (1.1), it follows easily that

$$\operatorname{cov}(\mathbf{b}_n) = \sigma^2(\mathbf{T}'_n\mathbf{T}_n)^{-1}, \tag{1.5}$$

and therefore $\mathbf{b}_n$ is weakly consistent (i.e., $\mathbf{b}_n \xrightarrow{P} \mathbf{B}$) if

$$(\mathbf{T}'_n\mathbf{T}_n)^{-1} \to \mathbf{0} \text{ as } n \to \infty. \tag{1.6}$$

If $\sigma \neq 0$, the condition (1.6) is also necessary for the weak consistency of $\mathbf{b}_n$ (cf. Drygas (1976)).

In time series theory, it is well known that every wide-sense stationary sequence $\{y_n\}$ with zero means and an absolutely continuous spectral distribution can be represented as

$$y_n = \text{l.i.m.}_{N\to\infty}\Sigma_{i=-N}^{N} a_{n-i}\epsilon_i, \tag{1.7}$$

where $\{\epsilon_n\}$ is an orthonormal sequence (i.e., $\sigma = 1$ in (1.1)), $\{a_n\}$ is a sequence of constants such that $\Sigma_{n=-\infty}^{\infty} a_n^2 < \infty$, and l.i.m. denotes limit in quadratic mean (cf. Doob (1953), page 499). From this representation, it follows that

---

[1,2] Research supported by the National Science Foundation.

AMS 1970 subject classifications: Primary 60F15; Secondary 60G35, 62J05, 62M10.

Key words and phrases. Least squares theory, linear processes, orthogonal random variables, moment inequalities, lacunary systems, almost sure convergence, law of the iterated logarithm.

$$E(\Sigma_{j=1}^{n} y_j)^2 = \Sigma_{i=-\infty}^{\infty}(\Sigma_{j=1}^{n} a_{j-i})^2 = o(n^2),$$

and therefore $\{y_n\}$ satisfies the weak law of large numbers:

$$n^{-1}\Sigma_{j=1}^{n} y_j \xrightarrow{P} 0. \tag{1.8}$$

The representation (1.7) also provides an important stochastic model in the engineering literature, where the sequence $\{\epsilon_n\}$ is a white noise sequence and $\{y_n\}$ is the output sequence obtained by passing $\{\epsilon_n\}$ through a linear filter defined by $\{a_n\}$ (cf. Kailath (1974)). We shall call the sequence $\{y_n\}$ in (1.7) a *linear process generated by* $\{\epsilon_n\}$.

In order to strengthen the weak consistency result on $\mathbf{b}_n$ into its strong consistency under the minimal assumption (1.6) on the design constants, or to strengthen the weak law (1.8) into the corresponding strong law, we have found it necessary to introduce additional structure into the noise sequence $\{\epsilon_n\}$. Indeed, Chen, Lai and Wei (1981) gave a counter-example to show that the condition (1.6) is not sufficient for the strong consistency of $\mathbf{b}_n$ in the Gauss-Markov model. A very useful additional assumption on $\{\epsilon_n\}$, which is satisfied by many important classes of random variables that are natural models for random noise, and which yields the desired strong limit theorems, takes the form of the following moment inequality, to be satisfied for some $p>2$ and all constants c_i:

$$E|\Sigma_{i=m}^{n} c_i \epsilon_i|^p \leqslant K_p(\Sigma_{i=m}^{n} c_i^2)^{p/2} \qquad \text{for all } n \geqslant m. \tag{1.9}$$

Given $p>0$, a sequence of random variables $\{\epsilon_n\}$ is called a *lacunary system of order* p, or an S_p *system*, if there exists a positive constant K_p such that the moment inequality (1.9) is satisfied for all constants c_i. The concept of S_p systems was introduced by Banach (1930) and Sidon (1934). If $\{\epsilon_n\}$ is an S_p system for all $p>0$, then it is called an S_∞ system. In view of (1.2), an equinormed orthogonal system is an S_2 system, and the moment inequality (1.9) can be regarded as an L_p extension of the L_2 property (1.2). In Section 2, we give some basic properties and examples of S_p systems, and in this connection, review some important moment inequalities in the literature. In particular, we also discuss how the moment inequality (1.9) in the case $p>2$ is related to the almost sure limiting behavior of the sequence $\{\Sigma_{i=1}^{n} c_i \epsilon_i\}$.

While the moment restriction (1.9) appears more restrictive than the equinormed orthogonal situation (1.2) in the sense that it considers the p^{th} absolute moment with $p>2$, it is also less restrictive than (1.2) in the sense that it replaces equality in (1.2) by an upper inequality ($\leqslant$). If we replace equality in (1.2) by a lower inequality ($\geqslant$), then we get a Bessel-type inequality. A sequence of random variables $\{\epsilon_n\}$ is said to satisfy the *Bessel inequality* if there exists $K>0$ such that for all constants c_i,

$$E(\Sigma_{i=m}^{n} c_i \epsilon_i)^2 \geqslant K \Sigma_{i=m}^{n} c_i^2 \qquad \text{for all } n \geqslant m \tag{1.10}$$

(cf. Gaposhkin (1966)). Since $E|Y|^p \geqslant (EY^2)^{p/2}$ for $p>2$, the inequality (1.10) in turn implies the existence of a positive constant $A_p>0$ such that for all constants c_i,

$$E|\Sigma_{i=m}^{n} c_i \epsilon_i|^p \geqslant A_p(\Sigma_{i=m}^{n} c_i^2)^{p/2} \qquad \text{for all } n \geqslant m. \tag{1.11}$$

Gaposhkin (1966) showed that if $\{\epsilon_n\}$ is an S_p system for some $p>2$ and if it also satisfies the Bessel inequality, then it is a *Banach system,* i.e., there exists $A(=A_1)$ such that (1.11) holds with $p=1$ for all constants c_i. Clearly, if $\{\epsilon_n\}$ is a Banach system, then for every $p \geqslant 1$, there exists $A_p>0$ such that (1.11) holds for all constants c_i.

Replacing equality in (1.2) by the upper inequality (1.9) (with $p=2$) and the lower inequality (1.10) enables us to substantially enlarge the equinormed orthogonal model for random noise and include random errors that are correlated and have different variances. Assuming (1.9) for some $p>2$ in addition often enables us to extend the mean square convergence properties in classical regression and time series models with equinormed orthogonal errors to the corresponding almost sure convergence properties. For example, as shown by Lai and Wei (1983), if the random errors ϵ_n in the linear process (1.7) form an orthonormal S_p system with $p>2$, then the weak law (1.8) can indeed be strengthened into the strong law, i.e.,

$$n^{-1}\Sigma_{j=1}^{n} y_j \to 0 \quad \text{a.s.}$$

To establish the strong consistency of the least squares estimate $\mathbf{b}_n = (b_{n1}, \ldots, b_{nk})'$ in the multiple regression model (1.3) when the random errors ϵ_i form an S_p system with $p>2$, we fix $j=1, \ldots, k$ and note that b_{nj} can be represented for all large n as

$$b_{nj} - \beta_j = (\Sigma_{i=1}^{n} a_{ni}\epsilon_i)/(\Sigma_{i=1}^{n} a_{ni}^2), \tag{1.12}$$

where $\{a_{ni} : 1 \leq i \leq n, n=1, 2, \ldots\}$ is a triangular array of constants such that

$$\Sigma_{i=1}^{m} a_{ni}a_{mi} = \Sigma_{i=1}^{m} a_{mi}^2 \qquad \text{for } n \geq m \tag{1.13}$$

(cf. Lai and Wei (1982), Lemma 2). Thus, to study the limiting behavior of the least squares estimate b_{nj}, it is useful to consider more generally linear transformations of the form

$$x_n = \Sigma_{i=-\infty}^{\infty} a_{ni}\epsilon_i \tag{1.14}$$

where a_{ni} are constants such that $\Sigma_{i=-\infty}^{\infty} a_{ni}^2 < \infty$ for every n. Since $\{\epsilon_n\}$ is an S_p system with $p>2$, the series in (1.14) indeed converges a.s. (see Section 2). Partial sums $x_n = \Sigma_{i=1}^{n} y_i$ of the linear process $\{y_i\}$ defined in (1.7) can also be expressed in the form (1.14). In Section 3, we consider the almost sure limiting behavior of such linear transformations of S_p systems and discuss applications of the results to regression and time series models.

2. Lacunary systems, Banach systems, and related moment inequalities. We now give some examples of S_p systems and Banach systems, and in this connection, also review some important moment inequalities in the literature.

Example 1. If $\{\epsilon_n\}$ are i.i.d. standard normal random variables, then since $E|\Sigma_{i=m}^{n} c_i\epsilon_i|^p = (\Sigma_{i=m}^{n} c_i^2)^{p/2} E|N(0,1)|^p$, $\{\epsilon_n\}$ is an S_∞ system and a Banach system.

Example 2. Let $\{\epsilon_n\}$ be i.i.d. Bernoulli random variables such that $P\{\epsilon_n = 1\} = P\{\epsilon_n = -1\} = 1/2$. Then by an inequality of Khintchine (1924), for every $p>0$, there exist positive constants A_p and B_p such that

$$A_p(\Sigma_{i=m}^{n} c_i^2)^{p/2} \leq E|\Sigma_{i=m}^{n} c_i\epsilon_i|^p \leq B_p(\Sigma_{i=m}^{n} c_i^2)^{p/2} \tag{2.1}$$

for all $n \geq m$ and all constants c_i. Thus, Khintchine's inequality implies that $\{\epsilon_n\}$ is an S_∞ system and a Banach system.

Khintchine's inequality was generalized to general independent random variables by Marcinkiewicz and Zygmund (1937) who showed that if ϵ_n are independent random variables with zero means, then for every $p \geq 1$, there exist positive constants A_p and B_p depending only on p such that

$$A_p E\{(\Sigma_{i=m}^{n} \epsilon_i^2)^{p/2}\} \leq E|\Sigma_{i=m}^{n} \epsilon_i|^p \leq B_p E\{(\Sigma_{i=m}^{n} \epsilon_i^2)^{p/2}\} \tag{2.2}$$

for all $n \geq m$. In the case $p>1$, the moment inequality (2.2) was extended from independent random variables to martingale difference sequences $\{\epsilon_n\}$ by Burkholder (1966). Some other important martingale extensions of the Marcinkiewicz-Zygmund inequality can be found in Burkholder's survey paper (1973) and the references therein.

Making use of Burkholder's inequality (2.2) for martingale difference sequences and Minkowski's inequality, Lai and Wei (1983) obtained

Example 3. Let $p \geq 2$, and let $\{\epsilon_n\}$ be a martingale difference sequence (i.e., $E\{\epsilon_n | \epsilon_j, j \leq n-1\} = 0$ for all n) such that $\sup_n E|\epsilon_n|^p < \infty$. Then $\{\epsilon_n\}$ is an S_p system. If furthermore $\inf_n E|\epsilon_n| > 0$, then it follows from Lemma 4 of Burkholder (1968) that $\{\epsilon_n\}$ is also a Banach system.

Let r be a positive even integer. A sequence of random variables $\{\epsilon_n\}$ is said to be *multiplicative of order r* if

$$E(\epsilon_{i_1} \dots \epsilon_{i_r}) = 0 \text{ for all } i_1 < i_2 < \dots < i_r. \tag{2.3}$$

When $r=2$, this reduces to orthogonal random variables and therefore forms an S_2 system if $\sup_i E\epsilon_i^2 < \infty$. For $r \geq 4$, Komlós (1972) obtained the following

Example 4. Let $r \geq 4$ be an even integer, and let $\{\epsilon_n\}$ be a multiplicative sequence of order r such that $\sup_i E\epsilon_i^r < \infty$. Then as shown by Komlós (1972), $\{\epsilon_n\}$ is an S_r system. Obviously, if $\inf_n E\epsilon_n^2 > 0$, then $\{\epsilon_n\}$ satisfies the Bessel inequality, and this in turn implies that $\{\epsilon_n\}$ is a Banach system since it is an S_r system ($r>2$) satisfying the Bessel inequality. Longnecker and Serfling (1978) introduced three different ways to weaken the multiplicative condition (2.3) and showed that these three different classes of weakly multiplicative systems of order r are also S_r systems if $\sup_i E\epsilon_i^r < \infty$. They also showed that certain stationary mixing sequences and Gaussian sequences are special cases of these weakly multiplicative sequences.

The following maximal inequality plays an important role in the theory of S_p systems with $p>2$.

LEMMA 1. (Móricz (1976)). *Let $p>0$ and $\alpha>1$. Let $\{x_n\}$ be a sequence of random variables. Suppose that there exist nonnegative constants d_i such that*

$$E|x_n - x_m|^p \leq \left(\Sigma_{i=m+1}^n d_i\right)^\alpha \qquad \textit{for } n>m\geq m_0. \tag{2.4}$$

Then there exists an absolute constant $C_{p,\alpha}$ such that

$$E\left(\max_{m\leq i\leq n} |x_i - x_m|^p\right) \leq C_{p,\alpha}\left(\Sigma_{i=m+1}^n d_i\right)^\alpha \qquad \textit{for } n>m\geq m_0. \tag{2.5}$$

As a consequence of the maximal inequality (2.5), we obtain the following corollary on the almost sure convergence of $\{x_n\}$ and also its order of magnitude in case of divergence (cf. Lai and Wei (1983), Lemma 3.2).

COROLLARY 1. *With the same notation and assumptions as in Lemma 1, define*

$$D_n = \Sigma_{i=m_0+1}^n d_i. \tag{2.6}$$

(i) *If $\lim_{n\to\infty} D_n < \infty$, then x_n converges a.s. and in the L_p-norm.*

(ii) *If $\lim_{n\to\infty} D_n = \infty$, then for every $\delta > 0$,*

$$x_n = o(\{D_n^{\alpha/p}(\log D_n)^{1/p}(\log\log D_n)^{(1+\delta)/p}\}) \quad a.s. \tag{2.7}$$

Remark. Suppose that $\{\epsilon_n\}$ is an S_p system for some $p>2$ and $\{c_n\}$ is a sequence of constants. Let $x_n = \Sigma_{i=1}^n c_i\epsilon_i$. Then (1.9) implies that $\{x_n\}$ satisfies (2.4) with $\alpha = p/2 > 1$, $d_i = K_p^{2/p}c_i^2$ and $m_0 = 1$. Therefore Lemma 1 and Corollary 1 are applicable to $\{x_n\}$.

The special case $p=\alpha=4$ in Lemma 1 was first established by Erdös (1943) for lacunary trigonometric series. The result of Erdös was subsequently extended by several authors (cf. Móricz (1976) and the references therein), and Móricz (1976) considered in addition to the case $\alpha>1$ in Lemma 1 also the cases $\alpha=1$ and $0<\alpha<1$. The latter two cases are quite different from the case $\alpha>1$; instead of the absolute constant $C_{p,\alpha}$ in (2.5), the corresponding maximal inequalities in these two cases involve constants of the form $C_{p,\alpha}(m,n)$. These results generalize the classical Rademacher-Mensov inequality for orthogonal random variables (cf. Doob (1953), page 156): If $\epsilon_1, \ldots, \epsilon_n$ are orthogonal random variables with finite variances $\sigma_1^2, \ldots, \sigma_n^2$, then

$$E\{\max_{1\leq j\leq n} (\Sigma_{i=1}^j \epsilon_i)^2\} \leq \left(\frac{\log 4n}{\log 2}\right)^2 \Sigma_{i=1}^n \sigma_i^2. \tag{2.8}$$

The following recent generalization of this kind of maximal inequalities is due to Móricz, Serfling and Stout (1982).

LEMMA 2. *Let* $g:\{1, \ldots, n\} \times \{1, \ldots, n\} \to [0,\infty)$ *such that for some* $Q \geq 1$

$$g(i,j) + g(j+1,k) \leq Qg(i,k) \quad \textit{for } i \leq j < k, \tag{2.9}$$

$$g(i,j) \leq g(i,j+1) \quad \textit{for } i \leq j. \tag{2.10}$$

Let $\epsilon_1, \ldots, \epsilon_n$ *be random variables such that for some* $p \geq 1$ *and* $\alpha \geq 1$,

$$E|\Sigma_{t=1}^j \epsilon_t|^p \leq g^\alpha(i,j) \quad \textit{for all } 1\leq i\leq j\leq n. \tag{2.11}$$

(i) *If* $\alpha > 1$ *and* $Q < 2^{(\alpha-1)/\alpha}$, *then there exists an absolute constant* $C_{p,\alpha,Q}$ *such that*

$$E(\max_{1\leq j\leq n} |\Sigma_{t=1}^j \epsilon_t|^p) \leq C_{p,\alpha,Q}\, g^\alpha(1,n). \tag{2.12}$$

(ii) *In the case* $\alpha = 1$, *we have the maximal inequality*

$$E(\max_{1\leq j\leq n} |\Sigma_{t=1}^j \epsilon_t|^p) \leq (\Sigma_{t=0}^{[\log n/\log 2]} Q^{t/p})^p\, g(1,n). \tag{2.13}$$

For the case $Q=1$, inequality (2.9) says that g is superadditive and implies (2.10). In this case, as pointed out by Longnecker and Serfling (1977), there exist nonnegative constants $d_1, \ldots, d_n$ such that

$$g(1,n) = \Sigma_{t=1}^n u_t \quad \text{and} \quad g(i,j) \leq \Sigma_{t=i}^j u_t \quad \text{for} \quad 1\leq i\leq j\leq n, \tag{2.14}$$

and therefore the maximal inequality (2.12) reduces to that of Lemma 1.

For the case where $g(i,j)$ takes the form $g(i,j) = g(j-i+1)$, (2.9) becomes

$$g(i) + g(j-i) \leq Q\, g(j) \quad \text{for } i\leq j. \tag{2.15}$$

Another maximal inequality of this nature but under an assumption different from (2.15) is

LEMMA 3. (Lai and Stout (1980)). *Let* $g:\{1, 2, \ldots\} \to (0, \infty)$ *be a function satisfying*

$$\liminf_{n\to\infty} g(Kn)/g(n) > K \quad \textit{for some integer } K\geq 2, \tag{2.16}$$

and

$$\textit{for all } \delta>0, \textit{ there exists } \rho=\rho(\delta)<1 \textit{ for which } \limsup_{n\to\infty}\{\max_{\rho n\leq i\leq n} g(i)/g(n)\} < 1+d. \tag{2.17}$$

Let $\epsilon_1, \epsilon_2, \ldots$ *be random variables such that for some* $p>0$,

$$E|\Sigma_{t=\nu+1}^{\nu+n} \epsilon_t|^p \leq g(n) \quad \textit{for all } \nu \geq 0 \textit{ and } n \geq 1. \tag{2.18}$$

Then there exists a positive constant C such that

$$E(\max_{1\leq j\leq n} |\Sigma_{t=\nu+1}^{\nu+j} \epsilon_t|^p) \leq C\, g(n) \quad \textit{for all } \nu \geq 0 \textit{ and } n \geq 1. \tag{2.19}$$

3. Linear transformations of S_p systems and their applications. In this section we first consider the multiple regression model (1.3) where the random errors ϵ_i form an S_p system with $p>2$, and apply Corollary 1 to establish the strong consistency of the least squares estimate $\mathbf{b}_n = (b_{n1}, \ldots, b_{nk})'$ under the assumption (1.6) on the design constants. This is the content of

COROLLARY 2. *Suppose that in the multiple regression model* (1.3) *the random variables* ϵ_i *form an* S_p *system for some* $p>2$. *Let* $\mathbf{V}_n = (v_{ij}^{(n)})_{1\le i,j\le k} = (\mathbf{T}'_n\mathbf{T}_n)^{-1}$. *Fix* $j=1, \ldots, k$. *If* $\lim_{n\to\infty} v_{jj}^{(n)} = 0$, *then for every* $\delta>1/p$,

$$b_{nj}-\beta_j = o(\{(v_{jj}^{(n)})^{1/2}|\log v_{jj}^{(n)}|^{1/p}(\log|\log v_{jj}^{(n)}|)^{\delta}\}) \quad \text{a.s.} \tag{3.1}$$

Proof. By (1.12), $b_{nj}-\beta_j = (\Sigma_{i=1}^n a_{ni}\,\epsilon_i)/(\Sigma_{i=1}^n a_{ni}^2)$ for all large n, where a_{ni} are constants satisfying (1.13). Let $x_n = \Sigma_{i=1}^n a_{ni}\epsilon_i$. *Since* $\{\epsilon_n\}$ is an S_p system, for $n>m$,

$$\begin{aligned} E|x_n-x_m|^p &= E|\Sigma_{i=1}^m(a_{ni}-a_{mi})\,\epsilon_i + \Sigma_{i=m+1}^n a_{ni}\epsilon_i|^p \\ &\le K_p\{\Sigma_{i=1}^m(a_{ni}-a_{mi})^2 + \Sigma_{i=m+1}^m a_{ni}^2\}^{p/2}. \end{aligned} \tag{3.2}$$

Let $D_n = \Sigma_{i=1}^n a_{ni}^2$, $d_n = D_n-D_{n-1}(D_0=0)$. It follows from (1.13) that for $n>m$

$$\Sigma_{i=1}^m(a_{ni}-a_{mi})^2 + \Sigma_{i=m+1}^n a_{ni}^2 = \Sigma_{i=1}^n a_{ni}^2 - \Sigma_{i=1}^m a_{mi}^2 = D_n-D_m,$$

and therefore by (3.2), for $n>m$

$$E|x_n-x_m|^p \le K_p(D_n-D_m)^{p/2} = K_p(\Sigma_{i=m+1}^n d_i)^{p/2}.$$

As $n\to\infty$, $D_n = 1/v_{jj}^{(n)} \to \infty$ (cf. Lai, Robbins and Wei (1978)), and therefore we can apply Corollary 1 (ii) to obtain that for every $\delta>1/p$,

$$x_n = o(\{D_n^{1/2}(\log D_n)^{1/p}(\log\log D_n)^{\delta}\}) \quad \text{a.s.} \tag{3.3}$$

proving the desired conclusion (3.1). □

Corollary 2 extends the result of Lai, Robbins and Wei (1978) who considered the special case $p=4$. The above proof also shows that the linear transformation $x_n=\Sigma_{i=1}^n a_{ni}\,\epsilon_i$ of an S_p system $\{\epsilon_i\}$ has the asymptotic behavior (3.3) if $D_n = \Sigma_{i=1}^n a_{ni}^2 \to \infty$ and if the constants a_{ni} satisfy (1.13).

More generally, let $\{a_{ni} : n\ge 1, -\infty<i<\infty\}$ be a double array of constants such that

$$\Sigma_{i=-\infty}^{\infty} a_{ni}^2 < \infty \text{ for every } n. \tag{3.4}$$

Thus, $\mathbf{a}_n = (a_{ni})_{-\infty<i<\infty} \in \ell^2$, and we shall let $\|\mathbf{a}_n\| = (\Sigma_{i=-\infty}^{\infty} a_{ni}^2)^{1/2}$ denote the ℓ^2 norm of $\mathbf{a}_n$. Let $\{\epsilon_n\}_{-\infty<n<\infty}$ be an S_p system with $p>2$. Define

$$x_n = \Sigma_{i=-\infty}^{\infty} a_{ni}\epsilon_i, \tag{3.5}$$

noting that the series in (3.5) converges a.s. and in the L_p norm in view of Corollary 1(i) and (3.4). By (1.9),

$$E|x_n-x_m|^p \le K_p\{\Sigma_{i=-\infty}^{\infty}(a_{ni}-a_{mi})^2\}^{p/2} = K_p\|\mathbf{a}_n-\mathbf{a}_m\|^p. \tag{3.6}$$

If furthermore $\{\epsilon_n\}$ satisfies the Bessel inequality, then $E(x_n-x_m)^2 \ge K\,\Sigma_{i=-\infty}^{\infty}(a_{ni}-a_{mi})^2$ by (1.10), and it then follows from (3.6) that

$$E|x_n-x_n|^p \le K'_p\{E(x_n-x_m)^2\}^{p/2}. \tag{3.7}$$

This inequality in turn enables us to relate the L_p properties of $\{x_n\}$ to its L_2 and spectral properties. Making use of this observation, Lai and Wei (1983) obtained the following

THEOREM 1. *Consider the linear process* y_n *defined in* (1.7) *where the random errors* ϵ_n *form an orthonormal* S_p *system with* $p>2$. *Let* f *be the spectral density of* $\{y_n\}$. *If* ess $\sup_{0\le\theta\le 2\pi} f(\theta) < \infty$, *then* $\{y_n\}$ *is an* S_p *system. Consequently,* $\Sigma_{i=1}^{\infty} c_i y_i$ *converges a.s. and in the* L_p *norm for all constants* c_i *such that* $\Sigma_1^{\infty} c_i^2 < \infty$.

In view of the inequality (3.6), the random variables $x_n = \Sigma_{i=-\infty}^{\infty} a_{ni}\epsilon_i$ satisfy moment inequalities of the type in Lemma 1 or 2 or 3 if the function $h(m,n) = \|\mathbf{a}_n - \mathbf{a}_m\|$ satisfies corresponding conditions of the type $h(m,n) \leqslant (\Sigma_{i=m+1}^{\infty} d_i)^{\alpha/p}$, or $h(m,n) \leqslant g^{\alpha/p}(m+1,n)$, or $h(m, n) \leqslant g^{1/p}(n-m)$ for $n>m$. Under such assumptions on $\|\mathbf{a}_n - \mathbf{a}_m\|$, we can therefore apply the maximal inequalities in these lemmas to obtain almost sure limit theorems of the type in Corollary 1 above or in Corollary 3.3 of Lai and Wei (1983) for linear transformations $x_n = \Sigma_{i=-\infty}^{\infty} a_{ni}\epsilon_i$ of S_p systems $\{\epsilon_n\}$ satisfying the Bessel inequality. Such maximal inequalities can also be applied in conjunction with exponential bound of the type

(3.8) $$P\{|x_n| > \tau(\theta)(D_n \log\log D_n)^{1/2}\} = 0(\exp(-\theta \log\log D_n)),$$

where $D_n = \Sigma_{i=-\infty}^{\infty} a_{ni}^2$, $\theta>1$ and $\tau(\theta)>0$, to establish laws of the iterated logarithm for x_n (cf. Lai and Wei (1982), Theorem 4). Using this approach and certain truncation techniques, Lai and Wei (1982) obtained the following law of the iterated logarithm for double arrays of independent random variables and applied the result to regression and time series problems.

THEOREM 2. *Let* $\ldots, \epsilon_{-1}, \epsilon_0, \epsilon_1, \ldots$ *be independent random variables such that*

(3.9) $$E\epsilon_n = 0 \text{ and } E\epsilon_n^2 = \sigma^2 \text{ for all } n, \text{ and } \sup_n E|\epsilon_n|^p < \infty \text{ for some } p > 2,$$

and let $\{a_{ni} : n \geq 1, -\infty<i<\infty\}$ *be a double array of constants such that* (3.4) *holds and*

(3.10) $$D_n = \Sigma_{i=-\infty}^{\infty} a_{ni}^2 \to \infty,$$

(3.11) $$\sup_n a_{ni}^2 = o(D_n(\log D_n)^{-r}) \quad \text{for all } r>0.$$

Let $x_n = \Sigma_{i=\infty}^{\infty} a_{ni}\epsilon_i$.

(i) *If there exist constants* $d_i \geqslant 0$ *and* $\lambda > 1/p$ *such that*

(3.12) $$\|\mathbf{a}_n - \mathbf{a}_m\| < (\Sigma_{i=m+1}^{\infty} d_i)^{\lambda} \quad \text{for } n>m \geqslant m_0, \text{ and}$$

(3.13) $$(\Sigma_{i=1}^{n} d_i)^{\lambda} = 0(D_n^{1/2}) \quad \text{as } n\to\infty,$$

then

(3.14) $$\limsup_{n\to\infty} |x_n|/(2D_n \log\log D_n)^{1/2} \leqslant \sigma \quad a.s.$$

(ii) *If* $\|\mathbf{a}_n - \mathbf{a}_m\| \leqslant g^{1/p}(n-m)$, *where* $g : \{1, 2, \ldots\} \to (0,\infty)$ *satisfies conditions* (2.16) *and* (2.17) *and* $g(n) = 0(D_n^{p/2})$, *then* (3.14) *still holds.*

REFERENCES

BANACH, S. (1930). Über einige Eigenschaften der lacunaren trigonometrische Reihen. *Studia Math. 2* 207–220.

BURKHOLDER, D. L. (1966). Martingale transforms. *Ann. Math. Statist. 37* 1494–1504.

BURKHOLDER, D. L. (1968). Independent sequences with the Stein property. *Ann. Math. Statist. 39* 1282–1288.

BURKHOLDER, D. L. (1973). Distribution function inequalities for martingales. *Ann. Probability 1* 19–42.

CHEN, G. C., LAI, T. L., and WEI, C. Z. (1981). Convergence systems and strong consistency of least squares estimates in regression models. *J. Multivariate Anal. 11* 319–333.

DOOB, J. L. (1953). *Stochastic Processes.* Wiley, New York.

DRYGAS, H. (1976). Weak and strong consistency of the least squares estimators in regression models. *Z. Wahrscheinlichkeitstheorie und Verw. Gebiete 34* 119–127.

ERDÖS, P. (1943). On the convergence of trigonometric series. *J. Math. Phys. (M.I.T.) 22* 37–39.

GAPOSHKIN, V. F. (1966). Lacunary series and independent functions. *Russian Math. Surveys 21*, No. 6, 1–82.

KAILATH, T. (1974). A view of three decades of linear filtering theory. *IEEE Trans. Inform. Theory IT-20* 146–181.

KHINTCHINE, A. (1924). Über einen Satz der Wahrscheinlichkeitsrechnung. *Fund. Math. 6* 9–20.

KOMLÓS, J. (1972). On the series $\Sigma c_k \varphi_k$. *Studia Sci. Math. Hungar. 7* 451–458.

LAI, T. L., ROBBINS, H., and WEI, C. Z. (1978). Strong consistency of least squares estimates in multiple regression. *Proc. Nat. Acad. Sci. USA 75* 3034–3036.

LAI, T. L., and STOUT, W. (1980). Limit theorems for sums of dependent random variables. *Z. Wahrscheinlichkeitstheorie und Verw. Gebiete 51* 1–14.

LAI, T. L., and WEI, C. Z. (1982). A law of the iterated logarithm for double arrays of independent random variables with applications to regression and time series models. *Ann. Probability 10* 320–335.

LAI, T. L., and WEI, C. Z. (1983). Lacunary systems and generalized linear processes. *Stochastic Processes Appl. 14* 187–199.

LONGNECKER, M. and SERFLING, R. J. (1977). General moment and probability inequalities for the maximum partial sum. *Acta Math. Acad. Sci. Hungar. 30* 129–133.

LONGNECKER, M. and SERFLING, R. J. (1978). Moment inequalities for S_n under general dependence restrictions with applications. *Z. Wahrscheinlichkeitstheorie und Verw. Gebiete 43* 1–21.

MARCINKIEWICZ, J. et ZYGMUND, A. (1937). Sur les fonctions independantes. *Fund. Math. 29* 60–90.

MÓRICZ, F. (1976). Moment inequalities and the strong laws of large numbers. *Z. Wahrscheinlichkeitstheorie und Verw. Gebiete 35* 299–314.

MÓRICZ, F., SERFLING, R., and STOUT, W. (1982). Moment and probability bounds with quasi-superadditive structure for the maximum partial sum. *Ann. Probability 10* 1032–1040.

SIDON, S. (1934). Ein Satz über Fourierische Reihen mit Lücken. *Math. Z. 32* 481–482.

Inequalities in Statistics and Probability
IMS Lecture Notes–Monograph Series Vol. 5 (1984), 173–177

STOCHASTIC MAJORIZATION OF THE LOG-EIGENVALUES OF A BIVARIATE WISHART MATRIX[1]

BY MICHAEL D. PERLMAN
University of Washington

Let $l = (l_1, l_2)$ and $\lambda = (\lambda_1, \lambda_2)$, where $\lambda_1 \geq \lambda_2 > 0$ are the ordered eigenvalues of $\mathbf{S}$ and Σ, respectively, and $\mathbf{S} \sim W_2(n,\Sigma)$ is a bivariate Wishart matrix. Let $\mathbf{m} = (m_1, m_2)$ and $\mu = (\mu_1, \mu_2)$, where $m_i = \log l_i$ and $\mu_i = \log \lambda_i$. It is shown that $P_\mu\{\mathbf{m} \notin B\}$ is Schur-convex in μ whenever B is a Schur-monotone set, i.e. $[\mathbf{x} \in B,\ \mathbf{x}$ *majorizes* $\mathbf{x}^*] \Rightarrow \mathbf{x}^* \in B$. This result implies the unbiasedness and power-monotonicity of a class of invariant tests for bivariate sphericity and other orthogonally invariant hypotheses.

1. Introduction. Let $\mathbf{S} \sim W_2(n,\Sigma)$ be a bivariate Wishart matrix with n degrees of freedom ($n \geq 2$) and expected value $n\Sigma$ (Σ positive definite). We shall be concerned with the power functions of orthogonally invariant tests for invariant testing problems such as the following:

$$
(1.1)\qquad
\begin{array}{ll}
H_{01}\colon \Sigma = \sigma^2\mathbf{I},\ \sigma^2\ \textit{arbitrary} & vs.\ K_1\colon \Sigma\ \textit{arbitrary} \\
H_{02}\colon \Sigma = \mathbf{I} & vs.\ K_2\colon \Sigma\ \textit{arbitrary} \\
H_{03}\colon \Sigma = \mathbf{I} & vs.\ K_3\colon \Sigma - \mathbf{I}\ \textit{positive definite} \\
H_{04}\colon \Sigma = \mathbf{I} & vs.\ K_4\colon \Sigma - \mathbf{I}\ \textit{negative definite.}
\end{array}
$$

Orthogonally invariant tests depend on $\mathbf{S}$ only through $l = (l_1, l_2)$, where $l_1 \geq l_2 (>0)$ are the ordered eigenvalues of $\mathbf{S}$. Because the power functions of such tests depend on Σ only through $\lambda = (\lambda_1,\lambda_2)$, where $\lambda_1 \geq \lambda_2$ (> 0) are the ordered eigenvalues of Σ, we may assume throughout this paper that $\Sigma = \mathbf{D}_\lambda \equiv diag(\lambda_1,\lambda_2)$.

The notions of majorization and Schur-convexity play an important role in determining such properties as unbiasedness and power monotonicity of invariant tests. To illustrate, consider the likelihood ratio test (LRT) for testing H_{01} (bivariate sphericity) *vs.* K_1. The acceptance region can be expressed in the equivalent forms

$$
(1.2)\qquad \{\mathbf{S} \mid tr\mathbf{S}/|\mathbf{S}|^{1/2} \leq c\} \Leftrightarrow \{l \mid (l_1 + l_2)/(l_1 l_2)^{1/2} \leq c\}.
$$

Since

$$
(1.3)\qquad tr\mathbf{S}/|\mathbf{S}|^{1/2} = (s_{11} + s_{22})/((s_{11}s_{22})^{1/2}|\mathbf{R}|^{1/2}) = (e^{t_1}+e^{t_2})/(e^{(t_1+t_2)/2}|\mathbf{R}|^{1/2}),
$$

where $\mathbf{S} = (s_{ij})_{i,j=1,2}$, $\mathbf{R}$ is the sample correlation matrix, and $t_i = \log s_{ii}$, and since s_{11}, s_{22}, and $\mathbf{R}$ are independent with $s_{ii} \sim \lambda_i\chi_n^2$ when $\Sigma = D_\lambda$, conditioning on $\mathbf{R}$ reduces the problem to the study of the power function of the LRT for equality of scale parameters ($\lambda_1 = \lambda_2$) based on the independent χ^2-variates s_{11} and s_{22} with equal degrees of freedom. It is easy to show that the joint density of $\mathbf{t} \equiv (t_1,t_2)$ is Schur-concave (in fact, permutation-invariant and log concave) with location parameter $\mu \equiv (\mu_1,\mu_2) \equiv (\log \lambda_1, \log \lambda_2)$, and that for fixed $\mathbf{R}$ the region

[1] This research was supported in part by National Science Foundation Grants MCS 80–02167 and MCS 83–01807.
AMS 1980 subject classifications. Primary 62H15; Secondary 62H10.
Key words and phrases: Bivariate Wishart distribution, log-eigenvalues, stochastic majorization, Schur function.

$$\{\mathbf{t}\,|\,(e^{t_1}+e^{t_2})/(e^{(t_1+t_2)/2}) \leqslant c\,|\mathbf{R}|^{1/2}\} \equiv \{\mathbf{t}\,|\,2\cosh((t_1-t_2)/2) \leqslant c\,|\mathbf{R}|^{1/2}\} \tag{1.4}$$

is Schur-monotone in $\mathcal{R}^2$ (see Definition 2.1). By a well-known theorem of Marshall and Olkin (1979, 11.E.5.a, p. 299) it follows that

$$1 - P_{\mu}\{2\cosh((t_1-t_2)/2) \leqslant c\,|\mathbf{R}|^{1/2}\,|\,\mathbf{R}\} \tag{1.5}$$

is a Schur-convex function of μ, which implies that the power of the LRT for bivariate sphericity is Schur-convex in μ. This in turn implies that the LRT is unbiased and that its power function increases monotonically as μ moves away from the null hypothesis line H_{01}: $\mu_1 = \mu_2$ at right angles.

The preceding conditional argument, due to Gleser (1966) for unbiasedness and to the author (cf. Marshall and Olkin (1979), pp. 387–8) for Schur-convexity, applies equally well to the LRT for p-variate sphericity, $p \geq 3$. Our goal is to extend these results to invariant tests other than the LRT, and to orthogonally invariant testing problems other than sphericity. In this note we show by means of a similar conditional argument that a quite general extension (Theorem 2.4) is possible in the bivariate case. The p-variate case appears more difficult, however, since it is not possible to express the eigenvalues of $\mathbf{S}$ directly in terms of its diagonal elements s_{ii} and correlation matrix $\mathbf{R}$, and a general result is not yet available for this case. Perlman (1982) gave a partial result in the p-variate case by a different argument.

2. Definitions and Main Result. We refer to Marshall and Olkin (1979) for the necessary general definitions and properties of majorization and Schur-convex functions. The following definitions and remarks concern related properties of regions in $\mathcal{R}^2$. We set $\mathbf{x} = (x_1,x_2)$.

Definition 2.1. A set $B \subseteq \mathcal{R}^2$ is *Schur-monotone in* $\mathcal{R}^2$ if [$\mathbf{x} \in \mathbf{B}$, $\mathbf{x}$ *majorizes* $\mathbf{x}^*$] $\Rightarrow$ $\mathbf{x}^* \in B$.

A Schur-monotone set in $\mathcal{R}^2$ is necessarily permutation-invariant ($(x_1,x_2) \in B \Rightarrow (x_2,x_1) \in B$), and a simple characterization is possible: B is Schur-monotone in $\mathcal{R}^2$ iff B is of the form

$$B = \{\mathbf{x}\,|\,|x_1 - x_2| \leqslant f(x_1+x_2)\} \tag{2.1}$$

for an arbitrary function f on $(-\infty,\infty)$.

Definition 2.2. Let $\mathcal{R}_0^2 = \{\mathbf{x}\,|\,x_1 \geqslant x_2\}$. A set $B \subseteq \mathcal{R}_0^2$ is *Schur-monotone in* $\mathcal{R}_0^2$ *if* [$\mathbf{x} \in B$, $\mathbf{x}$ *majorizes* $\mathbf{x}^*$, $\mathbf{x}^* \in \mathcal{R}_0^2$] $\Rightarrow \mathbf{x}^* \in B$.

A set B is Schur-monotone in $\mathcal{R}_0^2$ iff its symmetric extension is Schur-monotone in $\mathcal{R}^2$. Equivalently, B is Schur-monotone in $\mathcal{R}_0^2$ iff B is of the form

$$B = \{\mathbf{x}\,|\,0 \leqslant x_1 - x_2 \leqslant f(x_1+x_2)\} \tag{2.2}$$

for an arbitrary function f on $(-\infty,\infty)$.

The expressions (2.1) and (2.2) suggest the following characterizations of Schur-convex functions on $\mathcal{R}^2$ and $\mathcal{R}_0^2$, respectively:

FACT 2.3. A real-valued function β on $\mathcal{R}^2$ is Schur-convex on $\mathcal{R}^2$ iff β is permutation-invariant and $\beta(y+a, y-a)$ is increasing in $|a|$ for each fixed y in $(-\infty,\infty)$. A function β on $\mathcal{R}_0^2$ is Schur-convex on $\mathcal{R}_0^2$ iff $\beta(y+a, y-a)$ is increasing in $a \geq 0$ for each fixed y.

Let $m_i = \log l_i$, $i = 1,2$, where $l_1 \geqslant l_2$ are the eigenvalues of S. Set $\mathbf{m} = (m_1,m_2)$ and recall that $\mu = (\mu_1,\mu_2)$ where $\mu_i = \log \lambda_i$. The following is our main result.

THEOREM 2.4. *If B is a Schur-monotone region in $\mathcal{R}_0^2$, then*

$$\beta(\mu) \equiv P_\mu\{\mathbf{m} \notin B\}$$

is a Schur-convex function of μ on $\mathcal{R}_0^2$.

Proof. The proof extends the argument in the second paragraph of Section 1. We shall show that the event $\{\mathbf{m} \in B\}$, when expressed in terms of $t_1(\equiv \log s_{11})$ and $t_2(\equiv \log s_{22})$ for fixed $\mathbf{R}$, is a Schur-monotone region in $\mathcal{R}^2$, so that the conditional probability $P_\mu\{\mathbf{m} \notin B \mid \mathbf{R}\}$ is a Schur-convex function of μ on $\mathcal{R}_0^2$. This will immediately yield the desired result. By (2.2), the event $\{\mathbf{m} \in B\}$ is of the form

$$\{\mathbf{m} \mid 0 \leq m_1 - m_2 \leq f(m_1 + m_2)\} \tag{2.3}$$

for some function f on $(-\infty,\infty)$. Since $m_i = \log l_i$, (2.3) is equivalent to

$$\{l \mid 1 \leq (l_1/l_2) \leq g(l_1 l_2)\} \tag{2.4}$$

for some nonnegative function g on $[0,\infty)$. In the bivariate case, however, the ordered characteristic roots $l_1 \geq l_2$ of $\mathbf{S}$ are given by

$$\tfrac{1}{2}\{tr\mathbf{S} \pm [(tr\mathbf{S})^2 - 4|\mathbf{S}|]^{1/2}\}.$$

so that

$$\begin{aligned} l_1/l_2 &= \{tr\mathbf{S} + [(tr\mathbf{S})^2 - 4|\mathbf{S}|]^{1/2}\}^2/(4|\mathbf{S}|) \\ &= \{(s_{11} + s_{22}) + [(s_{11} + s_{22})^2 - 4 s_{11}s_{22}|\mathbf{R}|]^{1/2}\}^2/(4s_{11}s_{22}|\mathbf{R}|) \\ &= \{\cosh((t_1 - t_2)/2) + [\cosh^2((t_1 - t_2)/2) - |\mathbf{R}|]^{1/2}\}^2/|\mathbf{R}|. \end{aligned} \tag{2.5}$$

By (2.4) and (2.5), therefore, the event $\{\mathbf{m} \in B\}$ is equivalent to

$$\{ (\mathbf{t},\mathbf{R}) | \cosh((t_1 - t_2)/2) + [\cosh^2((t_1 - t_2)/2) - |\mathbf{R}|]^{1/2} \leq [|\mathbf{R}| g(e^{t_1+t_2}|\mathbf{R}|)]^{1/2} \}. \tag{2.6}$$

Since $y + [y^2 - |\mathbf{R}|]^{1/2}$ is increasing in y for $y \geq 1$ (note that $|\mathbf{R}| \leq 1$) and since $\cosh y$ is an increasing function of $|y|$, it follows that for fixed $\mathbf{R}$. (2.6) is of the form

$$\{\mathbf{t} \mid |t_1 - t_2| \leq h(t_1 + t_2)\} \tag{2.7}$$

for some function h on $(-\infty,\infty)$. By (2.1) it follows that (2.7) is a Schur-monotone region in $\mathcal{R}^2$, which completes the proof. □

3. Applications to the Power Functions of Invariant Tests. We shall apply Theorem 2.4 with B and β representing the acceptance region and power function of an orthogonally invariant test for each of the testing problems in (1.1).

The testing problem H_{01} vs K_1 in (1.1) can be re-expressed in terms of $\mu \equiv (\mu_1,\mu_2)$ as

$$H_{01}: \mu_1 = \mu_2 \; vs. \; K_1: \mu_1 \neq \mu_2.$$

The acceptance region of the likelihood ratio test (LRT) can be expressed in the equivalent forms

$$\begin{aligned} A_{01} &= \{l \mid (l_1 + l_2)/(l_1 l_2)^{1/2} \leq c\} \\ \Leftrightarrow B_{01} &= \{\mathbf{m} \mid 2\cosh((m_1 - m_2)/2) \leq c\} \end{aligned}$$

(cf. (1.2)–(1.4)). This is of the form (2.2), so B_{01} is a Schur-monotone region in $\mathcal{R}_0^2$. Thus Theorem 2.4 applies, so the power function of the LRT,

$$\beta(\mu) \equiv P_\mu\{\mathbf{m} \notin B_{01}\},$$

in Schur-convex in μ, as already seen in Section 1.

The testing problem H_{02} vs. K_2 can be re-expressed as

$$H_{02}: \mu_1 = \mu_2 = 0 \; vs. \; K_2: (\mu_1,\mu_2) \neq (0,0).$$

The acceptance region of the LRT can be written in the equivalent forms

$$A_{02} = \{l \mid \Sigma_{i=1}^{2} [\log(l_i/n) - (l_i/n) + 1] \geq c\}$$
$$\Leftrightarrow \; B_{02} = \{\mathbf{m} \mid \Sigma_{i=1}^{2} \gamma(m_i - \log n) \geq c\}.$$

where $\gamma(y) = y - e^y + 1$. Since γ is a concave function on $(-\infty,\infty)$, the symmetric extension of B_{02} to $\mathcal{R}^2$ is a convex, permutation-invariant region, hence B_{02} is a Schur-monotone region in $\mathcal{R}_0^2$. By Theorem 2.4, therefore, the power function of the LRT is Schur-convex in μ. Other invariant acceptance regions appropriate for testing H_{02} *vs* K_{02} include the regions

$$A_{(r)} = \{l \mid |\log l_1|^r + |\log l_2|^r \leq c\}$$
$$\Leftrightarrow B_{(r)} = \{\mathbf{m} \mid |m_1|^r + |m_2|^r \leq c\}$$

with $r > 0$. For $r \geq 1$, the symmetric extension of $B_{(r)}$ to $\mathcal{R}^2$ is convex and permutation-invariant, so the corresponding power function is Schur-convex in μ. Other acceptance regions possibly appropriate for this problem are the regions

$$A_{[r]} = \{l \mid (\log(l_1/l_2))^r + |\log l_1 l_2|^r \leq c\}$$
$$\Leftrightarrow B_{[r]} = \{\mathbf{m} \mid (m_1 - m_2)^r + |m_1 + m_2|^r \leq c\}$$

with $r > 0$. For each $r > 0$, $B_{[r]}$ is a Schur-monotone region in $\mathcal{R}_0^2$ by (2.2), so the corresponding power function is Schur-convex in μ. Note that for $0 < r < 1$, $B_{[r]}$ is not convex.

Next, we discuss a class of one-sided acceptance regions based on $tr(\mathbf{S}^r)$ appropriate for testing

$$H_{03}\colon \mu_1 = \mu_2 = 0 \; vs. \; K_3\colon \mu_1 \geq \mu_2 \geq 0.$$

For $-\infty \leq r \leq \infty$ define

$$T_r = [\tfrac{1}{2} tr(\mathbf{S}^r)]^{1/r} = [\tfrac{1}{2}(l_1^r + l_2^r)]^{1/r} = [\tfrac{1}{2}(e^{rm_1} + e^{rm_2})]^{1/r};$$

note that

$$T_\infty = l_1 = e^{m_1}, \qquad T_{-\infty} = l_2 = e^{m_2},$$
$$T_0 = |S|^{1/2} = (l_1 l_2)^{1/2} = e^{(m_1+m_2)/2},$$

by continuity. The equivalent acceptance regions

$$A_r = \{l \mid T_r \leq c\} \quad \Leftrightarrow \quad B_r = \{\mathbf{m} \mid T_r \leq c\}$$

are appropriate for testing H_{03} *vs.* K_3. For each $r \geq 0$ $(r \leq 0)$ the symmetric extension of $B_r(B_r^c)$ to $\mathcal{R}^2$ is convex and permutation-invariant, so by Theorem 2.4 the corresponding power function is Schur-convex (Schur-concave) in μ. [For $r = 0$, the distribution of $\mathbf{T}_0 = |\mathbf{S}|^{1/2}$ depends on μ only through $\mu_1 + \mu_2 \equiv \log|\Sigma|$, so the power function corresponding to B_0 is trivially both Schur-convex and Schur-concave in μ.]

Similarly, the equivalent regions $A_r^c \Leftrightarrow B_r^c$ complementary to $A_r \Leftrightarrow B_r$ are appropriate acceptance regions for testing

$$H_{04}\colon \mu_1 = \mu_2 = 0 \quad vs. \quad K_4\colon 0 \geq \mu_1 \geq \mu_2.$$

It follows from the preceding paragraph that the power function associated with the acceptance region B_r^c is Schur-convex (Schur-concave) for $r < 0$ $(r > 0)$.

We conclude this section by considering the LRT's for testing H_{03} *vs.* K_3 and H_{04} *vs.* K_4. It can be shown (cf. Perlman (1967)) that the acceptance region of the LRT for H_{03} *vs.* K_3 can be expressed in the equivalent forms

$$A_{03} = \{\, l \mid \Sigma_{\{i \mid l_i \geqslant n\}}[\log(l_i/n) - l_i/n + 1] \geqslant c\}$$
$$\Leftrightarrow B_{03} = \{\mathbf{m} \mid \Sigma_{\{i \mid m_i \geqslant \log n\}}[(m_i - \log n) - e^{(m_i - \log n)} + 1] \geqslant c\}$$
$$= \{\mathbf{m} \mid \Sigma_{i=1}^{2} \varphi(m_i - \log n) \geqslant c\}.$$

where
$$\varphi(y) = \begin{cases} y - e^y + 1 & \text{if } y \geq 0 \\ 0 & \text{if } y \leq 0. \end{cases}$$

Since φ is a concave function on $(-\infty,\infty)$, B_{03} is a Schur-monotone region in $\mathcal{R}_0^2$ and the associated power function is Schur-convex in μ. Similarly, the acceptance region of the LRT for H_{04} *vs*. K_4 can be expressed as

$$A_{04} = \{\, l \mid \Sigma_{\{i \mid l_i \leqslant n\}}[\log(l_i/n) - (l_i/n) + 1] \geqslant c\}$$
$$\Leftrightarrow B_{04} = \{\mathbf{m} \mid \Sigma_{i=1}^{2} \psi(m_i - \log n) \geqslant c\}.$$

where
$$\varphi(y) = \begin{cases} y - e^y + 1 & \text{if } y \leq 0 \\ 0 & \text{if } y \geq 0. \end{cases}$$

Since ψ is concave on $(-\infty,\infty)$, if follows as above that the power function associated with B_{04} is Schur-convex in μ.

Other power monotonicity properties of some of the tests discussed in this section may be found in Anderson and Das Gupta (1964), Das Gupta (1969), and Das Gupta and Giri (1971).

4. Concluding Remarks. The proof of Theorem 2.4 proceeded indirectly, expressing the eigenvalues l_1, l_2 of S in terms of its diagonal elements s_{11}, s_{22} and working with the relatively simple joint distribution of (s_{11}, s_{22}). This approach may not easily extend to the p-variate case ($p \geqslant 3$), so it may be preferable to work directly with the joint distribution of the eigenvalues $l_1, \ldots, l_p$ (cf. Muirhead (1982), Theorem 9.4.1). We admit, however, that we were unable to carry through the latter approach in the bivariate case.

Theorem 2.4 can be extended from probabilities of Schur-monotone regions to expectations of Schur-convex functions: if g is a Schur-convex function on $\mathcal{R}_0^2$ such that the expectations exist, then $E_\mu[g(\mathbf{m})]$ is a Schur-convex function of μ. This follows by a standard approximation argument.

REFERENCES

ANDERSON, T. W. and DAS GUPTA, S. (1964). A monotonicity property of the power functions of some tests of equality of two covariance matrices. *Ann. Math. Statist. 35* 1059–1063. [Correction: *ibid 36* (1965), 1318.]

DAS GUPTA, S. (1969). Properties of power functions of some tests concerning dispersion matrices of multivariate normal distirubtions. *Ann. Math. Statist. 40* 697–701.

DAS GUPTA, S. and GIRI, N. (1973). Properties of tests concerning covariance matrices of normal distributions. *Ann. Statist. 1* 1222–1224.

GLESER, L. J. (1966). A note on the sphericity test. *Ann. Math. Statist. 37* 464–467.

MARSHALL, A. W. and OLKIN, I. (1979). *Inequalities: Theory of Majorization and Its Applications*. Academic Press, New York.

MUIRHEAD, R. J. (1982). *Aspects of Multivariate Statistical Theory*. Wiley, New York.

PERLMAN, M. D. (1967). One-sided testing problems involving covariance matrices. Unpublished notes, Department of Statistics, Stanford University, Stanford, CA.

PERLMAN, M. D. (1982). Stochastic comparison of eigenvalues of Wishart matrices with applications to unbiasedness of tests for variance and covariance matrices. Invited address presented at the Symposium on Inequalities in Statistics and Probability, University of Nebraska, October 27–30, 1982.

Inequalities in Statistics and Probability
IMS Lecture Notes–Monograph Series Vol. 5 (1984), 178–186

APPROXIMATIONS AND ERROR BOUNDS IN STOCHASTIC PROGRAMMING

BY JOHN R. BIRGE and ROGER J.-B. WETS[1]
University of Michigan and University of Kentucky

We review and complete the approximation results for stochastic programs with recourse. Since this note is to serve as a preamble to the development of software for stochastic programming problelms, we also address the question of how to easily find a (starting) solution.

We consider the *stochastic program with (fixed) recourse* (Wets, 1983))

(0.1) find $x \in \mathscr{R}_+^{n_1}$ such that $Ax = b$ and $z = cx + \mathscr{Q}(x)$ is minimized

where A is $m_1 \times n_1$, $b \in \mathscr{R}^{m_1}$, and

(0.2) $$\mathscr{Q}(x) = E\{Q(x,\xi)\} = \int Q(x,\xi)\, P(d\xi)$$

with P a probability measure defined on $\Xi \subset \mathscr{R}^{m_2}$, and

(0.3) $$Q(x,\xi) = \inf_{y \in \mathscr{R}_+^{n_2}} \{qy \mid Wy = \xi - Tx\},$$

W is $m_2 \times n_2$, T is $m_2 \times n_1$, $q \in \mathscr{R}^{n_2}$ and $\xi \in \mathscr{R}^{m_2}$. We think of Ξ as the set of possible values of a random vector. Technically this means that Ξ is the support of the probability measure P. We shall assume that $\bar{\xi} = E\{\xi\}$ exists.

Many properties are known about problems of this type (Wets (1983)). For our purposes, the most important ones are

(0.4) $\xi \mapsto Q(x,\xi)$ is a convex piecewise linear function for all feasible x, i.e. $x \in K = K_1 \cap K_2$

where

$$K_1 = \{x \mid Ax = b, x \geqslant 0\}$$

$$K_2 = \{x \mid \text{for every } \xi \in \Xi, \text{ there exists a } y \geqslant 0 \text{ such that } Wy = \xi - Tx\},$$

and

(0.5) $x \mapsto Q(x,\xi)$ is a convex piecewise linear function which imples that

(0.6) $x \mapsto \mathscr{Q}(x)$ is a convex function, finite on K_2 (as follows from the integrability condition on Ξ).

It is also useful to consider an equivalent formulation of (0.1) that stresses the fact that choosing x corresponds to generating a *tender* $\chi = Tx$ to be bid by the decision maker against the outcomes ξ of the random events, viz.

(0.7) find $x \in \mathscr{R}_+^{n_1}$, $\chi \in \mathscr{R}^{m_2}$ such that $Ax = b$, $Tx = \chi$, and $z = cx + \psi(\chi)$ is minimized,

where

(0.8) $\Psi(\chi) = \mathrm{E}\{\psi(\chi,\xi)\} = \int \psi(\chi,\xi)\, \mathrm{P}(d\xi)$,

and

(0.9) $\psi(\chi,\xi) = \inf_{y \in \mathscr{R}_+^{n_2}} \{qy \mid Wy = \xi - \chi\}$.

---

1 Supported in part by a grant of the National Science Foundation.
AMS 1980 subject classifications. 90C15.
Key words and phrases: Stochastic programming, error bounds, recourse problem.

The functions ψ and Ψ have basically the same properties as Q and $\mathcal{Q}$, replacing naturally K_2 by the set

$$L_2 = \{\chi \mid \text{for every } \xi \in \Xi, \text{ there exists a } y \geq 0 \text{ such that } Wy = \xi - \chi\}.$$

Let z^* denote the optimal value of (0.1) or equivalently (0.7). We are interested in finding bounds on z^* by approximating $\mathcal{Q}$ or Ψ.

1. Lower Bounds. A lower bound for z^* can be obtained by solving the linear program

(1.1) find $x \geq 0, y \geq 0$ such that $Ax = b$, $Tx + Wy = \bar{\xi}$, and $cx + qy = z$ is minimized.

To see this note that (1.1) can also be expressed as

(1.2) find $x \in \mathcal{R}_+^{n_1}$ such that $Ax = b$ and $z = cx + Q(x,\bar{\xi})$ is minimized,

and with $\bar{z}$ denoting the optimal value of (1.2). We certainly have that $\bar{z} \leq z^*$ if we show that

(1.3) $$Q(\cdot,\bar{\xi}) \leq \mathcal{Q}(\cdot).$$

But this follows from (0.4) and Jensens' inequality:

(1.4) $$Q(x, \mathrm{E}\xi) \leq \mathrm{E}\{Q(x,\xi)\}$$

for every $x \in K_2$. There is another way to obtain this inequality, relying on the dual solution to (1.1):

(1.5) find $\sigma \in \mathcal{R}^{m_1}$, $\pi \in \mathcal{R}^{m_2}$ such that $\sigma A + \pi T \leq c$, $\pi W \leq q$, $\sigma b + \pi\bar{\xi} = w$ is maximized.

Let $(\bar{\sigma},\bar{\pi})$ be an optimal solution to this linear program. Since $\bar{\pi}W \leq q$, it follows again from the duality theory of linear programming that

$$Q(x,\xi) = \sup_{\pi\in\mathcal{R}^{m_2}}\{\pi(\xi - Tx) \mid \pi W \leq q\} \geq \bar{\pi}(\xi - Tx)$$

and also that, for $x \in K$,

$$\begin{aligned} cx + \mathcal{Q}(x) &\geq cx + \bar{\pi}\cdot E\xi - \bar{\pi}Tx = \bar{\pi}\xi + (c - \bar{\pi}T)x \\ &= \bar{\pi}\xi + \bar{\sigma}Ax = \bar{\pi}\bar{\xi} + \bar{\sigma}b = w_{\text{opt}} = \bar{z}. \end{aligned}$$

Hence

(1.6) $$\bar{z} \leq \inf_{x\in K} cx + \mathcal{Q}(x) = z^*.$$

Madansky (1960) was the first to point out that this type of reasoning provided error bounds for stochastic programs. We can refine this lower bound in a number of ways.

The first one is to use a sharper version of Jensens' inequality. Let $\mathcal{S}^\nu = \{\Xi_l, l = 1, \ldots, \nu\}$ be a partition of Ξ and let us denote by $\bar{\xi}^l$ the conditional expectation of ξ given that its values are in Ξ_l, i.e., $\bar{\xi}^l = E\{\xi | \xi \in \Xi_l\}$. Also, let $f_l = P(\Xi_l)$, i.e., f_l is the probability that $\xi \in \Xi_l$. The convexity of $Q(x,\cdot)$ yields

(1.7) $$Q(x,\bar{\xi}) \leq \Sigma_{l=1}^{\nu} f_l Q(x,\bar{\xi}^l) \leq E\{Q(x,\xi)\} = \mathcal{Q}(x)$$

as follows from a generalization of Jensens' inequality (Perlman (1974)). Denote by $\bar{z}^\nu$ the optimal value of the linear program:

(1.8) find $x \geq 0$ such that $Ax = b$, $Tx + Wy^l = \bar{\xi}^l$, $l = 1, \ldots, \nu$, and $cx + \Sigma_{l=1}^{\nu} f_l q y^l = z$ is minimized,

which can also be written in the form

(1.9) find $x \in \mathcal{R}_+^{n_1}$ such that $Ax = b$ and $z = cx + \Sigma_{l=1}^{\nu} f_l Q(x,\bar{\xi}^l)$ is minimized.

In view of (1.7), it follows that

(1.10) $$\bar{z} \leq \bar{z}^\nu \leq z^*.$$

The same reasoning shows that if $\mathcal{S}^{\nu'} = \{\Xi_k, k = 1, \ldots, \nu'\}$ is a finer partition of Ξ, i.e., for all $k = 1, \ldots, \nu'$, $\Xi_k \subset \Xi_l$, for some $\Xi_l \in \mathcal{S}^\nu$, and if $\bar{z}^{\nu'}$ is the optimal value of the linear program of type (1.8) that corresponds to this partition, then

$$\bar{z} \leqslant \bar{z}^\nu \leqslant \bar{z}^{\nu'} \leqslant z^*. \tag{1.11}$$

In fact the $\bar{z}^\nu$ converge to z^* provided that the partitions $\mathcal{S}^\nu$ are such that the probability measures they generate, viz.

$$P^\nu(A) = \Sigma_{\{l|\bar{\xi}^l \in A\}} P(\Xi_l),$$

converge in distribution to P, as follows from Theorem (3.9) of Wets (1983). The suggestion to rely on conditional expectations to refine (1.6) is due to Kall (1974) and to Huang, Ziemba and Ben-Tal (1977) who give a detailed analysis of these bounds when Ψ is separable.

Another method is to proceed as follows: For every $\xi \in \Xi$, and some $\hat{\xi} \in \operatorname{co} \Xi$ (the convex hull of Ξ), we define

$$\phi(\hat{\xi},\xi) = \inf cx + \hat{p}q\hat{y} + (1-\hat{p})qy_\xi \text{ such that } Ax = b, Tx + W\hat{y} = \hat{\xi}, \tag{1.12}$$
$$Tx + Wy_\xi = \xi, x \geqslant 0, \hat{y} \geqslant 0, y_\xi \geqslant 0$$

with $\hat{p} \in [0,1]$. If (0.1) is solvable, so is (1.12) for all $\xi \in \Xi$ as follows directly from Walkup and Wets (1967). Let x° solve (0.1) and for all ξ

$$y^\circ(\xi) \in \operatorname{argmin}_{y \in \mathcal{R}_+^{n_2}} \{qy | Wy = \xi - Tx^\circ\}.$$

It is well known that the $y^\circ(\xi)$ can be chosen so that as a function of ξ, $y^\circ(\cdot)$ is measurable, cf. Walkup and Wets (1967). Now let $\hat{\xi} = \bar{\xi}$ and $\bar{y}^\circ = E\{y^\circ(\xi)\}$. The triple $(x^\circ, \bar{y}^\circ, y^\circ(\xi))$ is a feasible solution of the linear program (1.12) when $\hat{\xi} = \bar{\xi}$. However in general it is not an optimal solution.

Whence

$$\phi(\bar{\xi},\xi) \leqslant cx^\circ + \hat{p}q\bar{y}^\circ + (1-\hat{p})qy^\circ(\xi) \tag{1.13}$$

and integrating this on both sides with respect to P we obtain

$$E\{\phi(\bar{\xi},\xi)\} \leqslant cx^\circ + \mathcal{Q}(x^\circ) = z^* \tag{1.14}$$

which gives us a new lower bound for z^*. This bound can be refined in many ways: first instead of using just one point $\bar{\xi}$ we can use a collection of points obtained as conditional expectations of a partition of Ξ. Second we can increase the number of points that are taken to build (1.12) as an approximation to (0.1). A detailed discussion appears in Birge (1982).

A lower bound of a somewhat different nature still using the convexity of Q, but not based on Jensens' inequality per se, can be obtained as follows. Let $\{\xi^l, l = 1, \ldots, \nu\}$ be a collection of points in Ξ and let

$$\pi^l \in \operatorname{argmax} [\pi(\xi^l - Tx) \mid \pi W \leqslant q].$$

Then $\pi^l \in \partial_\xi Q(x,\xi^l)$, i.e. the subgradient of Q with respect to ξ at ξ^l (for given x). We have that $Q(x,\xi^l) = \pi^l(\xi^l - Tx)$ and

$$Q(x,\xi) \geqslant \pi^l(\xi - Tx) \text{ for all } \xi \in \Xi. \tag{1.15}$$

The last inequality follows from the simple observation that

$$Q(x,\xi) = \sup [\pi(\xi - Tx) \mid \pi W \leqslant q]$$

and that π^l is a feasible, but not necessarily optimal, solution for the sup-problem defining Q. Since (1.15) holds for every l, we have

$$Q(x,\xi) \geqslant \max_{1 \leqslant l \leqslant \nu} \pi^l(\xi - Tx).$$

Integrating on both sides yields

(1.16) $$\mathcal{Q}(x) \geq E\{\max_{1 \leq \ell \leq \nu} \pi^{\ell}(\xi - Tx)\}.$$

In general finding the maximum for each ξ may be difficult. But we may assign each π^{ℓ} to a subregion of Ξ; this bound is not as tight as (1.16) but we can refine it by taking successively finer and finer partitions. However one should not forget that (1.16) involves a rather simple integral and the expression to the right could be evaluated numerically (to an acceptable degree of accuracy) without major difficultiies. Note that the calculation of this lower bound does not require the ξ^{ℓ} to be conditional expectations or chosen in any specific manner, however it should be obvious that a well chosen spread of the $\{\xi^{\ell}, \ell = 1, \ldots, \nu\}$ will give us sharper bounds. Also, the use of larger samples, i.e. by increasing ν, will also yield a better lower bound.

2. Upper Bounds. If $\mathcal{Q}(x)$ is easily computable, a simple upper bound is given by $z^* \leq c\hat{x} + \mathcal{Q}(\hat{x})$ for any feasible $\hat{x}$ in K. In particular, if $\bar{x}$ solves (1.1) and it turns out that $\bar{x} \in K$, then we have that

(2.1) $$\bar{z} = c\bar{x} + Q(\bar{x}, \bar{\xi}) \leq z^* \leq c\bar{x} + \mathcal{Q}(\bar{x}).$$

In general we cannot infer that $\bar{x} \in K$ simply from knowing that $\bar{x}$ solves (1.1), unless we know that we are dealing with a stochastic program with complete recourse, or more generally with relatively complete recourse, Wets (1983), i.e., when $K = \{x | Ax = b, x \geq 0\}$. Refinements of this bound, relying on different values of x may be found in Kall (1979) and Birge (1980) but they always involve the evaluation of $\mathcal{Q}(x)$.

Without evaluating $\mathcal{Q}$, we may find upper bounds for $\mathcal{Q}$ by considering the extreme points of coΞ. Let us assume in what follows that Ξ is compact, then so is its convex hull and $\Xi = \text{co}(\text{ext}\,\Xi)$ where ext Ξ are the extreme points of Ξ. Since $Q(x,\xi)$ is convex in ξ, we have that for all $\xi \in \Xi$

$$\begin{aligned} Q(x,\xi) &\leq \sup_{\xi \in \Xi} Q(x,\xi), \\ &= Q(x, e^{(x)}), \quad \text{for some } e^{(x)} \in \text{ext}\,\Xi. \\ &= \max_{e \in \text{ext}\,\Xi} Q(x,e). \end{aligned}$$

Now $e^{(x)}$ may depend on x, but we always have that

(2.2) $$\mathcal{Q}(x) \leq \max_{e \in \text{ext}\,\Xi} Q(x,e) = Q(x, e^{(x)}),$$

and hence

(2.3) $$z^* \leq \inf_{x \in K}[cx + (\max_{e \in \text{ext}\,\Xi} Q(x,e))].$$

If there are only a finite number of extreme points of Ξ, as is usually the case in practice, the function appearing on the right hand side of the inequality can be minimized without major difficulties. Let $\{e^j, j = 1, \ldots, J\} = \text{ext}\,\Xi$ be this finite collection of extreme points. We have to solve the mathematical program

(2.4) find $x \in \mathcal{R}_+^{n_1}$ and $\theta \in \mathcal{R}$ such that $Ax = b$, $Q(x, e^j) \leq \theta$ for $j = 1, \ldots, J$ and $cx + \theta$ is minimized.

The last condition can also be expressed as

$$\theta \geq qy^j, \; Wy^j = e^j - Tx, \; y^j \geq 0 \quad \text{for } j = 1, \ldots, J.$$

Thus (2.4) becomes equivalent to the linear program

(2.5) find $x \in \mathcal{R}_+^{n_1}$, $\theta \in R$ and $(y^j \in \mathcal{R}_+^{n_2}, j = 1, \ldots, J)$ such that $Ax = b$, $Tx + Wy^j = e^j$, $\theta \geq qy^j$ for $j = 1, \ldots, J$ and $cx + \theta$ is minimized.

The optimal value yields the upper bound for z^*.

This is a very crude bound. We can improve on this, as follows: every $\xi \in \Xi$ also belongs to co(ext Ξ). We can thus find $\{\lambda_j(\xi), j = 1, \ldots, J\}$ such that $\lambda_j(\xi) \geq 0$, $\Sigma_{j=1}^J \lambda_j(\xi) = 1$, and $\Sigma_{j=1}^J \lambda_j(\xi)\, e^j = \xi$. We write $\lambda_j(\xi)$ to indicate the dependence of the λ_j on ξ. By convexity of $Q(x,\cdot)$,

$$Q(x,\xi) \leq \Sigma_{j=1}^J \lambda_j(\xi)\, Q(x,e^j).$$

Taking the expectation on both sides yields

$$\begin{aligned}\mathcal{Q}(x) &\leq \textstyle\int_\Xi \Sigma_{j=1}^J \lambda_j(\xi)\, Q(x,e^j)\, P(d\xi)\\ &= \textstyle\int_\Lambda \Sigma_{j=1}^J \lambda_j Q(x,e^j)\, G(d\lambda)\end{aligned} \tag{2.6}$$

where G is the distribution function induced by P on $\Lambda = \{\lambda \in \mathcal{R}^J | \Sigma_{j=1}^J \lambda_j = 1, \lambda_j \geq 0\}$.

If coΞ is a simplex, then each $\xi \in \Xi$ is obtained by a unique convex combination of the extreme points. It is not difficult to actually derive G, calculate the last integral and then minimize the resulting function to obtain an upper bound for z^*. In general Ξ is not a simplex. We shall see later what to do in the general case, but there is an important class of problems that reduces to the case where Ξ is a simplex.

Suppose the random variables (of the m_2 vector) are independent. Then the distribution function (or the probability measure) is separable and (2.6) can be written as

$$\begin{aligned}\mathcal{Q}(x) &= \textstyle\int_{\alpha_{m_2}}^{\beta_{m_2}} P_{m_2}(d\xi_{m_2}) \ldots \int_{\alpha_2}^{\beta_2} P_2(d\xi_2) \int_{\alpha_1}^{\beta_1} P_1(d\xi_1)\, Q(x, (\xi_1,\xi_2, \ldots, \xi_{m_2}))\\ &\leq \textstyle\int_{\alpha_{m_2}}^{\beta_{m_2}} P_{m_2}(d\xi_{m_2}) \ldots \int_{\alpha_2}^{\beta_2} P_2(d\xi_2) \int_0^1 G_1(d\lambda_1)\, Q^1(x, (\lambda_1,\xi_2, \ldots, \xi_{m_2}))\end{aligned} \tag{2.7}$$

where

$$Q^1(x, (\lambda_1,\xi_2, \ldots, \xi_{m_2})) = (1-\lambda_1) Q(x, (\alpha_1,\xi_2, \ldots, \xi_{m_2})) + \lambda_1 Q(x, (\beta_1,\xi_2, \ldots, \xi_{m_2}))$$

and for each i, $\Xi_i = [\alpha_i,\beta_i]$ and $\Xi = \times_{i=1}^{m_2} \Xi_i$. Since $\xi_1 = (1 - \lambda_1)\, \alpha_1 + \lambda\beta_1$ we get the following expression for $\lambda_1(\xi_1)$:

$$\lambda_1 = (\xi_1-\alpha_1)/(\beta_1-\alpha_1), \quad \text{and} \quad 1-\lambda_1 = (\beta_1-\xi_1)/(\beta_1-\alpha_1).$$

Hence, with $\mu_1 = E\{\xi_1\}$,

$$\begin{aligned}&\textstyle\int_0^1 Q^1(x, (\lambda_1, \xi_2, \ldots, \xi_m))\, G_1(d\lambda_1) =\\ &((\beta_1-\mu_1)/(\beta_1-\alpha_1))\, Q(x, (\alpha_1, \xi_2, \ldots, \xi_{m_2})) + ((\mu_1-\alpha_1)/(\beta_1-\alpha_1))\, Q(x, (\beta_1,\xi_2, \ldots, \xi_{m_2}))\end{aligned} \tag{2.8}$$

which we can substitute in (2.7) for the integral with respect to λ_1. We can repeat this process for each ξ_i to obtain a bound on $\mathcal{Q}$ involving only the evaluation of the function $Q(x,\cdot)$ at the vertices of the retangular region Ξ.

The whole argument really boils down to the use of the simple inequality for real-valued convex functions ϕ of a random variable ξ, with distribution P on $[\alpha,\beta]$ and expectation μ.

$$\textstyle\int_\alpha^\beta \phi(\xi) P(d\xi) \leq ((\beta-\mu)/(\beta-\alpha))\phi(\alpha) + ((\mu-\alpha)/(\beta-\alpha))\phi(\beta) \tag{2.9}$$

This inequality is due to Edmundson. Madansky (1960) used it in the context of stochastic programs (with simple recourse) to obtain a simple version of (2.7). A much refined version of this upper bound can be obtained by partitioning the interval $[\alpha,\beta]$ and using (2.9) for each interval in the partition, substituting the end points of the subinterval for α and β, and the conditional expectation (with respect to this subinterval) for μ. In the case of stochastic programs with simple recourse this was carried out by Huang, Ziemba and Ben-Tal (1977) and by Kall and Stoyan (1982) who also consider stochastic problems of a more general nature.

Also, when P is not separable we can improve somewhat on (2.3) by observing that we can use (2.9) with respect to one random variable, say ξ_1. We have

$$\int Q(x,\xi)\,P(d\xi_1,\xi_2,\ \dots\ ,\xi_{m_2}) \leq \sup_{\{(\xi_2,\ \dots\ ,\xi_{m_2})\,|\,\xi\in\Xi\}} \int Q(x,\xi)\,P(d\xi_1,\xi_2,\ \dots\ ,\xi_{m_2})$$
$$= \sup_{\{(\bar{e}1,\ \dots\ ,\bar{e}^{\xi})\,|\,e^j=(e^j_1,\bar{e}^j)\in \mathrm{ext}\,\Xi\}}[((\beta_1-\mu_1(e^j))/(\beta_1-\alpha_1))\,Q(x,(\alpha_1,\bar{e}^j))$$
$$+ ((\mu_1(e^j)-\alpha_1)/(\beta_1-\alpha_1))\,Q(x,(\beta_1,\bar{e}j)]$$

where $\mu_1(e^j)$ is the conditional expectation of ξ_1 given $\bar{e}^j$ (the last (m_2-1) coordinates of e^j). From this it follows that

(2.10) $$\mathcal{Q}(x) \leq \min_{1\leq i\leq m_2} \sup_{\{\bar{e}^j|e^j\in \mathrm{ext}\,\Xi\}}[((\beta_1-\mu_1(e^j))/(\beta_1-\alpha_1))\,Q(x,(\alpha_i,e^{-j}))$$
$$+ ((\mu_i(e^j)-\alpha_i)/(\beta_i-\alpha_i))\,Q(x,(\beta_i,\bar{e}^j))],$$

where it must be understood that $\bar{e}^j$ consists of the (m_2-1) components of e^j that are not indexed by i. Further refinements through the partitioning of Ξ and the use of the corresponding conditional means, tighten up this inequality.

Another refinement of (2.3), in the case of nonseparable measure P, can be obtained by considering simplicial decompositions of Ξ, assuming naturally that Ξ admits such a decomposition (which means that Ξ should be polyhedral). Let $\mathcal{S} = \{\mathcal{S}^l,\ l = 1,\ \dots\ ,L\}$ be such a decomposition (technically $\mathcal{S}$ is a complex whose cells $\mathcal{S}^l$ are simplices). Let $\{e^l_0,\ \dots\ ,e^l_{m_2}\}$ be the vertices of the simplex $\mathcal{S}^l$, assuming that $\dim \Xi = m_2$. Then each $\xi \in \mathcal{S}^l$ determines a unique vector of barycentric coordinates $(\lambda^l_0,\ \dots\ ,\lambda^l_{m_2})$ such that

$$\xi = \Sigma_{j=0}^{m_2}\lambda^l_j e^l_j,\ \lambda^l_j \geq 0,\ \Sigma_0^{m_2}\lambda^l_j = 1.$$

On $\mathcal{S}^l$, we are thus given a simple formula for the relationship between the distribution of ξ and the induced distribution for the λ^l_js. We have

$$\int_{\mathcal{S}^l} Q(x,\xi)\,P(d\xi) \leq \int_\Lambda \Sigma_{j=0}^{m_2}\lambda_j Q(x,e^l_j)\,G_l(d\lambda) = \bar{Q}_l(x,\mathcal{S}^l)$$

where $\Lambda = \{\lambda \in \mathcal{R}^{m_2+1}\,|\Sigma_{j=0}^{m_2}\lambda^l_j = 1,\ \lambda^l_j \geq 0\}$ and G_l is the measure induced by the preceding transformation. If we assume that the measure P is absolutely continuous (with respect to the Lebesgue measure on $\mathcal{R}^{m_2}$), then P assigns zero measure to every face (of dimension less than m_2) of the simplices $\mathcal{S}^l$ and hence

(2.11) $$\mathcal{Q}(x) = \Sigma_l \int_{\mathcal{S}^l} Q(x,\xi)\,P(d\xi) \leq \Sigma_l \bar{Q}_l(x,\mathcal{S}^l).$$

This new upper bound can again be refined in two ways, first by considering finer simplicial decompositions, and second by considering for every ξ the smallest upper bound given by a number of possible simplicial representations. We sketch this out.

Suppose Ξ is a convex polytope (of dimension m_2) and $\{v^1,\ \dots\ ,v^l\}$ is a finite collection of points in Ξ that includes the extreme points of Ξ. Let $\mathcal{P}$ be the set of all (m_2+1) subsets of $\{v^1,\ \dots\ ,v^l\}$ such that $\mathrm{co}(v^{j_0},\ \dots\ ,v^{j_{m_2}})$ is a m_2-simplex. The convexity of $Q(x,\cdot)$ yields

$$Q(x,\xi) \leq \Sigma_{i=0}^{m_2}\lambda_{j_i} Q(x,v^{j_i})$$

where

$$\Sigma_{i=0}^{m_2}\lambda_{j_i} v^{j_i} = \xi,\ \Sigma_{i=0}^{m_2}\lambda_{j_i} = 1,\ \lambda_{j_i} \geq 0,$$

i.e. $\xi \in \mathrm{co}(v^{j_0},\ \dots\ ,v^{j_{m_2}})$. With $\mathcal{P}(\xi)$ denoting the elements of $\mathcal{P}$ that have ξ in their convex hull,

$$Q(x,\xi) \leq \inf_{\{(v^{j_0},\ \dots\ ,v^{j_{m_2}})\in \mathcal{P}(\xi)\,|\,\Sigma_{i=0}^{m_2}\lambda_{j_i}v^{j_i}=\xi\}} \Sigma_{i=0}^{m_2}\lambda_{j_i} Q(x,v^{j_i}).$$

Each element of $\mathcal{P}(\xi)$ induces a measure on Λ, we can integrate on both sides to obtain an upper bound on $\mathcal{Q}$ and thus also on z^*.

3. Getting a Starting Solution. The inequalities, and thus the resulting error bounds, presented above depend upon the chosen sample points of Ξ or the partitioning scheme used. Choices for initial samples can be based on the solutions of simplified problems in which the constraints have been relaxed. It is convenient to use here the version (0.7)–(0.8)–(0.9) of the original problem. We shall assume that we are dealing with stochastic programs with relatively complete recourse ($K = K_1$). In terms of (0.7) this means that if $x \in K_1$ and $\chi = Tx$, then $\chi = L_2$, cf. the expression for L_2 following (0.9).

Suppose χ^0 is a guess at the optimal tender, i.e. as part of a pair (x^0, χ^0) solving (0.7). Cost considerations might lead us to such a choice, but there is no guarantee that χ^0 is actually part of a feasible pair for problem (0.7), that we repeat here for convenience sake:

(0.7) find $x \in \mathscr{R}_+^{n_2}$, $\chi \in \mathscr{R}^{m_2}$ such that $Ax = b$, $Tx = \chi$, and $z = cx + \Psi(\chi)$ is minimized.

To obtain a feasible solution we might solve the linear program (with $h^+ \geqslant 0$, $h^- \geqslant 0$)

(3.1) find $x \in \mathscr{R}_+^{n_1}$, $u^+ \in \mathscr{R}_+^{m_2}$, $u^- \in \mathscr{R}_-^{m_2}$ such that $Ax = b$, $Tx + u^+ + u^- = \chi^0$, and $z = cx + h^+u^+ - h^-u^-$ is minimized.

We can use the resulting solution to start the optimization algorithm. In the case of simple recourse, a suitable choice of h^+ and h^- may be the vectors q^+ and q^- that determine the recourse costs. Recall that for stochastic programs with simple recourse, the function ψ as defined by (0.9) is given by $\psi(\chi, \xi) = \sum_{i=1}^{m_2} \psi_i(\chi_i, \xi_i)$ and

$$\psi_i(\chi_i, \xi_i) = \inf\{q_i^+ y_i^+ + q_i^- y_i^- \mid y_i^+ - y_i^- = \xi_i - \chi_i, y_i^+ \geqslant 0, y_i^- \geqslant 0\},$$
$$= \begin{cases} q_i^+(\xi_i - \chi_i) & \text{if } \chi_i \leqslant \xi_i, \\ q_i^-(\chi_i - \xi_i) & \text{if } \chi_i \geqslant \xi_i. \end{cases}$$

In this situation, we could proceed as follows: for every $i = 1, \ldots, m_2$, solve the single constraint stochastic program

(3.2) find $x \in \mathscr{R}_+^{n_1}$, $\chi_i \in \mathscr{R}$ such that $T_i x = \chi_i$, and $z_i = cx + \Psi_i(\chi_i)$ is minimized.

Here T_i is the i-th row of T and

$$\Psi_i(\chi_i) = E\{\Psi_i(\chi_i, \xi_i)\}.$$

This problem is equivalent to

(3.3) find $x \in \mathscr{R}_+^{n_1}$, $\chi_i \in \mathscr{R}$ such that $\chi_i = T_i x$, and

$$z_i = cx + \int_{\xi_i \geqslant \chi_i} q_i^+(\xi_i - \chi_i) F_i(d\xi_i) + \int_{\xi_i \leqslant \chi_i} q_i^-(\chi_i - \xi_i) F_i(d\xi_i)$$

with F_i denoting the marginal distribution function of ξ_i. The optimal solution of (3.2) is the pair (x^0, χ_i^0) such that

$$x_j^0 \geqslant 0 \text{ for } j = 1, \ldots, n,$$
$$\sum_{j=1}^{n_1} t_{ij} x_j^0 = \chi_i^0,$$
$$\theta \in -\partial\psi_i(\chi_i^0) = [q_i^+ - q_i F_i^+(\chi_i^0), q_i^+ - q_i F_i \chi_i^0)],$$
$$c_j - \theta t_{ij} \geqslant 0 \quad \text{for } j = 1, \ldots, n,$$
$$(c_j - \theta t_{ij}) x_j = 0 \quad \text{for } j = 1, \ldots, n,$$

where $q_i = q_i^+ + q_i^-$, $F_i(z) = P[\xi_i < z]$, and $F_i^+(z) = P[\xi_i \leqslant z]$.

In order to simplify the presentation, we make the following assumptions:

(i) F_i is strictly continuously increasing on its support,

(ii) $T_i \geqslant 0$,

(iii) $\inf_j c_j t_{ij}^{-1} \in [-q_i^-, q_i^+]$.

The last assumption is only introduced to render the problem nontrivial. Without such a condition the problem is either unbounded or of a type that has no practical interest. With this, we have

$$\theta = \inf_j(c_j/t_{ij}) = c_s/t_{is},$$
$$\chi_i = F_i^{-1}((q_i^+ - c_s/t_{is})/q_i).$$

This method gives us a starting vector χ^0, which we can then use to generate a feasible pair $(\hat{x}, \hat{\chi})$, as indicated at the beginning of this section. Some justification for this choice comes from the fact that we are solving for each i the problem "optimally". This boils down to finding the solution to a newsboy problem (having more than one supply source). For a detailed study of this class of problems, when viewed as simple stochastic programs, consult Wets (1974).

If we are not dealing with simple recourse we may still proceed in a very similar manner. For each i, the problem to be solved is

(3.4) find $x \in \mathscr{R}_+^{n_1}$, $\chi_i \in \mathscr{R}$ such that $T_i x = \chi_i$ and $cx + \int_{\Xi_i} \inf\{qy | W_i y = \xi_i - \chi_i\} dP_i(\xi_i)$ is minimized.

Here again P_i is the marginal distribution of ξ_i and $\Xi_i \subset \mathscr{R}$ its support. We note that the integrand above is

$$\int_{\xi_i < \chi_i} (q_j/w_{ij})_{\min}(\xi_i - \chi_i) dP_i(\xi_i) + \int_{\xi_i \geq \chi_i} (q_j/w_{ij})_{\max}(\xi_i - \chi_i) dP_i(\xi_i),$$

assuming here that

$$(q_j/w_{ij})_{\min} = \inf_{1 \leq j \leq n_2}(q_j/w_{ij}),$$
$$(q_j/w_{ij})_{\max} = \sup_{1 \leq j \leq n_2}(q_j/w_{ij}),$$

and the coefficients w_{ij} appearing in $(q_j/w_{ij})_{\min}$ and $(q_j/w_{ij})_{\max}$ are negative and positive respectively. The infimum in (3.4) then occurs at a point such that

$$0 = (\partial(cx)/\partial\chi_i) - ((q_j/w_{ij})_{\max} + ((q_j/w_{ij})_{\min})F(\chi_i) + (q_j/w_{ij})_{\max}$$

If we restrict χ_i to $\chi_i = t_{ij}x_j$ for fixed j,

$$\chi_i^0 \in \text{argmin}_j[(c_j/t_{ij})\chi_{ij} + \int_{\Xi_i} \inf\{qy | W_i y = \xi_i - \chi_{ij}\} dP_i(\xi_i)]$$

where

$$\chi_{ij} = F^{-1}(((q_j/w_{ij})_{\max} - (c_j/t_{ij}))/((q_j/w_{ij})_{\max} + (q_j/w_{ij})_{\min})).$$

Again this leads us to a vector χ^0. The intuitive justification for the use of this vector is the same as in the case of stochastic programs with simple recourse.

After the intial choice of χ^0, other values of χ may be chosen by minimizing the expected error in approximating the function $\Psi(\chi)$, by using an a priori distribution on χ. As new χ values are found in an optimization procedure, this distribution may be changed using Bayesian updates; in the case of simple recourse the expected error is easily measurable since $\psi(\chi)$ can be evaluated precisely on each subregion.

REFERENCES

Birge, J. (1980). Solution methods for stochastic dynamic linear programs, *Technical Report No. 80–29*, Systems Optimization Laboratory, Stanford University, Stanford, CA.

Birge, J. (1982). The value of the stochastic solution in stochastic linear programs with fixed recourse. *Mathematical Programming 24* 314–325.

Huang, C., Ziemba, W., and Ben-Tal, A. (1977). Bounds on the expectation of a convex function of a random variable: with applications to stochastic programming. *Operations Research 25* 315–325.

Kall, P. (1974). Approximations to stochastic programming with complete fixed recourse. *Num. Math 22* 333–339.

Kall, P. (1979). Computational methods for solving two-stage stochastic linear programming problems. *ZAMP 30* 261–271.

KALL, P. and STOYAN, D. (1982). Solving stochastic programming problems with recourse, including error bounds. *Math. Operationsforschung Statist. Ser. Optimization 13* 431–447.

MANDANSKY, A. (1960). Inequalities for stochastic linear programming problems. *Management Science 6* 197–204.

PERLMAN, M. (1974). Jensen's inequality for a convex vector-valued function on an infinite dimensional space. *J. Multivariate Analysis 4* 52–65.

WALKUP, D. and WETS, R. (1967). Stochastic programs with recourse. *SIAM J. Applied Math. 15* 1299–1314.

WETS, R. (1974). *Stochastic programming,* Lecture Notes.

WETS, R. (1983). Stochastic programming: approximation schemes and solution techniques, in *Mathematical Programming 1982: The State-of-the-Art*. Springer-Verlag, Berlin and New York.

Inequalities in Statistics and Probability
IMS Lecture Notes–Monograph Series Vol. 5 (1984), 187–192

COMPARING COHERENT SYSTEMS

BY HENRY W. BLOCK[1] AND WAGNER DE SOUZA BORGES[2]
University of Pittsburgh and Universidade de São Paulo

It is a well known engineering principle that "redundancy at the component level is more effective than redundancy at the system level." Here, redundancy simply means components are connected in parallel and the principle results from comparing the systems obtained when this parallel protocol is applied both at the component and systems levels. It is shown in this paper that if parallel or series protocols are ruled out, corresponding versions of the above principle are not possible. This question is examined both in structural as well as in reliability (stochastic) terms.

1. Introduction. Let $S = \{0,1, \ldots , m\}$ denote the set of all possible states of both the system and its components, and let $C = \{1, \ldots , n\}$ be the component set. The vector $\mathbf{x} = (x_1, \ldots , x_n) \in S^n$ represents the situation where components $1, \ldots , n$ are in states $x_1, \ldots , x_n$ respectively. In particular we write $\mathbf{k} = (k, \ldots , k)$ for $k \in S$.

The state of the system is a function of the component state vector $\mathbf{x} \in S^n$. A function $\phi: S^n \to S$ is called a multistate system structure (MSS) of order n provided it is nondecreasing, i.e. $\phi(\mathbf{x}) \leq \phi(\mathbf{y})$ whenever $x_i \leq y_i$ for all $i \in C$ ($\mathbf{x} \leq \mathbf{y}$).

We also use throughout the paper the following notational convention.

Notation 1.1. For $\mathbf{x}_i = (x_{i1}, \ldots , x_{in}) \in \mathcal{R}^n$, $i = 1, \ldots , k$ and $\psi: \mathcal{R}^k \to \mathcal{R}$ we let

$$\psi(\mathbf{x}_1, \ldots , \mathbf{x}_k) = (\psi(x_{11},x_{21}, \ldots , x_{k1}), \ldots , \psi(x_{1n},x_{2n}, \ldots , x_{kn})) \in \mathbf{R}^n. \tag{1.1}$$

Note that ϕ is an MSS of order n if and only if

$$\phi(\max_{1 \leq i \leq k}\mathbf{x}_i) \geq \max_{1 \leq i \leq k} \phi(\mathbf{x}_i) \text{ for all } \mathbf{x}_1, \ldots , \mathbf{x}_k \in S^n \text{ and } k \geq 2, \tag{1.2}$$

or equivalently

$$\phi(\min_{1 \leq i \leq k}\mathbf{x}_i) \leq \min_{1 \leq i \leq k} \phi(\mathbf{x}_i) \text{ for all } \mathbf{x}_1, \ldots , \mathbf{x}_k \in S^n \text{ and } k \geq 2, \tag{1.3}$$

where $\max_{1 \leq i \leq k}\mathbf{x}_i (\min_{1 \leq i \leq k}\mathbf{x}_i)$ is the vector of coordinatewise maximums (minimums). Inequality (1.2) expresses mathematically a well known engineering principle that states that "redundancy at the component level is more effective than redundancy at the system level", and (1.3) expresses a related dual principle. These principles are presented in their simplest form in Barlow and Proschan (1975).

We recall that the MSS of order k defined by $\psi(\mathbf{x}) = \max_{1 \leq i \leq k} x_i$ ($\psi(x) = \min_{1 \leq i \leq k} x_i$) for $\mathbf{x} \in S^k$ is called a parallel (series) system and note that using (1.1) the principle expressed by (1.2) ((1.3)) can be rewritten as follows. We express it in this form for ease in describing our subsequent results.

Principle 1.2. If ϕ is an MSS of order n and ψ is a parallel (series) system of order k, then the MSS of order $k \times n$ defined by

[1] Supported by ONR Contract N00014–76–C–0839.
[2] Supported in part by Conselho Nacional de Desenvolvimento Científico e Tecnológico (CNPq), processo n°. 200175–81.
AMS 1980 subject classifications. Primary 62N05; Secondary 60K10.
Key words and phrases: Coherent systems, redundancy, reliability.

$$\phi(\psi(\mathbf{x}_1, \ldots, \mathbf{x}_k)) \quad \text{for } \mathbf{x}_1, \ldots, \mathbf{x}_k \in S^n \tag{1.4}$$

is uniformly better (worse) than the MSS of order $k \times n$ defined by

$$\psi(\phi(\mathbf{x}_1), \ldots, \phi(\mathbf{x}_k)) \quad \text{for } \mathbf{x}_1, \ldots, \mathbf{x}_k \in S^n \tag{1.5}$$

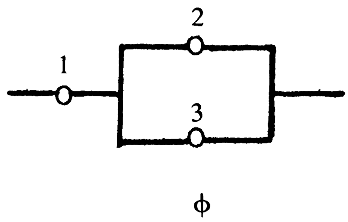
In this paper, we will consider the question of which of the two MSS's of order $k \times n$ defined in (1.4) and (1.5) for general ϕ and ψ is uniformly better. As an example to better visualize the two competing alternatives, assume that

$$\phi(x_1,x_2,x_3) = \min\{x_1,\max\{x_2,x_3\}\}$$

and

$$\psi(y_1,y_2,y_3,y_4) = \max\{y_1,\min\{y_1,y_2,y_3\}\}$$

for $x_i,x_j \in \{0,1\}$, $i = 1,2,3$, $j = 1,2,3,4$. Since ϕ and ψ can be represented respectively as



and

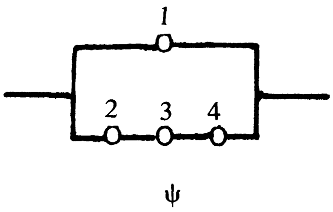


FIGURE 1.1.

the two alternatives are to build either the system illustrated in Figure 1.2 or in Figure 1.3.

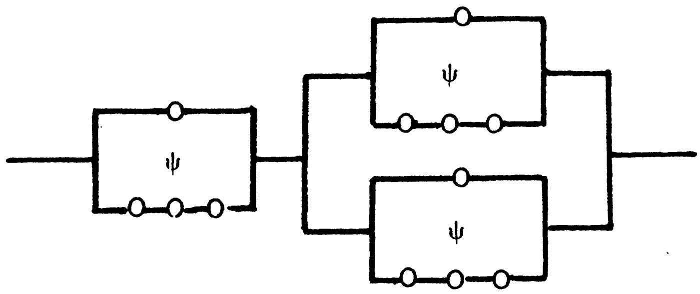


FIGURE 1.2.

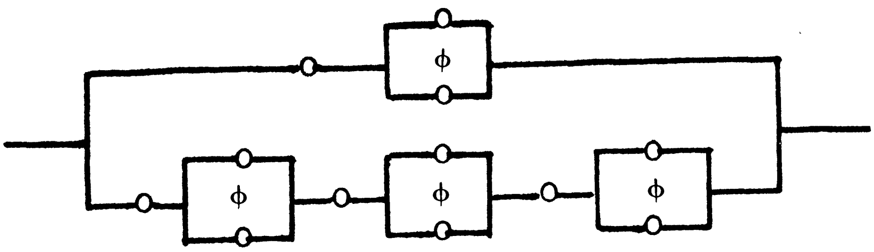


FIGURE 1.3.

The solution to this problem in the binary setting, i.e. when $S = \{0,1\}$ is given in Section 2. This is that if series and parallel systems are ruled out, neither of the resulting systems is uniformly better than the other. This is our main result which is given by Lemma 2.1. An interesting consequence of this result is given in Theorem 2.2.

In section 3 we consider the problem in the multistate setting and a weaker result is given in Proposition 3.1. An example is given to show that this cannot be improved upon in general but, if the specialized type of MSS of Barlow and Wu (1978) is considered, a direct analog of the binary result is obtainable. This is given in Proposition 3.2.

Finally in Section 4 we comment on the possibility of obtaining stochastic versions of the results given in the previous section. It is shown that even in the binary case only weak results can be achieved.

2. Binary System Structures. In this section we consider the binary setting where $S = \{0,1\}$ in which case an MSS is called a binary system structure (BSS). We assume that any BSS ϕ of order n considered here is coherent in the sense that for each $i \in C$ there is $\mathbf{x} \in \{0,1\}^n$ such that

$$\phi(x_1, \ldots, x_{i-1}, 0, x_{i+1}, \ldots, x_n) < \phi(x_1, \ldots, x_{i-1}, 1 x_{i+1}, \ldots, x_n). \tag{2.1}$$

We also recall from (3.6) of Chapter 1 of Barlow and Proschan (1975) that if ϕ is a coherent BSS of order n, then ϕ has a representation of the form

$$\phi(\mathbf{x}) = \min_{i \leqslant j \leqslant k} \max_{i \in K_j} x_i$$

where $\bigcup_{j=1}^{k} K_j = C$ and for all $i \neq j$, K_i is not a subset of K_j. These sets are called the min cut sets of ϕ and we refer the reader to Barlow and Proschan (1975) for properties of min cut sets and related notions.

Our main result will be a consequence of the following lemma.

LEMMA 2.1. *Let ϕ and ψ be coherent BSS's of orders $n \geqslant 2$ and $k \geqslant 2$, respectively.* (1) *If ϕ is not a parallel system and ψ is not a series system then there exist $\mathbf{x}_1, \ldots, \mathbf{x}_k$ in $\{0,1\}^n$ such that*

$$\phi(\psi(\mathbf{x}_1, \ldots, \mathbf{x}_k)) > \psi(\phi(\mathbf{x}_1), \ldots, \phi(\mathbf{x}_k)).$$

(2) *If ϕ is not a series system and ψ is not a parallel system then there exist $\mathbf{x}_1, \ldots, \mathbf{x}_k$ in $\{0,1\}^n$ such that*

$$\phi(\psi(\mathbf{x}_1, \ldots, \mathbf{x}_k)) < \psi(\phi(\mathbf{x}_1), \ldots, \phi(\mathbf{x}_k)).$$

Proof. 1) We will construct $\mathbf{x}_1, \ldots, \mathbf{x}_k$ in $\{0,1\}^n$ such that the desired inequality holds. Since ψ is not series, $k \geqslant 2$, and ψ is coherent we can find a min cut set K^{ψ} which contains at least two elements. Furthermore since ϕ is not parallel, $n \geqslant 2$, and ϕ is coherent there are at least two different min cut sets of ϕ; call them K_1^{ϕ} and K_2^{ϕ}. Now for each $i \in K^{\psi}$, choose K_1^{ϕ} or K_2^{ϕ} and construct $\mathbf{x}_i = (x_{i1}, \ldots, x_{in}) \in \{0,1\}^n$ defining $x_{ij} = 0$ if $j \in K^{\phi}$ (where K^{ϕ} is whichever one of K_1^{ϕ} or K_2^{ϕ} was chosen) and $x_{ij} = 1$ otherwise. Also construct $\mathbf{x}_i$ for $i \in K^{\psi}$ so that not all of them are associated with only one of K_1^{ϕ} or K_2^{ϕ}. For $i \notin K^{\psi}$ define $\mathbf{x}_i = \mathbf{1} = (1, \ldots, 1)$. Thus $(\phi(\mathbf{x}_1), \ldots, \phi(\mathbf{x}_k)) \in \{0,1\}^k$ has zeros for all the components $i \in K^{\psi}$ so that $\psi(\phi(\mathbf{x}_1), \ldots, \phi(\mathbf{x}_k)) = 0$. On the other hand $x_{ij} = 1$ for all $i \notin K^{\psi}$ so that $A = \{j\colon \psi(x_{1j}, x_{2j}, \ldots, x_{kj}) = 0\} = \{j;\, x_{ij} = 0 \text{ for all } i \in K^{\psi}\} = K_1^{\phi} \cap K_2^{\phi}$. But since K_1^{ϕ} and K_2^{ϕ} are different min cut sets, $K_1^{\phi} \cap K_2^{\phi}$ must be strictly contained in K_1^{ϕ} and K_2^{ϕ}. Thus A does not contain any min cut set of ϕ. Consequently $\phi(\psi(\mathbf{x}_1, \ldots, \mathbf{x}_k)) = 1$.

2) The second part of the lemma is proven similarly. □

The main result now follows easily.

THEOREM 2.2. *Let ϕ and ψ be coherent BSS's of orders $n \geqslant 2$ and $k \geqslant 2$* respectively. Then

$$\phi(\psi(\mathbf{x}_1, \ldots, \mathbf{x}_k)) = \psi(\phi(\mathbf{x}_1), \ldots, \phi(\mathbf{x}_k))$$

for all $\mathbf{x}_i \in \{0,1\}^n$, $i = 1, \ldots, k$, *if and only if* ϕ *and* ψ *are both parallel or both series.*

Proof. If the equality holds, it follows from Lemma 2.1 that: (i) either ϕ is parallel or ψ is series; and (ii) either ϕ is series or ψ is parallel. Combining (i) and (ii) we have that either ϕ and ψ are series or ϕ and ψ are parallel. Necessity of the equality follows immediately. □

Note 2.3. In proving Theorem 2.2 we used the contrapositive form of the two statements in Lemma 2.1. These results are that under the assumptions of the lemma: (i) If

$$\phi(\psi(\mathbf{x}_1, \ldots, \mathbf{x}_k) \leq \psi(\phi(\mathbf{x}_1), \ldots, \phi(\mathbf{x}_k)) \text{ for all } \mathbf{x}_1, \ldots, \mathbf{x}_k \in \{0,1\}^n,$$

then either ϕ is parallel or ψ is series. (ii) If

$$\phi(\psi(\mathbf{x}_1, \ldots, \mathbf{x}_k) \geq \psi(\phi(\mathbf{x}_1), \ldots, \phi(\mathbf{x}_k)) \text{ for all } \mathbf{x}_1, \ldots, \mathbf{x}_k \in \{0,1\}^n,$$

then either ϕ is series or ψ is parallel.

It is easy to show that the converses of (i) and (ii) above also hold.

As a special case the result of Theorem 2.4 of Chapter 1 of Barlow and Proschan (1975) follows from Note 2.3.

3. Multistate System Structures. We now examine the extent to which the results in the previous section can be generalized to the case of multistate system structures. Any MSS ϕ considered in this section will further satisfy the following two condtions: (i) $\phi(\mathbf{k}) = k$ for all $k \in S$; and (ii) for each $i \in C$ and $j \geq 1$ there exists $\mathbf{x} \in S^n$ such that

$$\phi(x_1, \ldots, x_{i-1}, j-1, x_{i+1}, \ldots, x_n) < \phi(x_1, \ldots, x_{i-1}, j, x_{i-1}, \ldots, x_n).$$

These will be called coherent MSS's of order n. This last concept coincides with the middle and most reasonable multistate concept of coherence discussed in Griffith (1980).

A full generalization of Theorem 2.2 is not possible in the multi-state case even under fairly strong conditions. We give however some weaker results and an instructive counterexample.

The first result is in the spirit of the remarks in Note 2.3.

Proposition 3.1. *Let* $\phi(\psi)$ *be a coherent MSS of order* n (k). Then

$$\phi(\psi(\mathbf{x}_1, \ldots, \mathbf{x}_k)) \leq \psi(\phi(\mathbf{x}_1), \ldots, \phi(\mathbf{x}_k)) \text{ for all } \mathbf{x}_1, \ldots, \mathbf{x}_k \in S^n$$

and all coherent MSS ψ (ϕ) *of order* $k(n)$, *if and only if* $\phi(\psi)$ *is a parallel (series) MSS.*

Proof. The "if" part is straightforward. For the "only if" let $\psi(\mathbf{x}) = \max_{1 \leq i \leq k} x_i$ for $\mathbf{x} \in S^k$. Obviously for this choice of ψ the reverse inequality holds. Thus

$$\phi(\max(\mathbf{x}_i, \ldots, \mathbf{x}_k)) = \max_{1 \leq i \leq k} \phi(\mathbf{x}_i).$$

By the same proof as that of Proposition 2.2 of Griffith (1980) the result follows. The proof of the dual result is similar. □

The following example shows that the generalization of Theorem 2.2 (and Note 2.3) is false even under stronger coherence assumptions.

Example 3.2. Let ϕ and ψ be identical MSS's defined as follows:

$$\phi(0,0) = \phi(0,1) = \phi(1,0) = \phi(0,2) = \phi(2,0) = 0,$$

$$\phi(1,1) = 1 \text{ and } \phi(1,2) = \phi(2,1) = \phi(2,2) = 2.$$

Then it is not hard to see that $\phi(\psi(\mathbf{x}_1,\mathbf{x}_2) = \psi(\phi(\mathbf{x}_1), \phi(\mathbf{x}_2))$ for all $\mathbf{x}_1,\mathbf{x}_2 \in \{0,1,2\}^2$. Moreover ϕ and ψ are coherent and even satisfy the strong coherence assumption of Griffith (1980). However neither ϕ nor ψ are either series or parallel.

If we consider the more restrictive multistate system structures proposed by Barlow and Wu (1978) we can obtain an extension of Theorem 2.2. An MMS ϕ of order n is of the type proposed by Barlow and Wu (1978) (BW-MSS) if it is of the form

$$\phi(\mathbf{x}) = \min_{1 \leqslant j \leqslant k} \max_{i \in K_j} x_i \quad \text{for } \mathbf{x} \in S^n,$$

where $\bigcup_{j=1}^{k} K_j = C$ and for $i \neq j$, K_i is not a subset of K_j. These functions are a particular subfamily of the coherent MSS's. Moreover for x_i binary, ϕ is a coherent BSS with min cut sets $K_1, \ldots, K_k$.

PROPOSITION 3.3. *Let ϕ and ψ be BW-MSS's of order $n \geqslant 2$ and $k \geqslant 2$, respectively. Then*

$$\phi(\psi(\mathbf{x}_1, \ldots, \mathbf{x}_k)) = \psi(\phi(\mathbf{x}_1), \ldots, \phi(\mathbf{x}_k)) \quad \text{for all } \mathbf{x}_1, \ldots, \mathbf{x}_k \in S^n$$

if and only if ϕ and ψ are both parallel or both series. The result remains true if equality holds for all $\mathbf{x}_1, \ldots, \mathbf{x}_k \in \{k_1, k_2\}^n$, where $0 \leqslant k_1 < k_2 \leqslant m$.

Proof. We need only show the result for the weaker assumption. As mentioned above ϕ and ψ reduce to coherent BSS's when restricted to $\{0,1\}^n$ and $\{0,1\}^k$, respectively. We consider $f(x) = (k_2 - k_1)^{-1}(x - k_1)$ so that when $x \in \{k_1, k_2\}$, $f(x) \in \{0,1\}$.

To prove sufficiency note that

$$\phi(\psi(f(\mathbf{x}_1), \ldots, f(\mathbf{x}_k))) = \phi(f(\psi(\mathbf{x}_1, \ldots, \mathbf{x}_k)))$$
$$= f(\phi(\psi(f(\mathbf{x}_1, \ldots, \mathbf{x}_k))) = f(\psi(\phi(\mathbf{x}_1), \ldots, \phi(\mathbf{x}_k)))$$
$$= \psi(f(\phi(\mathbf{x}_1)), \ldots, f(\phi(\mathbf{x}_k))) = \psi(\phi(f(\mathbf{x}_1)), \ldots, \phi(f(\mathbf{x}_k))).$$

Hence,

$$\phi(\psi(\mathbf{y}_1, \ldots, \mathbf{y}_k)) = \psi(\phi(\mathbf{y}_1), \ldots, \phi(\mathbf{y}_k)) \quad \text{for all } \mathbf{y}_1, \ldots, \mathbf{y}_k \in \{0,1\}^n,$$

and the result follows from Theorem 2.2.

Note 3.4. By similar methods, analogs of Lemma 2.1 and Note 2.3 can also be given for BW-MSS's.

4. Further Remarks. Stochastic versions of the results of the previous sections do not necessarily hold even in the binary setting. However, an analog of Proposition 3.1 can be obtained. We consider only the binary case although similar results hold in the multistate case.

We let ϕ and ψ be coherent BSS's of orders n and k respectively, and compare the reliability functions,

$$h_{\phi(\psi)}(\mathbf{P}_1, \ldots, \mathbf{P}_k) = E\, \phi(\psi(\mathbf{X}_1, \ldots, \mathbf{X}_k))$$

and

$$h_{\psi(\phi)}(\mathbf{P}_1, \ldots, \mathbf{P}_k) = E\, \psi(\phi(\mathbf{X}_1), \ldots, \phi(\mathbf{X}_k)),$$

of the two competing coherent BSS's of order $k \times n$ defined in (1.4), (1.5). Here $\mathbf{X}_i = (X_{i1}, \ldots, X_{in})$ for $i = 1, \ldots, k$ are independent random vectors of independent binary random variables, and $\mathbf{P}_i = (p_{i1}, \ldots, p_{in})$ for $i = 1, \ldots, k$ are defined by $p_{ij} = P\{X_{ij} = 1\}$.

We also let $h_\phi(p_1, \ldots, p_n) = E\, \phi(X_1, \ldots, X_n)$ $(h_\psi(q_1, \ldots, q_k) = \mathrm{E}\, \psi(Y_1, \ldots, Y_k))$ when $X_1, \ldots, X_n$ $(Y_1, \ldots, Y_k)$ are independent binary random variables and $p_i = P\{X_i = 1\}$ for $i = 1, \ldots, n$ $(q_j = P\{Y_j = 1\}$ for $j = 1, \ldots, k)$.

PROPOSITION 4.1. (1) *If ϕ is a parallel (series) BSS, then*

$$h_{\phi(\psi)}(\mathbf{P}_1, \ldots, \mathbf{P}_k) = h_\phi(h_\psi(p_{11}, \ldots, p_{k1}), \ldots, h_\psi(p_{1n}, \ldots, p_{kn})) \tag{4.1}$$

$$\leq (\geq) h_{\psi}(h_{\phi}(p_{11}, \ldots, p_{1n}), \ldots, h_{\phi}(p_{k1}, \ldots, p_{kn}))$$
$$= h_{\psi(\phi)}(\mathbf{P}_1, \ldots, \mathbf{P}_k)$$

for all $\mathbf{P}_1, \ldots, \mathbf{P}_k \in [0,1]^n$ *and any coherent BSS* ψ. (2) *Conversely if for any coherent BSS* ψ *and some* $\mathbf{P}_1, \ldots, \mathbf{P}_k \in (0,1)^n$, *inequality (4.1) holds, then* ϕ *is a parallel (series) BSS.*

Proof. (1) Follows from Proposition 3.1 by taking expectations. (2) Taking $\psi(\mathbf{x}) = \max_{1 \leq i \leq k} x_i$ for $\mathbf{x} \in \{0,1\}^k$, we have

$$h_{\phi(\psi)}(\mathbf{P}_1, \ldots, \mathbf{P}_k) - h_{\psi(\phi)}(\mathbf{P}_1, \ldots, \mathbf{P}_k)$$
$$= E[\phi(\psi(\mathbf{X}_1, \ldots, \mathbf{X}_k)) - \psi(\phi(\mathbf{X}_1), \ldots, \phi(\mathbf{X}_k))]$$
$$= E[\phi(\max_{1 \leq i \leq k} \mathbf{X}_i) - \max_{1 \leq i \leq k} \phi(\mathbf{X}_1)] \leq 0.$$

Hence,

$$\phi(\max_{1 \leq i \leq k} \mathbf{x}_i) = \max_{1 \leq i \leq k} \phi(\mathbf{x}_i) \quad \text{for all } \mathbf{x}_1, \ldots, \mathbf{x}_k \in \{0,1\}^n$$

and from Theorem 2.3, Chapter 1 of Barlow and Proschan (1975) ϕ must be a parallel BSS.

The dual statement is proved similarly. □

It is easy to check that ϕ and ψ are both parallel or series BSS's if and only if we have equality in (4.1) for all $\mathbf{P}_1, \ldots, \mathbf{P}_k \in [0,1]^n$. It is not true however that if equality holds in (4.1) for some $\mathbf{P}_1, \ldots, \mathbf{P}_k \in (0,1)^n$ then ϕ and ψ are both parallel or series BSS's. An example of this last fact can be constructed by simply taking ϕ and ψ identical, but neither being a parallel or series BSS, and taking $\mathbf{P}_1, \ldots, \mathbf{P}_k \in (0,1)^n$ ($k=n$) such that $p_{ij} = p_{ji}$ for all $i,j = 1, \ldots, n$. It is obvious that this construction provides equality in (4.1). It is also easy to show that if equality holds in (4.1) for some $\mathbf{P}_1, \ldots, \mathbf{P}_k \in (0,1)^n$ and either ϕ or ψ is parallel (series) then so is ψ or ϕ.

REFERENCES

BARLOW, R. E. and PROSCHAN, F. (1975). *Statistical Theory of Reliability and Life Testing*, Holt, Rinehart and Winston, New York.

BARLOW, R. E. and WU, A. (1978). Coherent Systems with Multistate Components, *Math. Oper. Res.* *3* 275–281.

GRIFFITH, W. E. (1980). Multistate Reliability Models, *J. Appl. Prob.* *17* 735–744.

Inequalities in Statistics and Probability
IMS Lecture Notes–Monograph Series Vol. 5 (1984), 193–198

MULTIVARIATE LIFE CLASSES AND INEQUALITIES

BY THOMAS H. SAVITS[1]
University of Pittsburgh

In this paper we review some univariate life classes which are useful in reliability theory. Recently some new characterizations of these classes have been given in terms of integral inequalities with respect to certain classes of function. These characterizations and their natural multivariate extensions are discussed. Some moment inequalities are then deduced.

1. Introduction. Various univariate classes of life distributions have been introduced in the context of mathematical reliability theory. Most of these classes have intuitive appeal, possess nice closure properties and lead to useful bound in estimating system reliability. The book by Barlow and Proschan (1975) gives an excellent discussion of these classes and their properties.

Recently there has been much interest in obtaining multivariate versions of these classes. Although there have been many different approaches, this review paper will focus on only three: the multivariate IFR class of Savits (1983); the multivariate IFRA class of Block and Savits (1980); the multivariate NBU class of Marshall and Shaked (1982). All three are based on recent characterizations which are expressable in terms of integral inequalities for certain classes of functions. Also, more importantly, all three classes possess many desirable closure properties.

All functions and sets in this paper are assumed to be Borel measurable. A subset A is said to be an upper set if $\mathbf{x} \in A$ and $\mathbf{y} \geqslant \mathbf{x}$ implies that $\mathbf{y} \in A$. A nonnegative function h is said to be log concave (on $\mathscr{R}_+^n$) if $h[\lambda\mathbf{x}+(1-\lambda)\mathbf{y}] \geqslant h^\lambda(\mathbf{x})h^{1-\lambda}(\mathbf{y})$ for all $\mathbf{x},\mathbf{y} \geqslant \mathbf{0}$ and all $0 < \lambda < 1$. A function ψ is said to be subhomogeneous (on $\mathscr{R}_+^n$) if $\psi(\alpha\, \mathbf{x}) \geqslant \alpha\psi(\mathbf{x})$ for all $\mathbf{x} \geqslant \mathbf{0}$ and all $0 < \alpha < 1$.

2. Review of Univariate Life Classes. Let T be a nonnegative random variable with survival function $\bar{F}(t) = P\{T>t\}$. Set $b = \inf\{t\geqslant 0; \bar{F}(t) = 0\}$ ($\inf \phi = +\infty$). For simplicity we assume $\bar{F}(0) = 1$.

Definition 1. (i) T is said to have an increasing failure rate (IFR) distribution if $\bar{F}(s+t)/\bar{F}(t)$ is nonincreasing in $t \in [0,b)$ for all $s \geqslant 0$. (ii) T is said to have an increasing failure rate average (IFRA) distribution if $\bar{F}(\alpha t) \geqslant \bar{F}^\alpha(t)$ for all $t \geqslant 0$, $0 < \alpha < 1$. (iii) T is said to have a new better than used (NBU) distribution if $\bar{F}(s+t) \leqslant \bar{F}(s)\bar{F}(t)$ for all $s,t \geqslant 0$. (iv) T is said to have a new better than used in expectation (NBUE) distribution if $\mu = E[T] < \infty$ and $\int_t^\infty \bar{F}(x)dx \leqslant \mu\, \bar{F}(t)$ for all $t \geqslant 0$.

These classes of distribution have been very useful in reliability theory (cf. Barlow and Proschan (1975) for a detailed discussion of their properties). It is known that IFR → IFRA → NBU → NBUE.

---

[1] Supported by ONR Contract N00014–76–C–0839.

AMS 1980 subject classifications. Primary 60K10; Secondary 62N05.

Key words and phrases: multivariate, inequalities, IFR, IFRA, NBU, NBUE, upper sets, log concave functions, subhomogeneous functions, log subhomogeneous functions.

The next two theorems are well known and list some useful equivalent conditions (see Barlow and Proschan (1975)).

THEOREM 1. *The following are equivalent: (a) T is IFR. (b) $\bar{F}$ is a Pólya frequency function of order two (PF_2); i.e., $\bar{F} \geqslant 0$ and*

$$\begin{vmatrix} \bar{F}(t_1-s_1) & \bar{F}(t_1-s_2) \\ \bar{F}(t_2-s_1) & \bar{F}(t_2-s_2) \end{vmatrix} \geqslant 0$$

for all $-\infty < t_1 < t_2 < \infty$, $-\infty < s_1 < s_2 < \infty$. (c) $\log \bar{F}(t)$ *is concave in $t \geqslant 0$.*

THEOREM 2. *The following are equivalent: (a) T is IFRA. (b) $\bar{F}^{1/t}(t)$ is nonincreasing in $t > 0$. (c) $-1/t \log \bar{F}(t)$ is nondecreasing in $t > 0$.*

Remark. If F has a density f, we define the hazard rate by $r(t) = f(t)/\bar{F}(t)$. Then T is IFR if and only if $r(t)$ is nondecreasing in $t \in [0,b)$; T is IFRA if and only if $(1/t)\int_0^t r(u)du$ is nondecreasing in $t \in [0,b)$.

3. Some Recent Characterizations. Within the past several years, various other characterizations of these classes have been discovered. The ones we list below are all expressed via integral inequalities with respect to certain function classes. Many of the known properties of the univariate classes follow easily from these characterizations.

THEOREM 3. *(i) T is IFR if and only if $E[h(\mathbf{x},T)]$ is log concave in $\mathbf{x}$ for all functions $h(\mathbf{x},t)$ which are log concave in $(\mathbf{x},t)$ and nondecreasing in t for each fixed $\mathbf{x}$. (ii) T is IFRA if and only if $E[h(T)] \leqslant E^{1/\alpha}[h^{\alpha}(T/\alpha)]$ for all nonnegative nondecreasing functions h and all $0 < \alpha < 1$. (iii) T is NBU if and only if $E[h(T/(\alpha+\beta))] \leqslant E[h^{\gamma}(T/\alpha)]\, E[h^{1/\gamma}(T/\beta)]$ for all nonnegative nondecreasing functions h, all $\alpha, \beta > 0$ and all $0 < \gamma < 1$.*

The first result is due to Savits (1983); the second to Block and Savits (1976); the third to Marshall and Shaked (1982). The IFRA characterizations given in Theorem 3.1(ii) was particular useful in solving the IFRA convolution problem (Block and Savits (1976)).

Although the original intuitive appeal of the univariate life classes is lost in the abstract characterizations given above, the multivariate extensions that naturally follow enjoy many desirable closure properites, as we shall see in the next section.

4. Multivariate Extensions. There are many different ways of obtaining multivariate extensions of the univariate life classes (e.g., see the review paper of Block and Savits (1981)); however, it is desirable that any such extension satisfy certain properties. We list below some such properties.

Let $\mathcal{C}$ denote a multivariate extension of a univariate class of distributions $\mathcal{C}_0$. By abuse of notation, we say that a random vector $\mathbf{T} \in \mathcal{C}$ if its distribution belongs to $\mathcal{C}$.

Properties.

(P0) A random variable $T \in \mathcal{C}$ if and only if $T \in \mathcal{C}_0$.

(P1) If $\mathbf{T} \in \mathcal{C}$, then all marginals belong to $\mathcal{C}$.

(P2) If $\mathbf{S},\mathbf{T} \in \mathcal{C}$ and are independent, then $(\mathbf{S},\mathbf{T}) \in \mathcal{C}$.

(P3) If $\mathbf{S},\mathbf{T} \in \mathcal{C}$ and are independent, then $\mathbf{S} + \mathbf{T} \in \mathcal{C}$ whenever it makes sense.

(P4) If $\mathbf{T}_n \in \mathcal{C}$ for each n and $\mathbf{T}_n \to \mathbf{T}$ in distribution, then $\mathbf{T} \in \mathcal{C}$.

(P5) If $(T_1, \ldots, T_n) \in \mathcal{C}$ and $a_i \geqslant 0$ $(1 \leqslant i \leqslant n)$, then $(a_1T_1, \ldots, a_nT_n) \in \mathcal{C}$.

(P6) If $(T_1, \ldots, T_n) \in \mathcal{C}$ and π is any permutation on $\{1, \ldots, n\}$, then $(T_{\pi(1)}, \ldots, T_{\pi(n)}) \in \mathcal{C}$.

(P7) If $\mathbf{T} \in \mathcal{E}$ and $\psi_1, \ldots, \psi_m$ are nonnegative nondecreasing subhomogeneous functions, then $(\psi_1(\mathbf{T}), \ldots, \psi_m(\mathbf{T})) \in \mathcal{E}$.

(P8) If $\mathbf{T} \in \mathcal{E}$ and $\psi_1, \ldots, \psi_m$ are nonnegative nondecreasing concave functions, then $(\psi_1(\mathbf{T}), \ldots, \psi_m(\mathbf{T})) \in \mathcal{E}$.

Property (P3) is included since all the univariate life classes described in Section 2 are closed under convolution. For the IFRA and NBU classes, property (P7) is natural since these classes are closed under the formation of coherent systems, which are included within the class of subhomogeneous functions. This is not true for the IFR class, however. On the other hand, the IFR class is closed under minimums and these are special examples of concave functions. Properties (P5) and (P6) are not as essential as the others.

Definition 2. (i) $\mathbf{T}$ is MIFR (in the sense of Savits (1983)) if $E[h(\mathbf{x},\mathbf{T})]$ is log concave in $\mathbf{x}$ for all log concave functions $h(\mathbf{x},\mathbf{t})$ which are nondecreasing and continuous in $\mathbf{t}$ for each $\mathbf{x}$. (ii) $\mathbf{T}$ is MIFRA (in the sense of Block and Savits (1980)) if $E[h(\mathbf{T})] \leqslant E^{1/\alpha}(h^{\alpha}(\mathbf{T}/\alpha)]$ for all continuous nonnegative nondecreasing functions h and all $0 < \alpha < 1$. (iii) $\mathbf{T}$ is MNBU (in the sense of Marshall and Shaked (1982)) if $\mathbf{E}[h(\mathbf{T}/(\alpha+\beta))] \leqslant E[h^{\gamma}(\mathbf{T}/\alpha)]\, E[h^{1-\gamma}(\mathbf{T}/\beta)]$ for all continuous nonnegative nondecreasing functions h and all $\alpha,\beta > 0$, $0<\gamma<1$.

Remark. It is shown in the above papers that the continuity assumption on h is not necessary.

THEOREM 4. *(i) The MIFR class satisfies properties (P0)–(P6) and (P8). (ii) The MIFRA and the MNBU class satisfy properties (P0)–(P7).*

The proofs of this theorem and related results are contained in the above cited papers. In particular, some useful equivalent formulations are given.

If $\mathbf{T}$ is a random vector, let $\mu(d\mathbf{y}) = P(\mathbf{T} \in d\mathbf{y})$ be its induced measure.

THEOREM 5. *(i)* $\mathbf{T}$ *is MIFR if and only if* $\mu[\lambda A+(1-\lambda)B] \geqslant \mu^{\lambda}(A)\mu^{1-\lambda}(B)$ for all upper convex sets A,B and all $0 < \lambda < 1$. *(ii)* $\mathbf{T}$ *is MIFRA iff* $\mu(\alpha A) \geqslant \mu^{\alpha}(A)$ for all upper sets A and all $0 < \alpha < 1$. *(iii)* $\mathbf{T}$ *is MNBU if and only if* $\mu((\alpha+\beta)A) \leqslant \mu(\alpha A)\mu(\beta A)$ *for all upper sets A and all* $\alpha, \beta > 0$.

It is known that MIFRA $\rightarrow$ MNBU, but the implication MIFR $\rightarrow$ MIFRA remains a conjecture.

5. Some Moment Inequalities. Before we consider some multivariate moment inequalities, let us first discuss the univariate case. If T is a nonnegative random variable, let $\mu_r = E[T^r]$ for $r > 0$.

Case (i). T is IFR. We consider functions of the form $h(r,t) = t^r/\phi(r)$ where ϕ is to be suitably chosen. In order to make use of Theorem 3(i) we need that h be log concave in (r,t). If ϕ is twice continuously differentiable, then a necessary and sufficient condition is that

$$r \cdot d/dr[\phi'(r)/\phi(r)] \geqslant 1. \tag{5.1}$$

One can easily check that this is true for $\phi(r) = r^r e^{-r}$. Thus we conclude that the "normalized moments" $\rho_r = \mu_r/(r^r e^{-r})$ are log concave in $r > 0$.

It is interesting to note that (5.1) is true for $\phi(r) = \Gamma(r)$ but false for $\phi(r) = \Gamma(r+1)$, which is the classical normalization factor. The class of ϕ which satisfy (5.1) with inequality replaced by equality is given by $\phi(r) = a\,r^r e^{-br}$ for $a > 0, -\infty < b < \infty$.

Case (ii). T is IFRA. If we let $h(x) = x^r$ in Theorem 3.1(ii), it can be easily shown that $(\rho_r)^{1/r}$ is nonincreasing in $r > 0$, where ρ_r are the same normalized moments given above.

Case (iii). T is NBU. Again letting $h(x) = x^r$ in Theorem 3.1(ii), we conclude that $\rho_{r+s} \leqslant \rho_r\rho_s$ for all $r,s > 0$.

It is convenient at this point to introduce some further definitions. A nonnegative function h is said to be log subhomogeneous (on $\mathcal{R}^n_+$) if $h(\alpha\mathbf{x}) \geqslant h^{\alpha}(\mathbf{x})$ for all $\mathbf{x} \geqslant \mathbf{0}$ and $0 < \alpha < 1$; it is said to be log subadditive (on $\mathcal{R}^n_+$) if $h((\alpha+\beta)\mathbf{x}) \leqslant h(\alpha\mathbf{x})h(\beta\mathbf{x})$ for all $\mathbf{x} \geqslant \mathbf{0}$, and $\alpha, \beta > 0$.

Using these definitions we summarize the above univariate results on ρ_r below.

(5.2)
- (i) T IFR $\rightarrow \rho_r$ is log concave in r.
- (ii) T IFRA $\rightarrow \rho_r$ is log subhomogeneous in r.
- (iii) T NBU $\rightarrow \rho_r$ is log subadditive in r.

A particularly interesting special case of (5.2) is the following. First note that in (5.2) we may replace ρ_r with $\rho^*_r = \rho_r e^{-r}$ since they have exactly the same properties. Now consider an exponential random variable with mean one. In this case $\rho^*_r = \Gamma(r+1)/r^r$. Since the exponential is in all life classes we deduce that

(5.3)
- (i) $\Gamma(r+1)/r^r$ is log concave in r.
- (ii) $\Gamma(r+1)/r^r$ is log subhomogeneous in r.
- (iii) $\Gamma(r+1)/r^r$ is log subadditive in r.

(Actually (i) $\rightarrow$ (ii) $\rightarrow$ (iii) but it is useful to list them separately). The result (5.3)(ii) was already proven in Marshall, Olkin and Proschan (1967), but our proof is much simpler.

We now contrast the results in (5.2) with the known classical univariate results given in Barlow and Proschan (1975). Let $\lambda_r = \mu_r/\Gamma(r+1)$. Then:

(5.4)
- (i) T IFR $\rightarrow \lambda_r$ is log concave in r.
- (ii) T IFRA $\rightarrow \lambda_r$ is log subhomogeneous in r.
- (iii) T NBU $\rightarrow \lambda_r$ is log subadditive in r.

Since $\rho_r = \lambda_r \cdot [\Gamma(r+1)/r^r] \cdot e^r$, the results (5.2) follow by combining the results (5.3) and (5.4). Hence the univariate results (5.2) are weaker than the univariate results (5.4). However, in some sense, they are asymptotically equivalent because, e.g., $(\rho_r/\lambda_r)^{1/r} \rightarrow 1$ as $r \rightarrow \infty$. This is the reason the irrelevant factor e^{-r} was introduced into the normalized moments ρ_r.

Although in the univariate case the results (5.2) are weaker than those of (5.4), there are no known generalizations of (5.4) in the multivariate setting. However, the results of (5.2) do generalize. The following multivariate moment relations follow from Definition 2 in exactly the same way as those derived from Theorem 3. If $\mathbf{r} = (r_1, \ldots, r_n)$, we let $\mu_{\mathbf{r}} = E[T_1^{r_1} \ldots T_n^{r_n}]$ and set $\rho_r = \mu_{\mathbf{r}}/(\Pi_{i=1}^n r_i^{r_i} e^{-r_i})$.

THEOREM 6. *Let* $\mathbf{T}$ *be a nonnegative random vector.*

(5.5)
- (i) *If* $\mathbf{T}$ *is MIFR, then* $\rho_{\mathbf{r}}$ *is log concave in* $\mathbf{r}$.
- (ii) *If* $\mathbf{T}$ *is MIFRA, then* $\rho_{\mathbf{r}}$ *is log subhomogeneous in* $\mathbf{r}$.
- (iii) *If* $\mathbf{T}$ *is MNBU, then* $\rho_{\mathbf{r}}$ *is log subadditive in* $\mathbf{r}$.

The results (5.5) are the best available at present. In particular they are valid for the MVE of Marshall and Olkin (1967) since the MVE is MIFR, MIFRA and MNBU.

6. Some Other Classes. The recent successful use of log concave functions to characterize the IFR class has suggested other variations on this theme. Although the full ramifications of this approach are being currently investigated, we illustrate with one interesting example.

Recall that in section five we defined a nonnegative function h to be log subhomogeneous if $h(\alpha\mathbf{x}) \geq h^{\alpha}(\mathbf{x})$ for all $0 < \alpha < 1$.

THEOREM 7. *$\mathbf{T}$ is MIFRA if and only if $E[h(\mathbf{x},\mathbf{T})]$ is log subhomogeneous in $\mathbf{x}$ for all functions $h(\mathbf{x},\mathbf{t})$ which are log subhomogeneous in $(\mathbf{x},\mathbf{t})$ and are nondecreasing in $\mathbf{t}$ for each fixed $\mathbf{x}$.*

Proof. Suppose $\mathbf{T}$ is MIFRA and let $h(\mathbf{x},\mathbf{t})$ be a log subhomogeneous function which is nondecreasing in $\mathbf{t}$ for each fixed $\mathbf{x}$. Then

$$E^{\alpha}[h(\mathbf{x},\mathbf{t})] \leq E[h^{\alpha}(\mathbf{x},\mathbf{T}/\alpha)] \leq E[h(\alpha\mathbf{x},\mathbf{T})].$$

The first inequality follows since $\mathbf{T}$ is MIFRA and the second follows since h is log subhomogeneous.

On the other hand suppose $E[h(\mathbf{x},\mathbf{T})]$ is log subhomogeneous in $\mathbf{x}$ for all log subhomogeneous functions $h(\mathbf{x},\mathbf{t})$ which are nondecreasing in $\mathbf{t}$ for each fixed $\mathbf{x}$. Let $h(\mathbf{t})$ be a nondecreasing function in $\mathbf{t}$ and define $H(r,\mathbf{t}) = h^{r}(\mathbf{t}/r)$ for $r > 0$. Then $H(r,\mathbf{t})$ is log subhomogeneous in $(r,\mathbf{t})$ and is nondecreasing in $\mathbf{t}$ for each fixed r. Hence $E[H(r,\mathbf{T})]$ is log subhomogeneous in r, i.e.,

$$E[H(\alpha r,\mathbf{T})] \geq E^{\alpha}[H(r,\mathbf{T})] \quad \text{or}$$

$$E[h^{\alpha r}(\mathbf{T}/\alpha r)] \geq E^{\alpha}[h^{r}(\mathbf{T}/r)].$$

Now set $r = 1$ to conclude that $\mathbf{T}$ is MIFRA.

In Block and Savits (1978), a new characterization of the NBUE class was given.

THEOREM 8. *T is NBUE if and only if $\mu = E[T]$ is finite and*

(6.1) $$\int_0^{\infty} h(z)\bar{F}(z)dz \leq \mu \int_0^{\infty} h(z)dF(z)$$

for all nonnegative nondecreasing functions h, where $\bar{F}(t) = P(T>t)$.

The author has recently proposed a multivariate extension of (6.1) and has shown that the resulting multivariate class satisfies properties (P0)–(P6).

REFERENCES

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

BLOCK, H. W. and SAVITS, T. H. (1976). The IFRA closure problem. *Ann. Probab. 4* 1030–1032.

BLOCK, H. W. and SAVITS, T. H. (1978). Shock models with NBUE survival. *J. Appl. Probab. 15* 621–628.

BLOCK, H. W. and SAVITS, T. H. (1980). Multivariate increasing failure rate average distributions. *Ann. Probab. 8* 793–801.

BLOCK, H. W. and SAVITS, T. H. (1981). Multivariate classes in reliability theory. *Math. of O.R. 6* 453–461.

MARSHALL, A. W. and OLKIN, I. (1967). A multivariate exponential distribution. *J. Amer. Statist. Assoc. 67* 30–44.

MARSHALL, A. W., OLKIN, I. and PROSCHAN, F. (1967). Monotonicity of ratios of means and other applications of majorization. *Inequalities,* ed. O. Shisha, 177–190. Academic Press, New York.

MARSHALL, A. W. and SHAKED, M. (1982). A class of multivariate new better than used distributions. *Ann. Probab. 10* 259–264.

SAVITS, T. H. (1984). A multivariate IFR class. To appear in *J. Applied Probab.*

Inequalities in Statistics and Probability
IMS Lecture Notes–Monograph Series Vol. 5 (1984), 199–205

APPLICATIONS OF A UNIFIED THEORY OF MONOTONICITY IN SELECTION PROBLEMS

BY ROGER L. BERGER[1] and FRANK PROSCHAN[2]
North Carolina State University and Florida State University

In this paper, the general monotonicity results concerning selection problems derived by Berger and Proschan (1984) are reviewed. They are then applied to several different formulations of the selection problem. These include comparison with a control and restricted subset selection problems. Several classes of selection rules previously proposed in the literature are shown to possess the monotonicity properties. In addition, a new class of rules for the restricted subset selection formulation is proposed and shown to possess the monotonicity properties.

1. Introduction. In this paper we study some monotonicity properties of ranking and selection rules.

Recall that in a selection problem the general goal is to determine which of several populations possesses the largest value of some parameter. Based on random observations from the populations, a selection rule selects a subset of the populations and leads to an assertion such as, "The population with the largest parameter is in the selected subset." (Different formulations of the selection problem entail different assertions resulting from the selection rule.) A reasonable selection rule should be more likely to choose populations with larger parameters rather than populations with smaller parameters. This property of selection rules is called monotonicity.

In this paper we study some general monotonicity properties of a broad class of selection rules in a unified manner. We also discuss applications of these general results to several different formulations of the selection problem.

In symbols, let $\mathbf{X} = (X_1, \ldots, X_n)$ be a random observation with distribution $F(\mathbf{x}; \lambda)$, where the unknown parameter vector $\lambda = (\lambda_1, \ldots, \lambda_n) \in \Lambda \subset \mathcal{R}^n$. The general goal of a selection problem is to decide which of the coordinates of λ are the largest or which are larger than a value λ_0 (possibly unknown). A (nonrandomized) selection rule $S(\mathbf{x})$ is any measurable mapping from the sample space $\mathcal{X}$ of $\mathbf{X}$ into the set of subsets of $\{1, \ldots, n\}$. Having observed $\mathbf{X} = \mathbf{x}$, the selection rule S asserts that the largest parameters are in $\{\lambda_i : i \in S(\mathbf{x})\}$. The subset $S(\mathbf{X})$ may be of fixed or random size depending on the formulation of the selection problem under consideration. See, for example, Bechhofer (1954) (fixed size), Gupta and Sobel (1958) (random size), and Gupta (1965) (random size).

Gupta (1965) calls a selection rule *montone* if

$$\lambda_i \geq \lambda_j \text{ implies } P_\lambda(i \in S(\mathbf{X})) \geq P_\lambda(j \in S(\mathbf{X})). \tag{1.1}$$

Monotonicity is a desirable property of a selection rule since the selected subset is supposed to consist of the large values of λ_i. On a case by case basis, various authors have shown that their proposed selection rules are monotone. Monotonicity has not been investi-

---

[1] This work was supported in part by the U.S. Army Research Office Grant No. DAAG 29–82–K–0168.
[2] Research sponsored by the Air Force Office of Scientific Research Grant AFOSR 82–K–0007.

AMS 1980 subject classifications. Primary, 62F07; secondary, 62H10.
Key words and phrases: Ranking and selection, decreasing in transposition, monotonicity, comparison with a control, restricted subset selection, just rules.

gated for some formulations of the selections problem even though it is a desirable property in all formulations.

In this paper we review the results in Berger and Proschan (1984) (BP(1984)). These results concern some general notions of monotonicity which include the previously discussed notion of Gupta (1965). BP(1984) show in a simple unified way that a broad class of selection rules (which includes rules proposed for various formulations of the selection problem) possess these monotonicity properties. BP(1984) also discuss the application of these results to selection rules proposed by Bechhofer (1954), Gupta and Sobel (1958), and Gupta (1965). In the present paper, we discuss the application of these results to other formulations of the selection problem and other classes of selection rules considered in the literature. Also, a new class of selection rules for the restricted subset selection problem is proposed and shown to possess these general monotonicity properties.

The monotonicity properties we consider are the following. Let $A = \{a_1, \ldots, a_k\}$ and $b = \{b_1, \ldots, b_k\}$ denote two subsets of $\{1, \ldots, n\}$ with $|A| = |B| = k$, where $|\cdot|$ denotes subset size. Subset A *is better than* B if, for some arrangements $a_{i(1)}, \ldots, a_{i(k)}$ and $b_{j(1)}, \ldots, b_{j(k)}$ of the elements of A and B, $\lambda_{a_{i(r)}} \geq \lambda_{b_{j(r)}}$ for every $r = 1, \ldots, k$. If A is better than B, then each of the following inequalities would be desirable for a selection rule:

(1.2) $$P_\lambda[|A \cap S(\mathbf{X})| \geq m] \geq P_\lambda[|B \cap S(\mathbf{X})| \geq m] \text{ for every } m \in \mathscr{R}^1;$$

(In words, P_λ [at least m elements of A are selected] $\geq P_\lambda$ [at least m elements of B are selected].)

(1.3) $$P_\lambda(A = S(\mathbf{X})) \geq P_\lambda(B = S(\mathbf{X}));$$

and

(1.4) $$P_\lambda(|A^c \cap S(\mathbf{X})| \leq m) \geq P_\lambda(|B^c \cap S(\mathbf{X})| \leq m) \text{ for every } m \in \mathscr{R}^1;$$

where A^c and B^c are the complements of A and B, respectively.

Some special cases are of particular interest. By setting $m = k$ in (1.2) we obtain $P_\lambda[A \subset S(\mathbf{X})] \geq P_\lambda[B \subset S(\mathbf{X})]$; i.e., the better subset is more likely than the worse subset to be included in the selected subset. From the special case $m = k = 1$ in (1.2), we obtain the classical montonicity property (1.1). By setting $m = 0$ in (1.3), we obtain $P_\lambda[A \supset S(\mathbf{X})] \geq P_\lambda[B \supset S(\mathbf{X})]$; i.e., the selected subset is more likely to be in the better subset than in the worse subset.

In Section 2 the class of selection rules is discussed. The assumptions regarding the distribution $F(\mathbf{x}; \lambda)$ are discussed in Section 3. The montonicity results from BP(1984) are presented and applied to three formulations of the selection problem in Section 4. In Section 5, an extension of these results to include additional parameters and statistics is presented and applied to another formulation of the selection problem.

2. A Class of Selection Rules. In this section, a broad class of selection rules is described. All of the rules in this class will have the general monotonicity properties (1.2), (1.3), and (1.4).

Let $\pi = (\pi_1, \ldots, \pi_n)$ denote a permutation of $(1, \ldots, n)$. For any $\mathbf{x} \in \mathscr{R}^n$, let $\mathbf{x} \circ \pi$ denote $(x_{\pi_1}, \ldots, x_{\pi_n})$. Let $I_C(\cdot)$ denote the indicator function of the set C.

A nonrandomized selection rule $S(\mathbf{x})$ can be defined by specifying its individual selection probabilities, $\psi_1(\mathbf{x}), \ldots, \psi_n(\mathbf{x})$. These are defined by $\psi_i(\mathbf{x}) = I_{S(\mathbf{x})}(i)$. We will be interested in selection rules which satisfy the following for every $\mathbf{x} \in \mathscr{R}^n$ and every $i \in \{1, \ldots, n\}$

(2.1) $$\text{If } \psi_i(\mathbf{x}) = 1 \text{ and } x_j \geq x_i, \text{ then } \psi_j(\mathbf{x}) = 1,$$

and

(2.2) $\psi_{\pi_i}(\mathbf{x}) = \psi_i(\mathbf{x} \circ \pi)$ for every permutation π. Rules satisfying (2.1) have been called *natural* in some selection literature (for example, Eaton, 1967).

Nagel (1970) and Gupta and Nagel (1971) defined and investigated a class of selection rules called just rules. A selection rule is *just* if, for every $i \in \{1, \ldots, n\}$, $\psi_i(\mathbf{x})$ is a nondecreasing function of x_i and a nonincreasing function of x_j, $j \neq i$. If a selection rule is just and satisfies (2.2), then the rule satisfies (2.1). To see this, let π be the permutation defined by $\pi_i = j$; $\pi_j = i$; and $\pi_r = r$, $r = 1, \ldots, n$, $r \neq i$ or j. Then if $x_j \geq x_i$,

$$\psi_j(\mathbf{x}) \geq \psi_j(\mathbf{x} \circ \pi) = \psi_{\pi_j}(\mathbf{x} \circ \pi \circ \pi) = \psi_i(\mathbf{x}).$$

The inequality follows from the justness, and the first equality follows from (2.2)

Almost all the selection rules which have been proposed in the literature for the models described in Section 3 are just rules satisfying (2.2). Thus almost all of the selection rules which have been proposed over the last thirty years satisfy the general monotonicity properties (1.2), (1.3), and (1.4); the results in Section 4 will give a simple unified proof of this fact, as well as other consequences.

3. The Model and Key Mathematical Ideas. In this section, the concept of a decreasing in transposition (DT) function is introduced. The effect of assuming that the density of $\mathbf{X}$ is DT is discussed.

Let π and π' be two permutations and i and j two elements of $\{1, \ldots, n\}$ such that $i < j$; $\pi_i < \pi_j$; $\pi'_i = \pi_j$; $\pi'_j = \pi_i$; and $\pi'_r = \pi_r$, $r = 1, \ldots, n$, $r \neq i$ or j. We say that π' is a *simple transposition* of π; in symbols $\pi >' \pi'$.

The concept of a decreasing in transposition function plays a central role in our derivation of the general monotonicity properties. A real valued function $g(\mathbf{x}\,;\lambda)$ on $\mathcal{R}^{2n}$ is *decreasing in transposition (DT)* if

(3.1) $g(\mathbf{x}\,;\lambda) = g(\mathbf{x} \circ \pi; \lambda \circ \pi)$ for every $\mathbf{x} \in \mathcal{R}^n$, $\lambda \in \mathcal{R}^n$ and every permutation π,

and

(3.2) $x_1 \leq \ldots \leq x_n$, $\lambda_1 \leq \ldots \leq \lambda_n$, and $\pi >' \pi'$ imply $g(\mathbf{x}\,;\lambda \circ \pi) \geq g(\mathbf{x}\,;\lambda \circ \pi')$.

Hollander, Proschan, and Sethuraman (1977) (HPS(1977)) present a detailed investigation of DT functions. The DT property is called arrangement increasing by Marshall and Olkin (1979).

We assume that the observation vector $\mathbf{X} = (X_1, \ldots, X_n)$ has a density $g(\mathbf{x}\,;\lambda)$ with respect to a measure $\sigma(\mathbf{x})$, where σ satisfies $\int_A d\sigma(\mathbf{x}) = \int_A d\sigma(\mathbf{x} \circ \pi)$ for each permutation π and Borel set $A \subset \mathcal{R}^n$. We assume that g is DT. HPS(1977) list several discrete and continuous densities which are DT. For example, if $g(\mathbf{x}\,;\lambda) = \pi_{i=1}^{n} h(x_i; \lambda_i)$ and h is TP_2, then g is DT.

By Theorem 4.1 of HPS(1977), if $\mathbf{X}$ has a DT density, then the coordinates of $\mathbf{X}$ are more likely to be in the same order as the coordinates of λ than in any other order. Furthermore, the probability of a rank order for $\mathbf{X}$ decreases as the rank order becomes more transposed from the order of λ. This behavior is typical in selection problems. Usually X_i is an estimate of λ_i and so the X_i's are expected to be in approximately the same order as the λ_i's. This leads to the use of selection rules satisfying (2.1). Most of the models considered in the selection literature are models with DT densities.

A class of selection problems considered in the literature which do *not* have DT densities are problems involving unequal sample sizes. See Berger and Gupta (1980) for several references. For example, in the problem of selecting the normal population with the largest normal mean, the density of the sample means will not be DT if the population variances are equal but the sample sizes are unequal. The density does not satisfy (3.1). Thus in their present form the results of BP(1984) do not apply to these selection problems.

4. Monotonicity Properties. In this section we state the monotonicity results of BP(1984). Then we apply these results to three different formulations of the selection problem.

THEOREM 4.1. *Suppose (a) the density $g(\mathbf{x}\ ;\ \lambda)$ of $\mathbf{X}$ is DT, (b) the individual selection probabilities of the selection rule S satisfy (2.1) and (2.2), (c) $A \subset \{1, \dots, n\}$, $B \subset \{1, \dots, n\}$, and A is better than B. Then*

$$P_\lambda[|A \cap S(\mathbf{X})| \geq m] \geq P_\lambda[|B \cap S(\mathbf{X})| \geq m] \text{ for every } m \in \mathscr{R}^1, \tag{4.1}$$

$$P_\lambda[A = S(\mathbf{X})] \geq P_\lambda[B = S(\mathbf{X})], \tag{4.2}$$

and

$$P_\lambda[|A^c \cap S(\mathbf{X})| \leq m] \geq P_\lambda[|B^c \cap S(\mathbf{X})| \leq m] \text{ for every } m \in \mathscr{R}^1. \tag{4.3}$$

The proof of Theorem 4.1 is given in BP(1984). It is based on the fact that the indicator functions of the desired events are DT functions of $\mathbf{X}$ and a vector of 1's and 0's indicating which elements of $\{1, \dots, n\}$ are in A (or B) and which are not. By the Composition Theorem of HPS(1977), the probabilities are DT functions of λ and the vector of 1's and 0's. The inequalities then follow, since the vector of 1's and 0's for B is a transposition of the corresponding vector for A.

We now present some examples of selection rules satisfying the conditions of Theorem 4.1.

Example 4.1. (Restricted subset selection). Santner (1975) introduced the restricted subset formulation of the selection problem. In this formulation, a subset of random size is selected. The size of the selected subset must not exceed m, a fixed constant satisfying $1 \leq m \leq n$. Santner (1975) proposed and studied a class of restricted subset selection rules. We will propose a class of rules which satisfy the conditions of Theorem 4.1. and thus possess the monotonicity properties (4.1), (4.2), and (4.3).

Santner (1975) proposed this class of restricted subset selection rules. Let $X_{[1]} \leq \dots \leq X_{[n]}$ denote the ordered values of $X_1, \dots, X_n$. Let $h^{-1}(z)$ be a nondecreasing real valued function of the real variable z satisfying $h^{-1}(z) \leq z$. Then a rule in Santner's class is defined by:

Include i in the selected subset if and only if

$$X_i \geq \max(X_{[n-m+1]}, h^{-1}(X_{[n]})). \tag{4.4}$$

Actually Santner places more restrictions on the function h^{-1} than we have stated but these conditions are all that are important for our discussion.

We propose the following class of restricted subset selection rules. For any $\mathbf{x} \in \mathscr{R}^n$, let $\mathbf{x}^i = (x_1^i, \dots, x_{n-1}^i)$ be the vector obtained by deleting x_i and arranging the remaining $n-1$ components of $\mathbf{x}$ in increasing order. Let p be a real valued function defined on $\mathscr{R}_*^{n-1} \equiv \{y \in \mathscr{R}^{n-1}: y_1 \leq \dots \leq y_{n-1}\}$ which is nondecreasing in each coordinate. Assume that $p(\mathbf{y}) \leq y_{n-1}$ for every $\mathbf{y} \in \mathscr{R}_*^{n-1}$. Finally we assume, as Santner (1975) did, that $g(\mathbf{x}\ ;\ \lambda)$ is a density with respect to Lebesgue measure on $\mathscr{R}^n$; thus no coordinates of $\mathbf{X}$ are tied with probability one. A class of restricted subset selection rules is defined by:

Include i in the selected subset if and only if

$$X_i \geq \max(X_{[n-m+1]}, p(\mathbf{X}^i)). \tag{4.5}$$

The class of restricted subset selection rules defined by (4.5) contains the class of rules defined by (4.4). Every rule in the class (4.5) also satisfies (2.1) and (2.2) and has the general monotonicity properties (4.1), (4.2), and (4.3) if the density of $\mathbf{X}$ is DT.

Let S be a rule in the class (4.5). To see that S satisfies (2.2), note that for any permutation π, $\mathbf{x} \in \mathscr{R}^n$, and $i \in \{1, \ldots, n\}$, $x_{\pi_i} = (\mathbf{x} \circ \pi)_i$ and $\mathbf{x}^{\pi_i} = (\mathbf{x} \circ \pi)^i$. To see that S satisfies (2.1), note that if $x_j \geq x_i$ (say, $x_i = x_{[r]}$ and $x_j = x_{[s]}$ where $r < s$) then $x_t^i = x_t^j$, $t = 1, \ldots, r-1$, $x_t^i \geqslant x_t^j$, $t = r, \ldots, s-1$, and $x_t^i = x_t^j$, $t = s, \ldots, n-1$. Thus, by the monotonicity of p, $p(\mathbf{x}^i) \geqslant p(\mathbf{x}^j)$. So if $\psi_i(\mathbf{x}) = 1$, then $x_j \geqslant x_i \geqslant \max(x_{[n-m+1]}, p(\mathbf{x}^i)) \geqslant \max(x_{[n-m+1]}, p(\mathbf{x}^j))$ and $\psi_j(\mathbf{x}) = 1$.

To see that Santner's (1975) class of restricted subset selection rules is a subset of the class (4.5), let h^{-1} be a function which defines a rule in (4.4). For $\mathbf{y} \in \mathscr{R}_*^{n-1}$ define $p(\mathbf{y}) = h^{-1}(y_{n-1})$. By use of the properties of h^{-1}, our restrictions on p are easily verified. If $x_i = x_{[n]}$ both the rule defined with p and the rule defined with h^{-1} include i in the selected subset. If $x_i \neq x_{[n]}$, then $x_{n-1}^i = x_{[n]}$ and $p(\mathbf{x}^i) = h^{-1}(x_{[n]})$. Thus Santner's rule from (4.4) with h^{-1} is equivalent to the rule from (4.5) defined with p.

Santner (1975) showed that every rule in the class he considered had the classical monotonicity property (1.1). Santner assumed that the coordinates of $\mathbf{X}$ are independent, the density of x_i is $g(x_i ; \lambda_i)$, and the family $g(x ; \lambda)$ is stochastically increasing. Under these same conditions, using a proof very similar to Santner's, we can show that every rule in the class (4.5) has the montonicity property (1.1). In addition, we can conclude, using Theorem 4.1, that any rule in the class (4.5) satisfies the monotonicity properties (4.1), (4.2), and (4.3) if the density of $\mathbf{X}$ is DT. Inequality (4.1) includes property (1.1) as a special case.

Example 4.2. (Comparison with a control). Lehmann (1961) formulated the comparison with a control problem in this way. A population is called *positive* if $\lambda_i \geqslant \lambda_0 + \Delta$ and *negative* if $\lambda_i \leqslant \lambda_0$, where $\Delta > 0$ and λ_0 are fixed constants. The general goal is to select a subset containing positive populations.

Lehmann (1961) derived minimax rules which minimize $\sup_\Lambda R(\lambda, S)$ subject to $\inf_{\Lambda'} T(\lambda, S) \geqslant \gamma$. Here γ is a fixed constant, Λ' is the subset of Λ for which at least one population is positive, R is either of two criteria concerning the number of negative populations selected, and T is any of four criteria concerning the number of positive populations selected.

One application of Lehmann's (1961) results is the following. Assume $X_1, \ldots, X_n$ are independent. Assume X_i is a sufficient statistic computed from a sample from the i-th population. Assume the density $g(x_i ; \lambda_i)$ of X_i possesses the monotone likelihood ratio property. Then the rule defined by $\psi_i(\mathbf{x}) = 1, \alpha, 0$ according as $X_i >, =, < C$ is minimax, where α and C are determined by $E_{\lambda_0+\Delta}\, \psi_i(X_i) = \gamma$.

The above assumptions imply that the density of $\mathbf{X}$ is DT. $\psi_i(\mathbf{x})$ will satisfy (2.1) and (2.2) if $\alpha = 0$ or $\alpha = 1$. This will be the case if $g(x ; \lambda)$ is a density with respect to Lebesgue measure. It will also be the case for certain values of λ_0 and Δ if $g(x ; \lambda)$ is a Poisson or binomial density. In each of these cases, Theorem 4.1 implies that the minimax rule will satisfy the montonicity properties (4.1), (4.2), and (4.3).

Example 4.3. (Just subset selection rules). In Section 2, it was shown that all just rules which satisfy (2.2) also satisfy (2.1). Thus, if the density of $\mathbf{X}$ is DT, any just rule satisfying (2.2) has the monotonicity properties (4.1), (4.2), and (4.3).

Historically, the concept of justness has been used only with the unrestricted subset selection formulation of Gupta (1965). For example, Bjornstad (1981) recently investigated a large class of just rules. But the concept of justness is equally appealing for other formulations of the selection problem. Indeed, the rules considered in Examples 4.1 and 4.2 are just.

5. Additional Parameters and Statistics. BP(1984) prove this more general monotonicity result applying to models which include other parameters besides λ and other statistics in addition to $\mathbf{X}$. Let $\mathbf{Y}$ be a statistic, possibly a vector, with sample space $\mathcal{Y}$. Let ν be a parameter, possibly a vector, with a set of possible values denoted by N.

THEOREM 5.1 *Assume that* $(\mathbf{X}, \mathbf{Y})$ *has a density* $g(\mathbf{x},\mathbf{y}; \lambda,\nu)$ *with respect to a measure* $\sigma(\mathbf{x}) \times \mu(\mathbf{y})$, *where* σ *satisfies* $\int_A d\sigma(\mathbf{x}) = \int_A d\sigma(\mathbf{x} \circ \pi)$ *for each permutation* π *and Borel set* $A \subset \mathcal{R}^n$. *Assume that for each* $\mathbf{y} \in \mathcal{Y}$ *and* $\nu \in N$, $g(\mathbf{x},\mathbf{y}; \lambda,\nu)$ *is a DT function of* $\mathbf{x}$ *and* λ.

Let $\psi_1(\mathbf{x}, \mathbf{y}), \ldots, \psi_n(\mathbf{x}, \mathbf{y})$ *denote the individual selection probabilities of a nonrandomized selection rule* $S(\mathbf{X}, \mathbf{Y})$. *Assume* (a) *for every* $\mathbf{y} \in \mathcal{Y}$, *if* $\psi_i(\mathbf{x}, \mathbf{y}) = 1$ *and* $x_j \geq x_i$, *then* $\psi_j(\mathbf{x}, \mathbf{y}) = 1$; (b) $\mathbf{y} \in \mathcal{Y}$, $\mathbf{x} \in \mathcal{R}^n$, $i \in 1, \ldots, n$, *and* π *a permutation imply* $\psi_{\pi_i}(\mathbf{x}, \mathbf{y}) = \psi_i(\mathbf{x} \circ \pi, \mathbf{y})$.

Let $A \subset \{1, \ldots, n\}$ *and* $B \subset \{1, \ldots, n\}$. *If* A *is better than* B, *then*

$$P_{\lambda,\nu}[|A \cap S(\mathbf{X}, \mathbf{Y})| \geq m] \geq P_{\lambda,\nu}[|B \cap S(\mathbf{X}, \mathbf{Y})| \geq m] \text{ for every } m \in \mathcal{R}^1, \tag{5.1}$$

$$P_{\lambda,\nu}[|A^c \cap S(\mathbf{X}, \mathbf{Y})| \leq m] \geq P_{\lambda,\nu}[|B^c \cap S(\mathbf{X}, \mathbf{Y})| \leq m] \text{ for every } m \in \mathcal{R}^1, \tag{5.2}$$

and

$$P_{\lambda,\nu}[A = S(\mathbf{X}, \mathbf{Y})] \geq P_{\lambda,\nu}[B = S(\mathbf{X}, \mathbf{Y})]. \tag{5.3}$$

The proof of Theorem 5.1 may be found in BP(1984).

Example 5.1. (Comparison with an unknown control). Tong (1969) formulated the problem of comparison with a control in this way. $X_0, X_1, \ldots, X_n$ are independent normal random variables with means $\lambda_0, \lambda_1, \ldots . \lambda_n$ and common known variance σ^2/N_0. The parameter λ_0 is the unknown control value. For $i = 1, \ldots, n$, λ_i is bad if $\lambda_i \leq \lambda_0 + \delta_1$ and λ_i is good if $\lambda_1 \geq \lambda_0 + \delta_2$, where $\delta_1 < \delta_2$ are known constants. The sample size N_0 is chosen so that the probability that all of the good populations are selected but none of the bad populations is selected is at least a preassigned value.

In our notation, $\mathbf{X} = (X_1, \ldots, X_n)$, $\lambda = (\lambda_1, \ldots, \lambda_n)$, $\mathbf{Y} = X_0$ and $\nu = \lambda_0$. Let $d = (\delta_1 + \delta_2)/2$. Tong (1969) showed that the selection rule which includes i in $S(\mathbf{X}, X_0)$ if and only if $X_i - X_0 > d$ is Bayes, minimax, and admissible among a class of translation invariant rules.

The conditions of Theorem 5.1 are easily verified for this selection rule and model. Thus $S(\mathbf{X}, X_0)$ possesses the general monotonicity properties (5.1), (5.2), and (5.3). For example, if A is the set of good parameters and B is any other set of equal size, then, by (5.3), A is more likely to be the selected set than is B.

In other applications, ν might include nuisance parameters, which have no bearing on which λ_i's are preferred, as well as control parameters, like λ_0. Similarly, $\mathbf{Y}$ might include estimates of nuisance parameters.

6. Conclusion. In this paper, we have reviewed the general monotonicity results for selection rules of BP(1984). By examples, we have indicated that almost all nonrandomized selection rules which have been proposed for models with DT densities possess the general monotonicity properties. Thus, results which have previously been derived on a case by case basis may now be obtained using this unified theory; in addition, other results may be obtained.

REFERENCES

BECHHOFER, R. E. (1954). A single-sample multiple decision procedure for ranking means of normal populations with known variances. *Ann. Math. Stat.* *25* 16–39.

BERGER, R. L. and GUPTA, S. S. (1980). Minimax subset selection rules with applications to unequal variance (unequal sample size) problems. *Scand. J. Statist.* *7* 21–26.

BERGER, R. L. and PROSCHAN, F. (1984). Monotonicity in selection problems: a unified approach. To appear in *Ann. Statist.* *12*.

BJØRNSTAD, J. F. (1981). A class of Schur procedures and minimax theory for subset selection. *Ann. Statist.* *9* 777–791.

EATON, M. L. (1967). Some optimum properties of ranking procedures. *Ann. Math. Statist.* *38* 124–137.

GUPTA, S. S. (1965). On some multiple decision (selection and ranking) rules. *Technometrics* *7* 225–245.

GUPTA, S. S. and NAGEL, K. (1971). On some contributions to multiple decision theory. *Statistical Decision Theory and Related Topics* (S. S. Gupta and J. Yackel, eds.), Academic Press, New York, 79–102.

GUPTA, S. S. and SOBEL, M. (1958). On selecting a subset which contains all populations better than a standard. *Ann. Math. Statist.* *29* 235–244.

HOLLANDER, M., PROSCHAN, F., and SETHURAMAN, J. (1977). Functions decreasing in transposition and their application in ranking problems. *Ann. Statist.* *5* 722–733.

LEHMANN, E. L. (1961). Some Model I problems of selection. *Ann. Math. Statist.* *32* 990–1012.

MARSHALL, A. W. and OLKIN, I. (1979). *Inequalities: Theory of Majorization and Its Applications*. Academic Press, New York.

NAGEL, K. (1970). On subset selection rules with certain optimality properties. Ph.D. thesis, Statistics Department, Purdue University, West Lafayette, IN.

SANTNER, T. J. (1975). A restricted subset selection approach to ranking and selection problems. *Ann. Statist.* *3* 334–349.

TONG, Y. L. (1969). On partitioning a set of normal populations by their locations with respect to a control. *Ann. Math. Statist.* *40* 1300–1324.

Inequalities in Statistics and Probability
IMS Lecture Notes–Monograph Series Vol. 5 (1984), 206–210

SELECTING THE t BEST CELLS OF A MULTINOMIAL USING INVERSE SAMPLING

BY PINYUEN CHEN and MILTON SOBEL[1]
Syracuse University and University of California at Santa Barbara

An inverse sampling procedure R is proposed for selecting the t "best" cells (i.e., cells with the largest cell probabilities) from a multinomial distribution with k cells ($1 \leqslant t < k$). Two different formulations of this selection problem are considered and the measure of distance in both formulations is the ratio of the largest and second largest cell probabilities. One formulation is of the usual type based on an empty indifference zone; in the other (new) formulation any collection of t cells from the union of the preference zone (for selection) and the indifference zone is called a correct selection. Type 2-Dirichlet integrals are used (i) to express the probability of correct selection as an integral with parameters only in the limits of integration, and (ii) to prove that the least favorable configuration for each of the formulations under R is the so-called slippage configurations with k–t equal cell probabilities and t cell probabilities slipped to the right by a common amount.

1. Introduction. One of the important applications of ranking and selection techniques is to select (without respect to order) the t best cells of a multinomial distribution with k cells. For the special case $t = 1$ the fixed sample size problem was first considered by Bechhofer, Elmaghraby and Morse (1959) and the inverse sampling procedure was first considered by Cacoullos and Sobel (1966). We are presently discussing fixed subset size problems and not considering the random subset size problem which was considered by Gupta and Nagel (1971) and more recently by Hu (1982). It is well known by people working in this area that the generalization of the fixed subset size problem to arbitrary t ($1 < t < k$) presents some serious difficulties (cf. the work of Lee (1975) and Hwang, Hsuan and Parned (1980) on this topic). In this paper we consider the corresponding problem for general $t \geqslant 1$ with an inverse sampling procedure.

Actually we consider two different formulations of the ranking and selection problem. The measure of distance in both formulations is the ratio of cell probabilities as in the previous references. Let

$$p_{[1]} \leqslant p_{[2]} \leqslant \ldots \leqslant p_{[k-t]} \leqslant p_{[k-t+1]} \leqslant \ldots \leqslant p_{[k]} \tag{1.1}$$

denote the ordered cell probabilities which sum to one. Let $\delta^* > 1$ and $P^* (\binom{k}{t}^{-1} < P^* < 1)$ denote specified constants. In the usual (or first) formulation we require a procedure R such that

$$P\{CS|R\} > P^* \text{ whenever } \delta \geqslant \delta^*, \tag{1.2}$$

where $\delta = p_{[k-t+1]}/P_{[k-t]}$.

Actually we need only consider configurations (1.1) with $p_{[k-t]} < p_{[k-t+1]}$ and in this case the definition of correct selection (CS) is clear, namely that we select the t cells with largest p-values.

We shall say the p-value is in the indifference zone (IZ) if it lies strictly between $p_{[k-t+1]}/\delta^*$ and $p_{[k-t+1]}$. The p-values $\geqslant p_{[k-t+1]}$ will be said to lie in the preference zone (PZ).

---

[1] This paper was supported by National Science Foundation Grant Number MCS82–02247.
AMS 1980 subject classifications. Primary, 62F07; secondary, 62E15.
Key words and phrases: Ranking and selection, inverse sampling, multinomial distribution, Type 2-Dirichlet integral.

Under the alternative (or second) formulation we consider any combination of the t cell probabilities, each $\geqslant p_{[k-t+1]}/\delta^*$, as being a correct selection and we use the terminology CSA for any such combination. Thus if we take any t cells from those in the union of the PZ and IZ as our selected subset, we call this a correct selection (CSA) under the second formulation.

Our goal is to show that the so-called least favorable configuration is the one with t cell probabilities slipped to the right by a common amount. We show below that under inverse sampling this is least favorable for both of the above formulations.

The main tool used in this paper is the fact that the P(CS) and also the P(CSA) can be expressed exactly in terms of type 2-Dirichlet integrals. This turns out to be highly useful because it is exact and because the p-values show up only in the limits of integration.

For both formulations we use the same sampling and the same decision procedure specified by a positive integer r to be determined (with the help of Dirichlet Tables).

Procedure R: Continue sampling one-at-a-time until t cells reach a frequency of at least r. As soon as this occurs we stop and select these t cells as being those with the t largest probabilities.

It is clear that under this procedure there can be no ties for the t-th position and hence the selected subset is well defined. Note that we are selecting the t best without respect to order, so that frequency ties present no difficulty.

2. P(CS) and Least Favorable Configuration for the First Formulation. It has been shown (Sobel, Uppuluri and Frankowski (1983)) that for a multinomial distribution the probability, when a specified cell (called the counting cell) reaches frequency m, that certain $(t-1)$ specified cells all have frequency $\geqslant r$ and the remaining $(k-t)$ cells all have frequency $< r$, is given exactly by the CD integral

$$\text{(2.1)} \qquad \mathrm{CD}_{\mathbf{a}}^{(t-1;\,k-t)}(r, m) = [\Gamma(m+(k-1)r)/\Gamma^{k-1}(r)\Gamma(m)] \int_0^{a_1} \cdots \int_0^{a_{t-1}} \int_{a_t}^{\infty} \cdots \int_{a_{k-1}}^{\infty} \Big(1 + \sum_{i=1}^{k-1} x_i\Big)^{-(m+(k-1)r)} \cdot \prod_{i=1}^{k-1} x_i^{r-1}\, dx_i$$

where $\mathbf{a} = (a_1, \ldots\ a_{t-1}, a_t, \ldots, a_{k-1})$ and $a_j = p_j/p_0$ is the ratio of the j-th cell probability to that of the counting cell; here we have assumed that the first $t-1$ cells form the specified set of size $t-1$ and that the counting cell is the last cell with probability p_0.

Using this probability interpretation with $m = r$ we can write the P(CS) for the first formulation above as the sum of t terms; in the j-th term we take the cell associated with $p_{[j]}$ as the counting cell ($j = k-t+1, k-t+2, \ldots\ k$). Hence we obtain

$$\text{(2.2)} \qquad P(\mathrm{CS}|R) = \sum_{j=k-t+1}^{k} \mathrm{CD}_{\mathbf{a}_j}^{(t-1;\,k-t)}(r, r)$$

where

$$\text{(2.3)} \qquad \mathbf{a}_j = (p_{[k]}/p_{[j]}, \ldots, p_{[j+1]}/p_{[j]}, p_{[j-1]}/p_{[j]}, \ldots, p_{[1]}/p_{[j]}).$$

Thus the sum of the two superscripts is the number of components of $\mathbf{a}$ and from (2.1) we see that the first $t-1$ components are upper limits in the CD integral, while the last $k-t$ components are lower limits.

The same probability interpretation gives us another expression for the P(CS), which we need for our result, by taking any one of the second set of $k-t$ specified cells as our counting cell. Thus we can also write

$$\text{(2.4)} \qquad P(\mathrm{CS}|R) = \sum_{j=1}^{k-t} \mathrm{CD}_{\mathbf{a}_j}^{(t,\,k-t-1)}(r, r),$$

where $\mathbf{a}_j$ is again given by (2.3) but j runs over a different set.

Under the first formulation the condition $\delta \geqslant \delta^*$ is equivalent to the inequality $p_{[k-t]} < p_{[k-t+1]}/\delta^*$ or equivalently the open interval $(p_{[k-t+1]}/\delta^*, p_{[k-t+1]})$ is empty. Using (2.2) and keeping the pairwise ratio of each two of the t largest p-values fixed we now consider $p_{[k-t]}$ as a variable. Since $p_{[k-t]}$ is in the numerator of the lower limit, the $P(\mathrm{CS})$ is monotonically decreasing in $p_{[k-t]}$. Hence we can decrease the $P(\mathrm{CS})$ by increasing the value of $p_{[k-t]}$ until it reaches $p_{[k-t+1]}/\delta^*$. The increase in the value of $p_{[k-t]}$ is offset by a decrease in $p_{[j]}$, as the sum of all the p-values has to remain equal to one. Since all the lower limits of the last $k-t$ integrals are increased, the value of the $P(\mathrm{CS})$ must decrease. Note that $p_{[j]} \geqslant p_{[k-t+1]} > p_{[k-t+1]}/\delta^*$; hence $p_{[k-t]}$ must reach its boundary first. The same argument allows us to increase in turn $p_{[k-t+1]}, p_{[k-t+2]}, \ldots, p_{[1]}$ up to the same boundary value, namely $p_{[k-t+1]}/\delta^*$.

We now use (2.4) with $p_{[j]} = p_{[k-t+1]}/\delta^*$ for $j = 1, 2, \ldots, k-t$. Consider $p_{[k-t+2]}$ as a variable with all the largest p-values as fixed, except that $p_{[k-t+1]}$ and the boundary $p_{[k-t+1]}/\delta^*$ can still vary. The $P(\mathrm{CS})$ has now been decreased to the value $P_1(\mathrm{CS})$ where

$$P_1\{\mathrm{CS}|R\} = (k-t)\mathrm{CD}_{\mathbf{a}}^{(t,\,k-t-1)}(r, r) \tag{2.5}$$

where

$$\mathbf{a} = (p_{[k]}/p_{[k-t+1]})\,\delta^*, \ldots, (p_{[k-t+2]}/p_{[k-t+1]})\delta^*, \delta^*, 1, \ldots, 1)$$

and the last $k-t-1$ components are all exactly 1. Since $p_{[k-t+2]}$ appears as the numerator of an upper limit of integration in (2.5), it follows that we can further decrease the $P(\mathrm{CS})$ by decreasing $p_{[k-t+2]}$ to $p_{[k-t+1]}$; actually $p_{[k-t+1]}$ was increasing so equality has to occur. Similarly we decrease all the larger p-values until they reach $p_{[k-t+1]}$.

The above argument proves the following

THEOREM 1. *The least favorable configuration for the first formulation is given by*

$$\begin{aligned} p_{[1]} = p_{[2]} = \ldots = p_{[k-t]} &= 1/(k-t+t\delta^*), \\ p_{[k-t+1]} = \ldots = p_{[k]} &= \delta^*/(k-t+t\delta^*). \end{aligned} \tag{2.6}$$

3. P(CSA) and the Least Favorable Configuration for the Second Formulation.

Consider the general configuration for the second formulation as follows:

$$p_{[1]} \leqslant \ldots \leqslant p_{[k-t-r]} \leqslant p_{[k-t+1]}/\delta^* < p_{[k-t-r+1]} \leqslant \ldots \leqslant p_{[k-t]} \leqslant \ldots \leqslant p_{[k]} \tag{3.1}$$

where r is the number of cell probabilities in the IZ.

The probability of correct selection under the second formulation can be written in the following Dirichlet integral form:

$$P(\mathrm{CSA}|R) = \Sigma^* \Sigma_{j=1}^{t} \mathrm{CD}_{\mathbf{a}_j}^{(t-1,\,k-t)}(r, r), \tag{3.2}$$

where Σ^* is over all possible subsets $\{p_{s_1}, p_{s_2}, \ldots, p_{s_t}\}$ of size t that can be taken from the set $\{p_{[k-t-r+1]}, \ldots, p_{[k]}\}$ of size $t+r$ and

$$\mathbf{a}_j = (p_{s_j})^{-1}(p_{s_1}, \ldots, p_{s_{j-1}}, p_{s_{j+1}}, \ldots, p_{s_t}, p_{[1]}, \ldots, p_{[k]}). \tag{3.3}$$

It should be noted that there are only $k-t$ components in (3.3) after p_{s_t}/p_{s_j} and that the numerators of these are taken from the set $\{p_{[1]}, p_{[2]}, \ldots, p_{[k]}\} - \{p_{s_1}, \ldots, p_{s_t}\}$ so that $\mathbf{a}_j$ in (3.3) has a total of $k-1$ components.

Using (3.2) and keeping the pairwise ratio of each of the $t+r$ largest p-values fixed, we now consider $p_{[k-t-r]}$ as a variable. Since $p_{[k-t-r]}$ is a numerator among the lower limits in (3.2), the $P(\mathrm{CSA})$ is monotonically decreasing in $p_{[k-t-r]}$. Hence we can decrease the $P(\mathrm{CSA})$ by increasing the values of $p_{[k-t-r-1]}, \ldots, p_{[1]}$ up to the common boundary value, namely $p_{[k-t+1]}/\delta^*$.

Thus we can restrict ourselves to the following configuration:

(3.4) $$p_{[1]} = \dots = p_{[k-t-r]} = p_{[k-t+1]}/\delta^* < p_{[k-t-r+1]} \leq \dots \leq p_{[k]}$$

for minimizing $P(\mathrm{CSA}|R)$ in (3.2).

It is clear from (3.4) that

$$p_{[1]}/p_{[k-t+1]} = \dots = p_{[k-t+r]}/p_{[k-t+1]} = (\delta^*)^{-1}.$$

Let $p_{[j]}/p_{[k-t+1]} = a_j$ for $j = k-t-r+2, \dots, k$ be kept as constants and let $p_{[k-t+1]}/p_{[k-t-r+1]} = x$ be the only variable in $P\{\mathrm{CSA}|R\}$ in (3.4) with the obvious restrictions that $\Sigma_{i=1}^{k} p_{[i]} = 1$ and $a_{k-t+1} = p_{[k-t+1]}/p_{[k-t+1]} = 1$. Then from (3.4) we can write

(3.5) $$P\{\mathrm{CSA}|R\} = \Sigma_1^* \Sigma_{\alpha=1}^{t} \mathrm{CD}_{\mathbf{a}_\alpha}^{(t-1;k-t)}(r, r)$$
$$+ \Sigma_2^* [\Sigma_{\beta=1}^{k-t-r} \mathrm{CD}_{\mathbf{a}_\beta}^{(t;k-t-1)}(r, r) + \Sigma_{\substack{\gamma=k-t-r+1 \\ \gamma \neq s_1, \dots, s_t}}^{k} \mathrm{CD}_{\mathbf{a}_\gamma}^{(t;k-t-1)}(r, r)]$$

where Σ_1^* is over the subsets $(p_{s_1}, \dots, p_{s_t})$ of size t which do not include $p_{[k-t-r+1]}$ and Σ_2^* is over those that do include $p_{[k-t-r+1]}$. In the former case (i.e., under Σ_1^*) the structure of $\mathbf{a}_\alpha$ is

(3.6) $$\mathbf{a}_\alpha = (p_{s_1}/p_{s_\alpha}, \dots, p_{s_{\alpha-1}}/p_{s_\alpha}, p_{s_{\alpha+1}}/p_{s_\alpha}, \dots, p_{s_t}/p_{s_\alpha}, 1/(\delta^* a_{s_\alpha}), \dots,$$
$$1/(\delta^* a_{s_\alpha}), 1/(x a_{s_\alpha}), a_{k-t-r+2}/a_{s_\alpha}, \dots, a_k/a_{s_\alpha})$$

where x appears in exactly one component and we are holding all the other ratios fixed. In the latter case, (i.e., under Σ_2^*) we use the alternative form (2.4) to write the relevant probabilities (i.e., the counting cells are taken from the set that is not selected) and we separate this sum into two parts according to whether the counting cell is among $p_{[1]}, \dots, p_{[k-t-r]}$ or is in the difference of the two sets $\{p_{[k-t-r+1]}, \dots, p_{[k]}\} - \{p_{s_1}, \dots, p_{s_t}\}$. In the first of these two parts we write $\mathbf{a}_\beta$ and its structure is

(3.7) $$\mathbf{a}_\beta = (p_{s_1}/p_\beta, \dots, \delta^*/x, \dots, p_{s_t}/p_\beta, 1/(\delta^* a_\beta), \dots, p_{\beta-1}/p_\beta, p_{\beta+1}/p_\beta, \dots, a_k/a_\beta)$$

where δ^*/x comes from the ratio $p_{[k-t-r+1]}/p_\beta$ and the remaining ratios are all fixed. In the second of these two parts we write $\mathbf{a}_\gamma$ and its structure is

(3.8) $$\mathbf{a}_\gamma = (p_{s_1}/p_\gamma, \dots, 1/(x a_\gamma), \dots, p_{st}/p_\gamma, 1/(\delta^* a_\gamma), \dots, a_k/a_\gamma),$$

where $(x a_\gamma)^{-1}$ comes from the ratio $p_{[k-t-r-1]}/p_r$ and the remaining ratios are all fixed.

Note that the total number of terms in $\Sigma_1^* \Sigma_{\alpha=1}^{t}$ is $\binom{t+r-1}{t} \cdot t = (t+r-1)!/[(t-1)!(r-1)!]$ and the total number of terms in the second part of Σ_2^*, namely in $\Sigma_2^* \Sigma_{\substack{\gamma=k-t-r+1 \\ \gamma \neq s_1, \dots, s_t}}^{k}$ is $\binom{t+r-1}{t-1} \cdot r = (t+r-1)!/[(t-1)!(r-1)!]$ also.

Actually we can set up a 1-1 correspondance between the terms in Σ_1^* and those in the second part of Σ_2^* as follows. Each term in Σ_1^* corresponds to a selected subset of size t and one of these t cells is used as a counting cell. Say we have $p_k, \dots, p_{k-t+1}$ and p_k is the counting cell to specify a single term in Σ_1^*. Then we take the selected subset to be $p_{[k-t-r+1]}, p_{[k-t+1]}, \dots, p_{[k-1]}$ and use $p_{[k]}$ as a counting cell to obtain a specific term in the second part of Σ_2^*, and this illustrates the correspondance of the terms. In Σ_1^* the varying lower limit is $(x a_\alpha)^{-1}$, which in our example is $(x a_k)^{-1}$ and in the corresponding term in the second part of Σ_2^* has the varying upper limit $(x a_\gamma)^{-1}$, which in our example is $(x a_k)^{-1}$. The other limits are all the same in corresponding terms. Hence the derivatives of corresponding terms cancel. Since the only remaining terms are those from the first part of Σ_2^* and these are all negative, it follows that $P(\mathrm{CSA}|R)$ is decreasing in x. Thus we decrease $P(\mathrm{CSA}|R)$ by lowering $p_{[k-t-r+1]}$ to $p_{[k-t+1]}/\delta^*$. Similarly we decrease in turn all the $p_{[k-t-r+2]}, \dots, p_{[k-t]}$ until they reach $p_{[k-t+1]}/\delta^*$. The above argument proves the following

THEOREM 2. *The least favorable configuration for the second formulation is given by*

(3.8) $$p_{[1]} = p_{[2]} = \ldots = p_{[k-t]} = 1/(k-t+t\delta^*),$$
$$p_{[k-t+1]} = \ldots = p_{[k]} = \delta^*/(k-t+t\delta^*),$$

exactly the same slippage configuration as in (2.6).

REFERENCES

BECHHOFER, R. E., S. A. ELMAGHRABY, and N. MORSE (1959). A single-sample multiple-decision procedure for selecting the multinomial event which has the largest probability. *Ann. Math. Statist. 30* 102–119.

CACOULLOS, T. and M. SOBEL (1966). An inverse sampling procedure for selecting the most probable event in a multinomial distribution, in *Multivariate Analysis: Proceedings of an International Symposium,* P. R. Krishnaiah, ed., Academic Press, New York, 423–455.

GUPTA, S. S. and K. NAGEL (1971). On some contributions to multiple decision theory, in *Statistical Decision Theory and Related Topics,* S. S. Gupta and J. Yackel, eds., Academic Press, New York, 79–102.

HU, S. P. (1982). Inverse sampling procedure for multinomial subset selection problems. Ph.D. Dissertation, Department of Mathematics, University of California at Santa Barbara, Santa Barbara, CA.

HWANG, F. K., F. HSUAN, and M. PARNED (1980). Some inequalities for the multinomial selection problem. Unpublished manuscript.

LEE, Y. J. (1975). On selecting the *t* most probable multinomial events: Single-sample procedures. Unpublished manuscript.

SOBEL, M., V. R. R. UPPULURI, and K. FRANKOWSKI (1983). *Selected Tables in Mathematical Statistics: Dirichlet Distribution - Type 2,* American Mathematical Society, Providence, RI (to appear).

Inequalities in Statistics and Probability
IMS Lecture Notes–Monograph Series Vol. 5 (1984), 211–227

ON SOME INEQUALITIES AND MONOTONICITY RESULTS IN SELECTION AND RANKING THEORY[1]

BY SHANTI S. GUPTA, DENG-YUAN HUANG and S. PANCHAPAKESAN
Purdue University, National Taiwan Normal University and Southern Illinois University

Several inequalities and monotonicity results have been obtained in the study of selection and ranking problems; these, in fact, are germane to the development of the theory. Basic to the setup of these problems is the assumption regarding some order relations such as stochastic ordering and the monotone likelihood ratio property. These and other related ideas, along with some basic inequalities that arise under these assumptions are reviewed. Further, some important inequalities relevant to selection from restricted families of distributions defined by some partial order relations (such as IFR and IFRA families) are also discussed. Several specific results relating to multivariate normal, multinomial and gamma distributions are also reviewed.

1. Introduction. Inequalities play a fundamental role in nearly all branches of mathematics—especially so in probability and statistics. The impact of basic inequalities such as those that carry the names of Cauchy-Schwarz, Chebyshev, Cramér-Rao, and Bonferroni in statistics is well known. Inequalities have been profitably used to obtain bounds for probabilities that are more tedious to compute or analytically impossible to handle. Especially in reliability problems, the limited assumptions that could be made about the nature of the life distributions of the components of a system as well as the structure of the system itself render inequalities not merely useful and desirable but essential. Since interest in inequalities pervades through nearly all branches of mathematics, significant contributions have been made by a very large number of researchers whose efforts span well over a century. From time to time, books and monographs have been written which are completely devoted to inequalities. The classic book of Hardy, Littlewood and Pólya (1934) is a remarkable collection of mathematical inequalities. Some important works that followed are Beckenbach and Bellman (1961), Godwin (1964), Kazarinoff (1961), Marshall and Olkin (1979), Mitrinović (1964, 1970), Pólya and Szegö (1951), Shisha (1967), and Tong (1980). Of these, the monographs of Marshall and Olkin (1979) and Tong (1980) contain the recent developments in the area of multivariate probability inequalities; this topic has seen a major growth in the last ten or fifteen years. In this connection we also refer to a recent review paper by Eaton (1982).

In selection and ranking problems, inequalities and monotonicity properties have a vital role to play. Consider the classical formulations of these problems in which one proposes a procedure which will guarantee a minimum probability of correct selection (PCS). This amounts to evaluating the PCS, determining the parametric configuration for which the PCS is minimum, and then determine the constants defining the procedure so that this minimum is at least a specified level P^*. Determining this configuration, known as a least favor-

---

[1] This research was supported by the Office of Naval Research Contract N00014-75-C-0455 at Purdue University. Reproduction in whole or in part is permitted for any purpose of the United States Government.

AMS 1980 subject classifications: Primary, 62F07; Secondary, 62N05.

Key words and phrases. Selection and ranking, stochastic ordering, monotone likelihood ratio, generalizations, probability of correct selection, expected subset size, sufficient conditions for monotonicity, restricted families, partial ordering, $\mathcal{H}$-ordering inequalities, multivariate normal, multinomial, gamma, exponential family, reliability theory.

able configuration (LFC), is a vital part of the analysis. Obviously, this involves establishing an inequality that the PCS for a certain parametric configurations does not exceed the PCS for any other configuration. In some situations, this can be established by demonstrating a monotonic behavior of the PCS. There are a number of problems in which the LFC cannot be analytically established; in such cases, recourse has been taken to obtain a good lower bound for the PCS first and then seek the LFC for this lower bound. Even when the LFC for the PCS can be analytically established, inequalities are further useful in obtaining conservative but easier-to-compute values for the constants of the procedure. Similar situations arise when we consider the worst configuration for any suitable performance characteristic such as the expected number of nonbest populations included in the selected subset. Additional uses of inequalities arise due to specific assumptions regarding the families of distributions under consideration; for example, distributions having an increasing failure rate (IFR) and increasing failure rate average (IFRA). For a general view of selection and ranking problems and the various formulations and goals that have been studied, we refer to Gupta and Panchapakesan (1979).

In this paper, we restrict our attention mainly to some inequalities and monotonicity properties that have typically arisen in the development of the selection and ranking theory. Basic to the setup of these problems is the assumption regarding some order relations such as stochastic ordering and the monotone likelihood property. These and other related ideas, along with some basic inequalities that arise under these assumptions are discussed in Section 2. In reliability models, partial order relations such as convex ordering, star ordering and tail ordering play an important role. Section 3 deals with restricted families of distributions defined by such partial order relations and some important inequalities obtained in the investigation of selection problems for such families. Interesting inequalities appear in the study of selection rules for normal, multinomial and gamma distributions. These are discussed in Section 4.

2. Ordered Families of Distributions. Inherent to a selection and ranking problem is the choice of a ranking parameter, say, θ. The natural setup consists of k populations that are described by their associated probability distributions P_{θ_i}, $i = 1, \ldots, k$, where $\theta_i \epsilon \Omega$, a subset of the real line. In other words, these populations belong to a family $\mathcal{P} = \{P_\theta\}$ indexed by $\theta \epsilon \Omega$. A reasonable procedure can be proposed if we have some knowledge of the structural properties of this family. For example, if $X_1, \ldots, X_k$ are observations from the k populations, we would like to say that large values of X generally go with large values of θ. Such statements bring in order relations for distributions belonging to the family. We will now formalize such concepts and state some monotonicity results.

2.1. Stochastic Ordering and Monotone Likelihood Ratio Property. Let X be a real valued random variable with distribution P_θ, $\theta \epsilon \Omega$. Then the family $\mathcal{P} = \{P_\theta\}$, $\theta \epsilon \Omega$, is said to be *stochastically increasing* (SI) in θ if for $\theta_1 < \theta_2$, the distributions P_{θ_1} and P_{θ_2} are distinct, and for any real number a,

$$P_{\theta_1}[X > a\} \leq P_{\theta_2}[X > a]. \tag{2.1}$$

It is well known that a stronger property is that of *monotone likelihood ratio* (MLR) introduced by Karlin and Rubin (1956) and this is equivalent to the frequency function having *total positivity of order 2* (TP_2). The concept of total positivity is, however, more general and is not restricted to frequency functions (see Karlin, 1968).

A basic result of Lehmann (1959, p. 112, Problem 11) can be stated as follows.

THEOREM 2.1. *Let* $\{P_\theta\}$, $\theta \epsilon \Omega$, *be an SI family of distributions and let* $\psi(x)$ *be a real valued function nondecreasing in* x. *Then* $E_\theta[\psi(X)]$ *is nondecreasing in* θ.

A straight forward generalization of this theorem independently obtained by Alam and Rizvi (1966) and Mahamunulu (1967) is given below.

THEOREM 2.2. *Let* $\{P_\theta\}$, $\theta \epsilon \Omega$ *be an SI family of distributions. Let* $X_1, \ldots, X_k$ *be independent random variables.* X_i *having the distribution* P_{θ_i}, $\theta_i \epsilon \Omega$, $i = 1, \ldots, k$. *Then* $E_\theta \psi(X_1, \ldots, X_k)$ *is nondecreasing in each component of* $\theta = (\theta_1, \ldots, \theta_k)$ *if* $\psi(x_1, \ldots, x_k)$ *is nondecreasing in each of its arguments.*

Theorem 2.2 has been successfully applied to many selection problems. For suitably chosen $\psi(x_1, \ldots, x_k)$, the expectation $E_\theta \psi(X_1, \ldots, X_k)$ becomes the PCS. The monotonicity property of the expectation enables one to obtain the LFC.

Another generalization of Theorem 2.1 in a different direction is due to Gupta and Panchapakesan (1972) who considered a class of subset selection rules defined through a class of functions h. For evaluating the infimum of the PCS, we need to minimize over θ the expectation $E_\theta[\psi(X,\theta)]$. The following theorem of Gupta and Panchapakesan (1972) gives a sufficient condition for the monotonicity of $E_\theta[\psi(X,\theta)]$.

THEOREM 2.3. *Let* $F(\cdot,\theta)$, $\theta \epsilon \Omega$, *be a family of absolutely continuous distributions on the real line* $\mathcal{R}$ *with continuous densities* $f(\cdot,\theta)$ *and let* $\psi(x,\theta)$ *be a bounded real valued function possessing first partial derivatives* ψ_x *and* ψ_θ *with respect to* x *and* θ, *respectively, and satisfying certain regularity conditions C. Then* $E_\theta[\psi(X,\theta)]$ *is nondecreasing in* θ *provided that for all* $\theta \epsilon \Omega$,

$$(2.2) \qquad f(x,\theta)\psi_\theta(x,\theta) - [(\partial/\partial\theta)F(x,\theta)]\psi_x(x,\theta) \geq 0 \qquad \text{a.e. } x,$$

where the regularity conditions C are: (i) *for all* $\theta \epsilon \Omega$, $\psi_x(x,\theta)$ *is Lebesgue integrable on* $\mathcal{R}$; *and* (ii) *for every* $[\theta_1, \theta_2] \subset \Omega$ *and* $\theta_3 \epsilon \Omega$, *there exists* $g(x)$ *depending only on* θ_1, θ_2, θ_3 *such that*

$$|\psi_\theta(x,\theta)f(x,\theta_3) - [(\partial/\partial\theta)F(x,\theta)]\psi_x(x,\theta_3)| \leq g(x)$$

for all $\theta \epsilon [\theta_1,\theta_2]$ *and* $g(x)$ *is Lebesgue integrable on* $\mathcal{R}$.

Remark 2.4. (1) If $\psi(x,\theta) = \psi(x)$ for all $\theta \epsilon \Omega$, the sufficient condition (2.2) reduces to $[(\partial/\partial\theta)F(x,\theta)]\psi_x(x) \leq 0$, which is satisfied by the hypotheses of Theorem 2.1 since $\{F_\theta\}$ is SI and $\psi(x)$ is nondecreasing in x.

(2) For the class of procedures defined by Gupta and Panchapakesan (1972), $\psi(x,\theta) = F(h(x);\theta)$ and (2.2) becomes

$$(2.3) \qquad f(x;\theta)[(\partial/\partial\theta)F(h(x),\theta)] - h'(x)f(h(x);\theta)[(\partial/\partial\theta)F(x,\theta)] \geq 0$$

where $h'(x) = (d/dx)\,h(x)$.

(3) This condition has been specialized to the cases of (i) location parameter, (ii) scale parameter, and (iii) convex mixtures of distributions by Gupta and Panchapakesan for the purposes of specific applications.

(4) An analogue of this theorem for discrete distributions is given by Panchapakesan (1969), who has given in another paper (1978) sufficient conditions for monotonicity when Ω is a countable set.

(5) The monotonicity of $E_\theta[\psi(x,\theta)]$ in θ is strict if strict inequality holds in (2.3) on a set of positive Lebesgue measure.

(6) Obvious modifications in Theorems 2.1 through 2.3 give monotonicity in the opposite direction.

For subset selection rules the expected subset size has been used as a performance characteristic. We naturally want to know the worst configuration in the sense that it maximizes the expected subset size. The following theorem (discussed and proved without a formal statement) of Gupta and Panchapakesan (1972) gives a sufficient condition for the expected subset size to be maximized at an equi-parameter configuration.

THEOREM 2.5. *Let $X_1, \ldots, X_k$ be independent random variables, X_i having an absolutely continuous distribution $F(\cdot,\theta_i)$, $\theta_i \in \Omega$, with continuous density $f(\cdot,\theta_i)$. Let $\psi(x,\theta)$ be a bounded function possessing the first partial derivatives ψ_x and ψ_θ with respect to x and θ, respectively, and satisfying the regularity conditions of Theorem 2.3. Define* $B(\theta_1, \ldots, \theta_k) = \sum_{i=1}^k E_{\theta_i}[\prod_{r=1, r\neq i}^k \psi(X,\theta_r)]$. Then

$$B(\theta \mid \theta_1 \leq \ldots \leq \theta_k) \leq B(\theta \mid \theta_1 = \ldots = \theta_k) \tag{2.4}$$

provided that, for all $\theta_i \leq \theta_j$ and a.e. x, the following holds:

$$[(\partial/\partial\theta_i)\psi(x,\theta_i)]f(x,\theta_j) - [(\partial/\partial x)\psi(x,\theta_j)][(\partial/\partial x)F(x,\theta_i)] \geq 0. \tag{2.5}$$

Remarks 2.6. As in the case of Theorem 2.3, Gupta and Panchapakesan (1972) have specialized this for (i) location parameter, (ii) scale parameter, and (iii) convex mixtures. For their class of procedures, $\psi(x,\theta_i) = F(h(x);\theta_i)$, $i = 1, \ldots, k$. For location and scale parameter cases, the usual choices are $h(x) = x + b$, $b \geq 0$, and $h(x) = ax$, $a \geq 1$, respectively. In these cases, the left-hand side of (2.3) is zero for all x; thereby showing that $E_\theta[\psi(X,\theta)]$ is independent of θ. Further, the condition (2.5) in these cases reduces to the monotone likelihood ratio property, a result directly proved by Gupta (1965). □

Now, we note that Theorem 2.2 is a simple generalization of Theorem 2.1 to $\mathcal{R}^k$, the k-dimensional Euclidian space. We now consider various generalizations of the concepts of stochastic ordering and monotone likelihood ratio to distributions in higher dimensions. To this end, we introduce the following definitions.

Definition 2.7. A function ψ is defined on $\mathcal{R}^k$ is said to be increasing with respect to a partial order relation "$\prec$" if $\mathbf{x}_1 \prec \mathbf{x}_2$ implies $\psi(\mathbf{x}_1) \leq \psi(\mathbf{x}_2)$ for all $\mathbf{x}_1, \mathbf{x}_2 \in \mathcal{R}^k$.

Definition 2.8. A set S in $\mathcal{R}^k$ is said to be an increasing set if its indicator function is increasing, that is, if $\in S$ and $\mathbf{x}_1 \prec \mathbf{x}_2$, then $\mathbf{x}_2 \in S$.

Let $\mathbf{X}$ be a k-dimensional random vector with distribution P_θ in $\mathcal{R}^k$, where $\theta = (\theta_1, \ldots, \theta_k)$. Let $P_\theta(S) = P_\theta(\mathbf{X} \in S)$ for any measurable set S.

Definition 2.9. A distribution P_θ is said to have stochastically increasing property (SIP) in θ if $P_{\theta_1}(S) \leq P_{\theta_2}(S)$ for every monotone increasing measurable set S and for every $\theta_1 \prec \theta_2$.

The following lemma is due to Lehmann (1955).

LEMMA 2.10. *A family of distributions P_θ has SIP in θ if and only if $E_{\theta_1}\psi(\mathbf{X}) \leq E_{\theta_2}\psi(\mathbf{X})$ for all increasing integrable functions $\psi(\mathbf{X})$ and $\theta_1 \prec \theta_2$.*

The following theorem follows easily from Lemma 2.10.

THEOREM 2.11. *Let the distribution of* $\mathbf{X}$ *have SIP in* θ *and let* $\psi(\mathbf{x},\theta)$ *be increasing in* $\mathbf{x}$ *and* θ. *Then* $E_\theta\psi(\mathbf{X},\theta)$ *is increasing in* θ.

When we have independence, it is easily verified that the MLR property implies SIP (Lehmann, 1955). When we deal with correlated random variables $X_1, \ldots, X_k$, it is natural to look for a generalized concept of MLR in higher dimensions. For a density $f(x,\theta)$ in the one-dimensional case, the MLR property says that

$$f(x_1,\theta_1)f(x_2,\theta_2) - f(x_1,\theta_2)f(x_2,\theta_1) \geq 0 \tag{2.6}$$

for every $x_1 \leqslant x_2$ and $\theta_1 \leqslant \theta_2$. We can rewrite (2.6) in the form

$$f(\mathbf{x},\theta) \geqslant f(\mathbf{x},(1,2)\theta) \tag{2.7}$$

where $f(\mathbf{x},\theta) = \Pi_{i=1}^{2} f(x_i;\theta_i)$, $\theta = (\theta_1,(\theta_2)$, and $(1,2)\theta$ is the vector obtained from θ by interchanging θ_1 and θ_2. This provides the motivation for the following definition of Property M by Eaton (1967).

Definition 2.12. A family of real valued density functions $\{f_\alpha(\mathbf{x};\theta)\}$, $\alpha \in \mathcal{A}$, *is said to have Property M if, for each* $\alpha \in \mathcal{A}$ *and for each pair* (i,j), $1 \leqslant i \neq j \leqslant k$, *the following holds:*

$$x_i \geqslant x_j \text{ and } \theta_i \geqslant \theta_j \Rightarrow f_\alpha(\mathbf{x};\theta) \geqslant f_\alpha(\mathbf{x};(i,j)\theta). \tag{2.8}$$

Eaton (1967) has given a necessary and sufficient condition for a class of densities to possess Property M. Bechhofer, Kiefer and Sobel (1968, p. 41) in their monograph on sequential identification and selection rules define a *rankability condition* which is the same as Property M. Hollander, Proschan and Sethuraman (1977) have defined a concept of *decreasing in transposition* (DT) which is also same as Property M; however, their motivation comes from finding classes of functions which share certain properties of Schur functions. In fact, when $g(\mathbf{x},\theta) = h(\mathbf{x}-\theta)$, g is DT on $\mathcal{R}^{2k}$ if and only if h is Schur-concave on $\mathcal{R}^k$. Finally, Marshall and Olkin (1979, p. 160) have also used DT functions but they call them *arrangement increasing* (AI) functions.

It is important to note that, unlike in the case of one-dimensional distributions, Property M does not imply SIP. The following simple example of Hsu (1977) illustrates this point.

Example 2.13. $\mathbf{X} = (X_1,X_2)$ *has the following probability distribution* $f_{(\theta_1,\theta_2)}(x_1,x_2)$ *for four permissible values of* $\theta = (\theta_1,\theta_2)$;

$$f_{(1,2)}(5,6) = 1 - f_{(1,2)}(6,5) = 0.9,$$
$$f_{(2,1)}(5,6) = 1 - f_{(2,1)}(6,5) = 0.1,$$
$$f_{(3,4)}(5,6) = 1 - f_{(3,4)}(6,5) = 0.6,$$
$$f_{(4,3)}(5,6) = 1 - f_{(4,3)}(6,5) = 0.4,$$

Further, we can have SIP without Property M; this is true in one-dimension also. Finally, it is possible to have both SIP and Property M as it is the case with the multinomial distribution.

Another generalization of MLR is given by Gupta and Huang (1980) who obtained for a family of densities having this generalized MLR property an essentially complete class of multiple decision rules.

Definition 2.14. A probability density $f(\mathbf{x};\theta)$ *is said to have a generalized monotone likelihood ratio (GMLR) in* $\mathbf{x}$, *if for every* i *and all fixed* x_j, $j = 1, \ldots, k$, $j \neq i$, $f(\mathbf{x};\theta_1)/f(\mathbf{x};\theta_2)$ *is nondecreasing in* x_i, *where*

$$\theta_m = (\theta_{m1}, \ldots, \theta_{mk}),\ m = 1,2;\ \theta_{1j} = \theta_{2j} \text{ for all } j \neq i, \text{ and } \theta_{1i} > \theta_{2i}.$$

What we have discussed so far are some basic assumptions that are usually made regarding the underlying family, and the monotonicity behavior of the expectations of certain functions. Also of relevance here is the concept of stochastic majorization and inequalities obtained by majorization. One definition of stochastic majorization is to say that **X** is stochastically majorized by **Y** if $E(\psi(\mathbf{X})) \leqslant E(\psi(\mathbf{Y}))$ for all Schur-convex functions ψ; of course, there are other possible definitions (see Marshall and Olkin, 1977, chapter 11). Majorization techniques can be used to show that $E[\psi(\mathbf{X})] \leqslant E[\psi(\mathbf{Y})]$ for several other families of functions ψ. The relevance of these results to selection problems is obvious, when $\psi(\mathbf{X})$ is the indicator function of the event "a correct selection is made." For several useful inequalities in this direction, we refer to Chapters 12 and 13 of Marshall and Olkin (1977).

3. Restricted Families of Distributions. By restricted families of distributions, we mean a family of distributions $\mathcal{F}$ each member of which is partially ordered in a sense with respect to a given distribution G. Such families do arise naturally in reliability studies. More commonly known families of this type are those with increasing failure rate (IFR) and increasing failure rate on the average (IFRA) and naturally those with corresponding decreasing properties. In dealing with such classes we do not know the exact forms of the distributions that belong to $\mathcal{F}$, but we do know the nature of the partial order relation and the distribution G. Precisely this knowledge enables one to find bounds for quantities of interest such as the probability of survival and mean life in terms of G. Inequalities are thus very important in reliability studies. As a matter of no surprise, significant contributions to inequalities for restricted families have been made by researchers in mathematical reliability—Barlow, Marshall and Proschan, to mention a few. Typical of these problems is the use of order statistics. Many important order statistics inequalities that arise in inference problems of reliability are reviewed by Gupta and Panchapakesan (1974).

Selection procedures for restricted families of distributions were first studied by Barlow and Gupta (1969). When we have k populations from $\mathcal{F}$, we can generally evaluate (under some additional assumptions) the infimum of the PCS in terms of the known G by establishing appropriate inequalities. We describe in this section such inequalities and explain the contexts of the selection problems. For the purpose of describing these results, we need to introduce some definitions.

Assuming that all our distributions are absolutely continuous, we now define some of the special order relations of interest to us. F and G denote distribution functions.

Definitions 3.1. (i) *F is said to be convex with respect to* (w.r.t.) *G (written $F \underset{c}{\prec} G$) if and only if $G^{-1}F(x)$ is convex on the support of F.* (ii) *F is star shaped* w.r.t. *$G \underset{*}{\prec} G$) if and only if $F(0) = G(0) = 0$ and $G^{-1}F(x)/x$ is increasing in $x \geqslant 0$ on the support of F. (iii) F is tail ordered* w.r.t. *$G(F \underset{t}{\prec} G)$ if and only if $F(0) = G(0) = ½$, and $G^{-1}F(x) - x$ is nondecreasing on the support of F.*

If $G(x) = 1-e^{-x}$, $x \geqslant 0$, then (i) defines the class of IFR distributions studied by Barlow, Marshall and Proschan (1963) while (ii) defines the class of IFRA distributions studied by Birnbaum, Esary and Marshall (1966). Convex ordering was studied by van Zwet (1964). Doksum (1969) has used the tail ordering. It is easy to verify that the above order relations are all partial order relations. One can also easily see that convex ordering implies star ordering. Without the assumption of the common median zero, the definition (iii) has been used by Bickel and Lehmann (1979) to define an *ordering by spread* with the germinal concept attributed to Brown and Tukey (1946) by them. This kind of ordering has also been perceived by Saunders and Moran (1978) in the context of a neurobiological problem and is called *ordering by dispersion* by them. We now give a formal definition below.

Definition 3.2. G is more dispersed than F($F \underset{d}{\prec} G$) *if*

$$G^{-1}(\beta) - G^{-1}(\alpha) \geq F^{-1}(\beta) - F^{-1}(\alpha) \quad \text{for all } 0 < \alpha < \beta < 1. \tag{3.1}$$

By setting $x = F^{-1}(\beta)$ and $y = F^{-1}(\alpha)$, it is easy to see that (3.1) is equivalent to saying that $G^{-1}F(t) - t$ is increasing in t. However, (3.1) presents the idea more clearly, that is, any two percentage points of G are at least as far apart as the corresponding percentage points of F.

Finally, we define a general partial order relation through a class of real functions introduced by Gupta and Panchapakesan (1974). The star and tail orderings can be obtained as special cases.

Definition 3.3. Let $\mathcal{H} = \{h(x)\}$ be a class of real valued functions $h(x)$ defined on the real line. Let F and G be distributions on the real line such that $F(0) = G(0)$. We say that F is $\mathcal{H}$-ordered w.r.t *G*($F \underset{\mathcal{H}}{\prec} G$) *if*

$$G^{-1}F(h(x)) \geq h(G^{-1}F(x)) \tag{3.2}$$

for all $h \in \mathcal{H}$ and all x on the support of F.

All the order relations we have defined so far can easily be verified to be partial order relations in that they satisfy only reflexivity and transitivity. It can be seen immediately from the above definition that, if $\mathcal{H} = \{ax,\ a \geq 1\}$ and $F(0) = G(0) = 0$, we get the star ordering and that the tail ordering is obtained by taking $\mathcal{H} = \{x+b,\ b \geq 0\}$ and $F(0) = G(0) = \frac{1}{2}$. Also, if we do not include $F(0) = G(0)$ in the definition, then the dispersion ordering become a special case.

The next theorem gives the basic inequality of Gupta and Panchapakesan (1974) and some related inequalities.

THEOREM 3.4. *Let $X_0, X_1, \ldots, X_p (Y_0, Y_1, \ldots, Y_p)$ be independent and identically distributed, each with distribution function F(G), and let $F \underset{\mathcal{H}}{\prec} G$. Then the following inequalities hold.*

(a) $P[h(X_0) \geq X_i, i=1, \ldots, p] \geq P[h(Y_0) \geq Y_i, i=1, \ldots, p]$,

(b) $P[X_0 \geq h(X_i), i=1, \ldots, p] \leq P[Y_0 \geq h(Y_i), i=1, \ldots, p]$,

(c) $P[h(X_0) \leq X_i, i=1, \ldots, p] \leq P[h(Y_0) \leq Y_i, i=1, \ldots, p]$,

(d) $P[X_0 \leq h(X_i), i=1, \ldots, p] \geq P[Y_0 \leq h(Y_i), i=1, \ldots, p]$.

Proof. We will prove (a). The other inequalities can be established similarly. Let $\varphi = G^{-1}F$. Then

$P[h(X_0) \geq X_i, i=1, \ldots, p]$

$= P[\varphi(h(X_0)) \geq \varphi(X_i), i=1, \ldots, p]$, since φ is nondecreasing

$\geq P[h(\varphi(X_0)) \geq \varphi(X_i), i=1, \ldots, p]$, since $F \underset{\mathcal{H}}{\prec} G$

$= P[h(Y_0) \geq Y_i, i=1, \ldots, p]$, since $\varphi(X_i)$ is stochastically equal to Y_i, $i=0, 1, \ldots, p$. □

The inequalities (a) through (d) of the above theorem can be rewirtten respectively as

$$\int F^p(h(x))\, dF(x) \geq \int G^p(h(x))\, dG(x), \tag{3.3}$$

$$\int F^p(h^{-1}(x))\, dF(x) \leq \int G^p(h^{-1}(x))\, dG(x), \tag{3.4}$$

$$\int [1-F(h(x))]^p\, dF(x) \leq \int [1-G(h(x))]^p\, dG(x), \tag{3.5}$$

and

$$\int [1-F(h^{-1}(x))]^p\, dF(x) \geq \int [1-G(h^{-1}(x))]^p\, dG(x), \tag{3.6}$$

where h^{-1} is assumed to exist and the integrals extend over the supports of the relevant distributions. Gupta (1966) obtained essentially these inequalities for any $p > 0$ under a set of hypotheses which amounts to $\mathcal{H}$-ordering. Also, in selection and ranking problems, we typically get the probabilities,

$$P[h(X_0) \geq X_i, i=1, \ldots, p] \text{ and } P[X_0 \leq h(X_i), i=1, \ldots, p].$$

These are same as the left-hand side probabilities in (a) and (d) of Theorem 3.4 if we assume that $h(x) \geq x$. This is satisfied for natural choices of $h(x)$ in the procedures. It should be noted that $h(x) \geq x$ in the special classes of $\mathcal{H}$ yielding star and tail ordering.

Interesting special inequalities are obtained by considering special pairs of F and G in Theorem 3.4. We mention here a few of them relevant to selection rules, thus generally applying inequalities (a) and (d) of Theorem 3.4.

Suppose $X_1, \ldots, X_n$ are i.i.d. with distribution F and $Y_1, \ldots, Y_n$ are i.i.d. with distribution G. Let $F \underset{\mathcal{H}}{\prec} G$. Let $F_{[j]}$ and $G_{[j]}$ denote the cdf's of the jth order statistic of the X_i and the Y_i respectively. Define

$$B_{j,n}(x) = [n!/(j-1)!(n-j)!] \int_0^x u^{j-1}(1-u)^{n-j} du$$

so that

$$F_{[j]}(x) = B_{j,n}(F(x)) = B_{j,n}F(x). \tag{3.7}$$

Since

$$G_{[j]}^{-1}F_{[j]}(x) = [B_{j,n}G]^{-1}B_{j,n}F(x) = G^{-1}F(x), \tag{3.8}$$

we see that order statistics preserve $\mathcal{H}$-ordering. So we get

$$\int F_{[j]}^p(h(x))dF_{[j]}(x) \geq \int G_{[j]}^p(h(x))\, dG_{[j]}(x) \tag{3.9}$$

and

$$\int [1-F_{[j]}(h^{-1}(x))]^p\, dF_{[j]}(x) \geq \int [1-G_{[j]}(h^{-1}(x))]^p\, dG_{[j]}(x). \tag{3.10}$$

Barlow and Gupta (1969) studied subset selection procedures for selecting the distribution with the largest (smallest) α-quantile from $k = p+1$ distributions that are star ordered w.r.t. G. In their procedures, $h(x) = ax$, $a \geq 1$. With this choice of $h(x)$, the right-hand sides of (3.9) and (3.10) become the infimum of PCS in these two cases. Specializing these inequalities further to the case of IFRA distributions, we get the following corollary.

COROLLARY 3.5. *Let $F_{[j]}$ denote the cdf of the jth order statistic in a random sample of n observations from an IFRA distribution F. Then*

$$\int_0^\infty F_{[j]}^p(ax)\, dF_{[j]}(x) \geq \int_0^\infty G_{[j]}^p(ax)\, dG(_{[j]}(x) \tag{3.11}$$

and

$$\int_0^\infty [1-F_{[j]}(x/a)]^p\, dF_{[j]}(x) \geq \int_0^\infty [1-G_{[j]}(x/a)]^p\, dG(_{[j]}(x), \tag{3.12}$$

where

$$G_{[j]}(x) = \Sigma_{t=j}^{n}\binom{n}{t}[1-e^{-x}]^t e^{-(n-t)x} = B_{j,n}(1-e^{-x}). \tag{3.13}$$

Barlow, Gupta and Panchapakesan (1969) have tabulated the values of a^{-1} for which the right-hand sides of (3.11) and (3.12) are equal to P^* (the guaranteed minimum PCS) for selected values of p, n, j and P^*. Gupta and Panchapakesan (1975) studied a similar quantile selection procedure for selecting the largest quantile for distributions that are star ordered w.r.t. the standard normal distribution folded at the origin. In this case, the inequality (3.11) holds with $G_{[j]}(x) = B_{j,n}(2\phi(x)-1)$, where $\phi(x)$ is the standard normal cdf. The values of a^{-1} for which the right-hand side of (3.11) is equal to P^* are tabulated by Gupta and Panchapakesan (1975) for selected values of p, n, j and P^*.

It is easy to verify that the folded normal distribution is an IFR and therefore an IFRA distribution. So we can obtain further inequalities by taking $F_{[j]}(x) = B_{j,n}(2\phi(x)-1)$ in the above corollary.

We can get similar inequalities for F and G such that $F \underset{d}{\prec} G$. We have to take $h(x) = x+b$, $b > 0$, in (3.5) and(3.6). More inequalities can be obtained by considering $F_{[j]}$ and $G_{[j]}$ with special choices of G. Thse inequalities occur in selection procedures of Barlow and Gupta (1969) for selection in terms of medians for a class of distributions (not defined in this paper) and the procedures of Gupta and Panchapakesan (1974) who have used the logistic distribution for G.

Remarks 3.6. Suppose we take $\mathcal{H} = \{ax,\ a \geq 1\}$ in Theorem 3.4. Then, letting $Z_1 = \max\{X_1/X_0, \ldots, X_p/X_0\}$, $Z_2 = \min\{X_1/X_0, \ldots, X_p/X_0\}$, $W_1 = \max\{Y_1/Y_0, \ldots, Y_p/Y_0\}$ and $W_2 = \min\{Y_1/Y_0, \ldots, Y_p/Y_0\}$, we get

$$
\begin{aligned}
&P[Z_1 \leq a] \geq P[W_1 \leq a],\\
&P[Z_1 \leq a^{-1}] \leq P[W_1 \leq a^{-1}],\\
&P[Z_2 \geq a] \leq P[W_2 \geq a],\\
&P[Z_2 \geq a^{-1}] \geq P[W_2 \geq a^{-1}].
\end{aligned}
\tag{3.14}
$$

In other words, we have inequalities for the distribution functions (and hence for quantiles) of the maximum and the minimum of certain correlated ratios of variables with distributions F and G.

In the case of $\mathcal{H} = \{x+b,\ b \geq 0\}$, we let $Z_1' = \max\{X_1-X_0, \ldots, X_p-X_0\}$, $Z_2' = \min\{X_1-X_0, \ldots, X_p-X_0\}$, $W_1' = \max\{Y_1-Y_0, \ldots, Y_p-Y_0\}$ and $W_2' = \min\{Y_1-Y_0, \ldots, Y_p-Y_0\}$. Then, we get

$$
\begin{aligned}
&P[Z_1' \leq b] \geq Pr[W_1' \leq b],\\
&P[Z_1' \leq -b] \leq Pr[W_1' \leq -b],\\
&P[Z_2' \geq b] \leq Pr[W_2' \geq b],\\
&P[Z_2' \geq -b] \geq Pr[W_2' \leq -b],
\end{aligned}
\tag{3.15}
$$

We will come back to these inequalities in Section 4.3.

4. Inequalities for Specific Distributions. We are mainly interested in certain inequalities relating to multivariate normal, multinomial and gamma distributions that occur in ranking and selection problems. Of course, these are of interest otherwise too.

4.1 Inequalities for Multivariate Normal Distribution. A probability expression that occurs frequently in selection problems is $P[X_1 \leq a_1, \ldots, X_k \leq a_k]$ where $X_1, X_2, \ldots, X_k$ are identically distributed but correlated. Most familiar of these and perhaps most often used in practice are the cases where $X_1, \ldots, X_k$ have a joint k-variate normal and t distributions. Evaluation of these probability integrals are difficult to accomplish as k gets large when there is no special pattern of the associated covariance matrix Σ. In such cases, inequalities which give good bounds become more attractive. There are numerous results in the literature in this direction. We will mention here only two results, namely, those of Anderson (1955) and Slepian (1962). For a detailed account of these and other related inequalities and references, the reader is referred to the book of Tong (1980) and the recent survey paper of Eaton (1982). To state Anderson's theorem, let us define a partial ordering $\underset{\sim}{\prec}$ for covariance matrices of the same order by $\psi \underset{\sim}{\prec} \Sigma$ if $\Sigma - \psi$ is positive semidefinite.

THEOREM *4.1* (Anderson, 1955). *Let* $\mathbf{X} = (X_1, \ldots, X_k)$ *and* $\mathbf{Y} = (Y_1, \ldots, Y_k)$ *be k-variate normally distributed random vectors with common mean vector zero and covariance matrices* Σ *and* ψ *respectively and let E be a convex set symmetric about the origin. Then* $\psi \precsim \Sigma$ *implies* $P[\mathbf{Y} \in E] \geq P[\mathbf{X} \in E]$.

As we have pointed out earlier, inequalities have been used in selection problems typically to obtain the infimum of the PCS or a lower bound for it. One result that has been used very often at some stage of the problem is the Slepian inequality stated below.

THEOREM 4.2. (Slepian Inequality). *If* $\mathbf{X} = (X_1, \ldots, X_k)$ *has the k-variate normal distirubiton with nonsingular covariance matrix* $\Sigma = (\sigma_{ij})$, *with* $\sigma_{ii} = 1, i=1, \ldots, k$, *then for any constants* $c_1, \ldots, c_k$, *the probability* $P[X_1 \leqslant c_1, \ldots, X_k \leqslant c_k]$ *is strictly increasing as a function of each* σ_{ij} *for* $i \neq j$. *In particular, if* $\sigma_{ij} > 0, i, j = 1, \ldots, k$, *then*

$$P[X_i \leqslant c_i, i=1, \ldots, k] > \Pi_{i=1}^{k} P[X_i \leq c_i].$$

Motivated by a design problem with a selection and ranking goal, Rinott and Santner (1977) obtained an inequality that combines the aspects of the results of Anderson and Slepian; namely, for $d > 0$.

$$\int\int \phi^n(d+x+\alpha y)\, \phi^m(d+x) d\phi(x) d\phi(y) \leqslant \int \phi^{n+m}(d+x) d\phi(x) \tag{4.1}$$

where $\phi(x)$ is the standard normal cdf, m and n are integers such that $m+1 \geqslant n \geqslant 1$, and all integrals are from $-\infty$ to ∞. It can also be shown that the left-hand side of (2.8) is decreasing in $|\alpha|$ for any $d \geqslant 0$.

4.2 Inequalities for Multinomial Distributions. Let $\mathbf{X} = (X_1, \ldots, X_k)$ have the multinomial distribution given by

$$P(\mathbf{X} = \mathbf{x}) = n! \pi_{i=1}^{k} (\theta_i^{x_i}/x_i!) \tag{4.2}$$

where

$$\mathbf{x} = (x_1, \ldots, x_k), \Sigma_{i=1}^{k} x_i = n \text{ and } \Sigma_{i=1}^{k} \theta_i = 1.$$

Define

$$C(\theta_1, \ldots, \theta_m) = P[X_i \geqslant c_i, i=1, \ldots, m] \tag{4.3}$$

where $\Sigma_{i=1}^{m} c_i \leqslant n$ and $m \leqslant \min(k-1, n)$. The results of Alam (1970a) are summarized in the following theorem.

THEOREM 4.3. $C(\theta_1, \ldots, \theta_m)$ *is nondecreasing in* θ_i, $i=1, 2, \ldots, m$. *Further, for* $c_i = c_j$,

$$C_{ijt}(\theta_1, \ldots, \theta_m) \leqslant C(\theta_1, \ldots, \theta_m) \leqslant C_{ij}(\theta_1, \ldots, \theta_m) \tag{4.4}$$

where $C_{ij}(\theta_1, \ldots, \theta_m)$ *is obtained from* $C(\theta_1, \ldots, \theta_m)$ *by replacing* θ_i *and* θ_j *with their average, and* $C_{ijt}(\theta_1, \ldots, \theta_m)$ *is obtained from* $C(\theta_1, \ldots, \theta_m)$ *by substituting t for* θ_i *and* $\theta_i + \theta_j - t$ *for* θ_j *where* $0 \leqslant t \leqslant \min(\theta_i, \theta_j)$.

Let us assume here and in what follows on multinomial distribution that $\theta_1 \leqslant \theta_2 \leqslant \ldots \leqslant \theta_k$. From Theorem 4.3, we have

$$\begin{aligned} &P[X_1 \geqslant c, \ldots, X_k \geqslant c | \theta_1, \ldots, \theta_1, \theta^*] \\ &\leqslant P[X_1 \geqslant c, \ldots, X_k \geqslant c | \theta_1, \ldots, \theta_k] \\ &\leqslant P[X_1 \geqslant c, \ldots, X_k \geqslant c | \overline{\theta}, \ldots, \overline{\theta}] \end{aligned} \tag{4.5}$$

where $c \leqslant n/k$, $\theta^* = 1-(k-1)\theta_1$ and $\overline{\theta} = \Sigma\theta_i/k$.

Using a representation of $P[X_1 \geqslant c, \ldots, X_k \geqslant c | \theta_1, \ldots, \theta_k]$ in terms of the Dirichlet integral, the inequalities in (4.5) can be obtained as a special case of Theorem 1 of Olkin (1972) which shows the Dirichlet integral to be a Schur function. More general results are available in Marshall and Olkin (1979, p. 306).

Bechhofer, Elmaghrabi and Morse (1959) considered a single sample selection procedure to select the most probable cell with a minimum guaranteed probability P^* that the selected cell will be the one associated with θ_k whenever $\theta_k/\theta_{k-1} \geq \delta > 1$. The rule R proposed by Bechhofer, Elmaghrabi and Morse takes a sample of N observations and selects the cell that yields the largest number of observations using randomization to break ties. The PCS is given by

$$\begin{aligned} \text{PCS} &= P[X_k > X_j, j \neq k] + \tfrac{1}{2}\Sigma_{i \neq k} P[X_k = X_i, X_k > X_j, j \neq i] \\ &\quad + \ldots + 1/k\, P[X_k = X_{k-1} = \ldots = k_1] \\ &= \psi(\theta_1, \theta_2, \ldots, \theta_k), \text{ say.} \end{aligned} \tag{4.6}$$

The following result of Kesten and Morse (1959) gives the LFC.

THEOREM 4.4. *With the above assumptions and notations,*

$$\psi(\theta_1, \ldots, \theta_k \mid \theta_k/\theta_{k-1} \geq \delta > 1) \geq \psi(\theta_1^*, \ldots, \theta_k^*) \tag{4.7}$$

where $\theta_1^* = \ldots = \theta_{k-1}^* = (\delta+k-1)^{-1}$ *and* $\theta_k^* = \delta(\delta+k-1)^{-1}$.

Cacoullos and Sobel (1966) used an inverse sampling rule for the same selection problem. Observations are obtained sequentially until one of the k cells has a prespecified count N. This particular cell is then identified as the most probable cell. In this case, the PCS can be written as a Dirichlet integral and the LFC is the same as that of the single sample procedure of Bechhofer, Elmaghrabi and Morse (1959). Alam (1971) considered a different stopping rule, namely, the observations are taken sequentially until the difference between the highest and the next highest cell counts is equal to r. For $k=2$,

$$\text{PCS} = \lambda^r/(1+\lambda^r) \tag{4.8}$$

where $\lambda = \theta_2/\theta_1$. For $k > 2$, there is no exact result. Alam (1971) gives a lower bound, namely,

$$\text{PCS} \geq 1 - \Sigma_{i=1}^{k-1} \lambda_i^r/(1+\lambda_i^r) \tag{4.9}$$

where $\lambda_i = \theta_i/\theta_k$, $i=1, \ldots, k-1$. An improved bound, namely, $\theta_k^r/\Sigma_{i=1}^k \theta_i^r$, is recently given by Levin and Robbins (1981).

Going back to the single sample procedure of Bechhofer, Elmaghraby and Morse (1959) for selecting the most probable cell, the LFC is sought subject to $\theta_k/\theta_{k-1} \geq \delta > 1$. If we are interested in selecting the least probable cell, then the analogous problem will be to get the LFC whenever $\theta_2/\theta_1 \geq \delta > 1$. The analogous procedure will select the cell with the least count using randomization to break ties. In this case, a minimum P* for the PCS cannot be guaranteed for all P*. This is shown by Alam and Thompson (1972) who proposed a modified indifference-zone. Their rule is still to select the cell with the least count. Let $\psi'(\theta_1, \ldots, \theta_k)$ denote the PCS for this rule. Then their LFC result can be stated as follows:

$$\psi'(\theta_1, \ldots, \theta_k \mid \theta_2 - \theta_1 \geq c) \geq \psi'(\theta_1^*, \ldots, \theta_k^*) \tag{4.8}$$

where $0 < c < (k-1)^{-1}$, $\theta_1^* = [1-(k-1)c]/k$, *and* $\theta_2^* = \ldots = \theta_k^* = (1+c)/k$.

We get additional probability inequalities via subset selection rules. Gupta and Nagel (1967) discussed single sample subset selection rules for selecting the most (least) probable cell. If we denote the cell counts by $X_1, \ldots, X_k$, their rules R_1 and R_2 for the most and the least probable cell, respectively, are as follows:

Select the cell with count X_i if and only if

$$R_1: \qquad X_i \geqslant \max(X_1, \ldots, X_k) - d$$

$$R_2: \qquad X_i \leq \min(X_1, \ldots, X_k) + c$$

where c and *d* are nonnegative integers chosen suitably to guarantee the specified minimum PCS.

The PCS for R_1 is given by

$$P(CS|R_1) = F(k,n,d;\theta_1, \ldots, \theta_k) = \Sigma(\nu_1 \overset{n}{\ldots} \nu_k)\Pi_{i=1}^{k}\theta_i^{\nu_i} \tag{4.9}$$

where the summation is over all k-tuples $(\nu_1, \ldots, \nu_k)$ such that the ν_i are nonnegative $\Sigma\nu_i = n$ and $\nu_i \leqslant \nu_k + d$, $i=1, \ldots, k-1$. In the case of R_2, $P(CS|R_2) = G(k,n,c;\theta_1, \ldots, \theta_k)$ is given by the summation in (4.9) extending over k-tuples $(\nu_1, \ldots, \nu_k)$ such that the ν_i are nonnegative, $\Sigma\nu_i = n$ and $\nu_i \geqslant \nu_1 - c$, $i=2, \ldots, k$.

We now summarize the inequality results of Gupta and Nagel (1967) in the following lemmas and theorems.

LEMMA 4.5. *$F(k,n,d;\theta_1, \ldots, \theta_k)$ satisfies the following inequalities:*

(1) *For* $1 \leqslant i < j < k$, *and* $0 < \epsilon \leqslant \theta_i$,

$$F(k,n,d;\theta_1, \ldots, \theta_k) \geqslant F(k,n,d;\theta_1, \ldots, \theta_i-\epsilon, \ldots, \theta_j+\epsilon, \ldots, \theta_k).$$

(2) *For* $1 \leqslant i < k$, *and* $0 < \epsilon \leqslant \theta_k$,

$$F(k,n,d;\theta_1, \ldots, \theta_k) \geqslant F(k,n,d;\theta_1, \ldots, \theta_i+\epsilon, \ldots, \theta_k-\epsilon).$$

It should be noted that Lemma 4.5 is true even if the order is disturbed in the configurations on the right hand side of the inequalities. The next theorem on the LFC is a consequence of Lemma 4.5.

THEOREM 4.6. *Let r be the smallest integer for which $\theta_i > 0$ and let s be the largest integer such that $\theta_j < \theta_k$. For a configuration minimizing $F(k,n,d;\ \theta_1, \ldots, \theta_k)$ we have $r \geqslant s$. Furthermore, if $r = k-1$, then $r > s$.*

In other words, Theorem 4.6 says that the worst configuration is of the type $(0, \ldots, 0, \alpha, \beta, \ldots, \beta)$, $\alpha \leqslant \beta$.

LEMMA 4.7. *$G(k,n,c;\theta_1, \ldots, \theta_k)$ satisfies the following inequalities:*

(1) *For* $1 < i < j \leqslant k$, *and* $0 < \epsilon \leqslant \theta_i$,

$$G(k,n,c;\theta_1, \ldots, \theta_k) \geqslant G(k,n,c;\theta_1, \ldots, \theta_i-\epsilon, \ldots, \theta_j+\epsilon, \ldots, \theta_k).$$

(2) *For* $1 < j \leqslant k$, *and* $0 < \epsilon \leqslant \theta_j$,

$$G(k,n,c;\theta_1, \ldots, \theta_k) \geqslant G(k,n,c;\theta_1+\epsilon, \ldots, \theta_j-\epsilon, \ldots, \theta_k).$$

As in the case of Lemma 4.5, here also the statements are true even if the order is disturbed in the configuration. The following theorem is a consequence of Lemma 4.7.

THEOREM 4.8. *$G(k,n,c;\ \theta_1, \ldots, \theta_k)$ is minimized at a configuration of the type $\theta_1 = \ldots = \theta_{k-1} \leqslant \theta_k$.*

Now, let us consider s independent multinomial distributions each with k cells. Let $\theta_i = (\theta_{i1}, \ldots, \theta_{ik})$ be the vector of the cell probabilities of π_i, the ith distribution, $i = 1, \ldots, s$. We also assume that, for each i, $\theta_{i1} \leq \ldots \leq \theta_{ik}$.

Definition 4.9. We say that θ_i majorizes $\theta_j (\theta_i \underset{m}{\succ} \theta_j)$ if $\Sigma_{\alpha=r}^{k}\theta_{i\alpha} \geqslant \Sigma_{\alpha=r}^{k}\theta_{j\alpha}$ for $r = 1, \ldots, k$ with equality holding for $r = 1$.

Definition 4.10. If a function φ satisfies the property that $\varphi(\mathbf{x}) \leqslant \varphi(\mathbf{y})(\varphi(\mathbf{x}) \geqslant \varphi(\mathbf{y}))$ whenever $\mathbf{x} \underset{m}{\succ} \mathbf{y}$, then φ is called a Schur-concave (Schur-convex) function.

If $\theta_i \succ_m \theta_j$, it implies that $H(\theta_i) \leq H(\theta_j)$, where $H(\theta_i) = -\Sigma_{\alpha=1}^{k} \theta_{i\alpha} \log \theta_{i\alpha}$ is the *Shannon entropy function* associated with π_i.

Suppose we take n independent observations from each multinomial distribution. Let $x_{i\alpha}$ denote the number of outcomes in the cell with probability $\theta_{i\alpha}$ in π_i, $\alpha = 1, \ldots, k$; $i = 1, \ldots, s$. Define

$$
(4.10) \qquad Q_j(n,k,s;\theta_1, \ldots, \theta_s)
$$
$$
= P\{\varphi(X_{j1}/n, \ldots, X_{jk}/n) \geq \max_{1 \leq \alpha \leq s} \varphi(X_{\alpha 1}/n, \ldots, X_{ik}/n) - d\}, j = 1, \ldots, s,
$$

where φ is a Schur-concave function and $d > 0$.

Gupta and Wong (1976) investigated a subset selection rule for selecting the population whose cell probability vector majorizes that of any other, assuming that one such exists. The special case of $k = 2$ multinomial distributions with the Shannon entropy function as a particular choice of φ was earlier considered by Gupta and Huang (1976). The following theorem relates to the properties of the procedure of Gupta and Wong (1976).

THEOREM 4.11. *If* $\theta_i \succ_m \theta_j$, *then* $Q_i(n,k,s;\theta_1, \ldots, \theta_s) \leq Q_j(n,k,s;\theta_1, \ldots, \theta_s)$. *Further, if* $\theta_i \succ_m \theta_j$ *for all* $j=1, \ldots, s$, *then* $Q_i(n,k,s;\theta_1, \ldots, \theta_s) \geq Q_j(n,k,s;\theta_1 = \ldots = \theta_s)$.

4.3. Inequalities for the Gamma Distribution. Let

$$
(4.11) \qquad \gamma(m,x) = \int_0^x t^{m-1} e^{-t} dt
$$

and

$$
(4.12) \qquad \Gamma(m,x) = \Gamma(m) - \gamma(m,x), m > 0.
$$

Of course,

$$
(4.13) \qquad f(x;m) = [\Gamma(m)]^{-1} e^{-t} t^{m-1}, x \geq 0, m > 0,
$$

is the gamma density where m is the shape parameter. For $0 < m < 1$, continued fraction expansions can be obtained (see, for example, Khovanskii, 1956) for $x^{-m}e^x\gamma(m,x)$ and $x^{-m}e^x\Gamma(m,x)$. Let $P_n(m,x)/Q_n(m,x)$ and $P'_n(m,x)/Q'_n(m,x)$ be the nth convergents of these two expansions respectively.

In the case of $\gamma(m,x)$, Gupta and Waknis (1965) obtained the system of inequalities:

$$
(4.14) \qquad P_n(m,x)/Q_n(m,x) < e^x x^{-m} \gamma(m,x)
$$
$$
< P_n(m,x)/Q_n(m,x) + x^n(n+1+m)/(n+m)_{n+1}(n+1+m-x), n = 1, 2, \ldots,
$$

where $x < n+m+1$ is a necessary restriction only on the inequalities on the right-hand side of (4.14) and where $(n)_r = n(n-1) \ldots (n-r+1), r \geq 1$, and

$$
(4.15) \qquad P_n(m,x)/Q_n(m,x) = m^{-1}[1 + x/(1+m) + \Sigma_{j=2}^{n-1} x^j/(1+m) \ldots (j+m)].
$$

In the case of $\Gamma(a,x)$, the even order convergents form a monotonic increasing sequence and the odd order convergents form a monotonic decreasing sequence, both converging to $e^x x^{-m}\Gamma(m,x)$. So a system of inequalities can be generated by bounding $e^x x^{-m}\Gamma(m,x)$ by successive convergents. These bounds are discussed in Gupta and Waknis (1965). These bounds in turn can be used to get bounds on the integrals

$$
(4.16) \qquad \int_0^\infty F^p(cx;m) f(x;m) dx
$$

and

$$
(4.17) \qquad \int_0^\infty [1-F(bx;m)]^p f(x;m) dx
$$

where $F(x,m)$ is the cdf of the gamma distribution. The integrals (4.16) and (4.17) with $c \geq 1$ and $0 < b \leq 1$ are the infima of the PCS for the subset selection rules of Gupta (1963) and Gupta and Sobel (1962).

Now, let $X_0, X_1, \ldots, X_p$ be independent identically distributed each having a gamma distribution with density $f(x;m)$ given by (4.13). Let

$$Z_1 = \max(X_1/X_0, \ldots, X_p/X_0), \qquad Z_2 = \min(X_1/X_0, \ldots, X_p/X_0). \tag{4.18}$$

Let $G_m(y)$ and $H_m(y)$ denote the cdf's of Z_1 and Z_2, respectively. We note that the integrals in (4.16) and (4.17) are $G_m(c)$ and $1-H_m(b)$, respectively. Alam (1970b) proved that, for $m > 1$, $H_m(y)$ is increasing in m for $y > 1$ and is decreasing in m for $y < 1$. Alam's proof involves a fair amount of analytical details. Further, Alam has no comment on the behavior of $G_m(y)$. The following theorem provides validity of Alam's result for $m > 0$ and establishes the monotonicity behavior of G_m and H_m for a larger class of distributions.

THEOREM 4.12. *Let $X_0, X_1, \ldots, X_p$ be i.i.d. nonnegative random variables each having the distribution F_λ, where $\{F_\lambda\}$ is a star-preceding family in $\lambda \in \Lambda$* [i.e., $F_{\lambda_2} \underset{*}{\prec} F_{\lambda_1}$ for $\lambda_1 < \lambda_2$]. *Let G_λ and H_λ be the cdf's of Z_1 and Z_2 defined in (4.18). Then $G_\lambda(y)$ and $H_\lambda(y)$ are both increasing in λ for $y > 1$ and decreasing in λ for $y < 1$.*

Proof. Since $F_{\lambda_2} \underset{*}{\prec} F_{\lambda_1}$ for $\lambda_1 < \lambda_2$, the conclusions of the theorem follow immediately from the inequalities (3.14) of Remarks 3.6. □

Remarks 4.13. In the case of the gamma family $\{F_m\}$, it is known that F_m convex precedes in $m > 0$; see van Zwet (1964), p. 60. Since the convex ordering implies the star ordering, Alam's result readily follows from Theorem 4.12. As we pointed out earlier, in subset selection procedures, we typically encounter $G_m(y)$ for $y < 1$ and $H_m(y)$ for $y > 1$. That the monotonicity properties of $G_m(y)$ and $H_m(y)$ in these cases can be established by the star-ordering property of the gamma distribution was known though not formally demonstrated; see McDonald (1969) and Panchapakesan (1978) who have given different alternative proofs in the case of integral m for $p = 1$ and $p \geq 1$ respectively. Finally, the monotonicity property of $H_m(y)$ is applied to evaluate the infimum of the PCS for the inverse sampling procedure of Cacoullos and Sobel (1966) for selecting the most probable multinomial cell.

For the Gamma distribution with density in (4.13), let $\xi_m(\alpha)$ and $\xi_m(\beta)$ denote the αth and the βth quantiles, where $0 < \alpha < \beta < 1$. For $m_1 < m_2$, as pointed out earlier, $F_{m_2} \underset{*}{\prec} F_{m_1}$. This is equivalent to

$$F_{m_1}^{-1}(\beta)/F_{m_1}^{-1}(\alpha) \geq F_{m_2}^{-1}(\beta)/F_{m_2}^{-1}(\alpha); \tag{4.19}$$

in other words, $\xi_m(\beta)/\xi_m(\alpha)$ decreases in m, a result obtained by Saunders and Moran (1978) using a fairly long direct method. They have also shown that, for $m_1 < m_2$, F_{m_2} is more dispersed than F_{m_1}; in other words, $\xi_m(\beta) - \xi_m(\alpha)$ increases in m. Also, we can now apply the inequalities in (3.15) to obtain new inequalities for the distribution functions of the maximum and the minimum of certain correlated differences.

4.4 *Inequalities Arising From A Two Stage Selection Procedure.* Gupta and Miescke (1982) studied sequential selection procedures with elimination which are based on vector-at-a-time sampling. They showed that the 'natural' terminal decisions are optimum in a fairly decision-theoretic sense. To describe the inequalities that are obtained, let $\pi_1, \ldots, \pi_k$

be k independent populations with densities f_{θ_i}, $\theta_i \in \Omega$, with respect to the Lebesgue measure on the real line $\mathcal{R}$ or any counting measure on a lattice in $\mathcal{R}$, where $\mathcal{F} = \{f_\theta\}$, $\theta \in \Omega$, is a one-parameter exponential family. Let $X_{i1}, X_{i2}, \ldots$ be independent observations from π_i, $i=1, \ldots, k$. For fixed $n < m$, let $U_i = X_{i1} + \ldots + X_{in}$, $V_i = X_{i,n+1} + \ldots + X_{i,m}$, and $W_i = U_i + V_i$, $i=1, \ldots, k$. Further, for fixed $S \subseteq \{1, \ldots, k\}$, permutation symmetric Borel set $A \subseteq \mathcal{R}^k$, and $i \in s$, define

$$(4.20) \qquad q_i = P_\theta\{V_i = \max_{j \in s} V_j\},$$
$$r_i = P_\theta\{W_i = \max_{j \in S} W_j \,|\, (U_1, \ldots, U_k) \in A\}.$$

THEOREM 4.14. *For* $s = \{i_1, \ldots, i_m\}$

(1) $\theta_{i_j} \leq \theta_{i_t}$ *implies that* $r_{i_j} \leq r_{i_t}$ *and* $q_{i_j} \leq q_{i_t}$, $j,t = 1, \ldots, m; j \neq t$, *and*

(2) *the vector* $r = (r_{i_1}, \ldots, r_{i_m})$ *majorizes the vector* $q = (q_{i_1}, \ldots, r_{i_m})$.

4.5 An Ordering Theorem and Its Specific Application. Let $X_1, \ldots, X_p$ be *conditionally independent and identically distributed* random variables, that is, their joint distribution F is a mixture of the form

$$(4.21) \qquad F(x_1,x_2, \ldots, x_p) = \int \Pi_{i=1}^p F_1(x_i,z)dF_2(z) = E[\Pi_{i=1}^p F_1(x_i,Z)].$$

where F_1 (for given z) and F_2 are distribution functions. The following theorem is due to Tong (1977a).

THEOREM 4.15. *Let* $\mathbf{a} = (a_1,a_2, \ldots, a_r)$ *and* $\mathbf{b} = (b_1,b_2, \ldots, b_r)$ *be vectors of nonnegative integers such that* $a_1 \geqslant a_2 \geqslant \ldots \geqslant a_r$ *and* $b_1 \geqslant b_2 \geqslant \ldots \geqslant b_r$ *with* $\Sigma_{j=1}^r a_j = \Sigma_{j=1}^r b_j = p$. *If* $X_1, \ldots, X_p$ *are conditionally i.i.d. random variables and if* $\mathbf{a} \underset{m}{\succ} \mathbf{b}$, *then*

$$(4.22) \qquad \Pi_{j=1}^r P[X_i \in A, i = 1, \ldots, a_j] \geqslant \Pi_{j=1}^r P[X_i \in A, i = 1, \ldots, b_j]$$

holds for every Borel measurable set A.

Now, if $Y_1,Y_2, \ldots, Y_p$ are i.i.d. random variables and Z is independent of the Y_i, then it is known (see Tong, 1977b, Theorem 2) that $X_i = \phi(Y_i,Z)$, $i = 1,2, \ldots, p$, are conditionally i.i.d. for any Borel measurable function ϕ. This fact together with Theorem 4.15 can be used to obtain bounds on the PCS under the indifference zone formulation and the subset selection approach in view of the fact that the PCS for many classical rules (see Gupta and Panchapakesan, 1972) is a cumulative probability of conditionally i.i.d. random variables.

Tong (1977a) has also discussed a special form of Theorem 4.15 and its applications to several specific multivariate distributions. Applications to multiple decision situations besides selection and ranking are discussed by Tong (1977b).

REFERENCES

ALAM, K. (1970a). Monotonicity properties of the multinomial distribution. *Ann. Math. Statist. 41* 315–317.

ALAM, K. (1970b). A monotonicity property of the distribution of the studentized smallest chi-square. *Ann. Math. Statist. 41* 318–320.

ALAM, K. (1971). On selecting the most probable category. *Technometrics 13* 843–850.

ALAM, K. and RIZVI, M. H. (1966). Selection from multivariate populations. *Ann. Inst. Statist. Math. 18* 307-318.

ALAM, K. and THOMPSON, J. R. (1972). On selecting the least probable multinomial event. *Ann. Math. Statist. 43* 1981–1990.

ANDERSON, T. W. (1955). The integral of a symmetric unimodal function over a symmetric convex set and some probability inequalities. *Proc. Amer. Math. Soc. 6* 170–176.

BARLOW, R. E. and GUPTA, S. S. (1969). Selection procedures for restricted families of distribution. *Ann. Math. Statist. 40* 905–917.

BAROLOW, R. E., GUPTA, S. S. and PANCHAPAKESAN, S. (1969). On the distribution of the maximum and minimum of ratios of order statistics. *Ann. Math. Statist. 40* 918–934.

BARLOW, R. E., MARSHALL, A. W. and PROSCHAN, F. (1963). Properties of probability distributions with monotone hazard rate. *Ann. Math. Statist. 34* 375–389.

BECHHOFER, R. E., ELMAGHRABI, S. and MORSE, N. (1959). A single-sample multiple-decision procedure for selecting the multinomial event which has the highest probability. *Ann. Math. Statist. 30* 102–119.

BECHHOFER, R. E., KIEFER, J. and SOBEL, M. (1968). *Sequential Identification and Ranking Procedures*. The University of Chicago Press, Chicago.

BECKENBACH, E. F. and BELLMAN, R. (1961, 1965). *Inequalities,* 1st ed., 2nd ed. Springer-Verlag, Berlin and New York.

BICKEL, P. J. and LEHMANN, E. L. (1979). Descriptive statistics for nonparametric models IV. Spread. *Contributions to Statistics: Jaroslav Hajek Memorial Volume* (Jana Jureckova, ed.), D. Reidel Publishing Co., Boston, 33–40.

BIRNBAUM, Z. W. ESARY, J. D. and MARSHALL, A. W. (1966). A stochastic characterization of wear-out for components and systems. *Ann. Math. Statist. 37* 816–825.

BROWN, G. and TUKEY, J. W. (1946). Some distributions of sample means. *Ann. Math. Statist. 7* 1–12.

CACOULLOS, T. and SOBEL, M. (1966). An inverse-sampling procedure for selecting the most probable event in a multinomial distribution. *Multivariate Analysis* (P. R. Krishnaiah, ed.), Academic Press, New York, 423–455.

DOKSUM, K. (1969). Starshaped transformations and the power of rank tests. *Ann. Math. Statist. 40* 1167-1176.

EATON, M. L. (1967). Some optimum properties of ranking procedures. *Ann. Math. Statist. 38* 124–137.

EATON, M. L. (1982). A review of selected topics in multivariate probability inequalities. *Ann. Statist. 10* 11–43.

GODWIN, H. A. (1964). *Inequalities on Distribution Functions*. Griffin's Statistical Monographs and Courses, No. 16. Hafner Publishing Company, New York.

GUPTA, S. S. (1963). On a selection and ranking procedure for gamma populations. *Ann. Inst. Statist. Math. 14* 199–216.

GUPTA, S. S. (1965). On some multiple decision (selection and ranking) rules. *Technometrics 7* 225–245.

GUPTA, S. S. (1966). On some selection and ranking procedures for multivariate normal populations using distance functions. *Multivariate Analysis* (P. R. Krishnaiah, ed.), Academic Press, New York, 457–475.

GUPTA, S. S. and HUANG, D-Y. (1976). On subset selection procedures for the entropy function associated with the binomial populations. *Sankhyā Ser. A 38* 153–173.

GUPTA, S. S. and HUANG, D-Y. (1980). An essentially complete class of multiple decision procedures. *J. Statist. Planning and Inference 4* 115–121.

GUPTA, S. S. and MIESCKE, K-J. (1982). On the problem of finding a best population with respect to a control in two stages. *Statistical Decision Theory and Related Topics III,* Vol. 2 (S. S. Gupta and J. Berger, eds.), Academic Press, New York, 473-496.

GUPTA, S. S. and NAGEL, K. (1967). On selection and ranking procedures and order statistics from the multinomial distribution. *Sankhyā Ser. B 29* 1–34.

GUPTA, S. S. and PANCHAPAKESAN, S. (1972). On a class of subset selection procedures. *Ann. Math. Statist. 43* 814–822.

GUPTA, S. S. and PANCHAPAKESAN, S. (1974). Inference for restricted families: (a) multiple decision procedures; (b) order statistics inequalities. *Reliability and Biometry: Statistical Analysis of Lifelength* (F. Proschan and R. J. Serfling, eds.), SIAM, Philadelphia, 503–596.

GUPTA, S. S. and PANCHAPAKESAN, S. (1975). On a quantile selection procedure and associated distribution of ratios of order statistics from a restricted family of probability distributions. *Reliability and Fault Tree Analysis: Theoretical and Applied Aspects of System Reliability and Safety Assessment* (R. E. Barlow, J. B. Fussell and N. D. Singpurwalla, eds.), SIAM, Philadelphia, 557–576.

GUPTA, S. S. and PANCHAPAKESAN, S. (1979). *Multiple Decision Procedures: Theory and Methodology of Selecting and Ranking Populations*. Wiley, New York.

GUPTA, S. S. and SOBEL, M. (1962). On selecting a subset containing the population with the smallest variance. *Biometrika 49* 495–507.

GUPTA, S. S. and WAKNIS, M. R. (1965). A system of inequalities for the incomplete gamma function and the normal integral. *Ann. Math. Statist.* *36* 139–149.

GUPTA, S. S. and WONG, W-Y. (1976). Subset selection procedures for finite schemes in information theory. *Colloquia Mathematica Societatis János Bolyai: 16. Topics in Information Theory,* Conference held at Keszthely, August 1975, 279–291.

HARDY, G. H., LITTLEWOOD, J. E. and PÓLYA, G. (1934, 1952). *Inequalities,* 1st ed., 2nd ed. Cambridge University Press, London and New York.

HOLLANDER, M, PROSCHAN, F. and SETHURAMAN, J. (1977). Functions decreasing in transposition and their applications in ranking problems. *Ann. Statist.* *5* 722–733.

HSU, J. C. (1977). On Some Decision-Theoretic Contributions to the Problem of Subset Selection. Ph.D. Thesis. Department of Statistics, Purdue University, West Lafayette, Indiana.

KARLIN, S. (1968). *Total Positivity, Vol. 1*. Stanford University Press, Stanford.

KARLIN, S. and RUBIN, H. (1956). Distributions possessing a monotone likelihood ratio. *J. Amer. Statist. Assoc.* *51* 637–643.

KAZARINOFF, N. D. (1961). *Geometric Inequalities*. Random House, New Mathematical Library, *4*, New York.

KESTEN, H. and MORSE, N. (1959). A property of the multinomial distribution. *Ann. Math. Statist.* *30* 120–127.

KHOVANSKII, A. N. (1956). *The Application of Continued Fractions*. (Translated by P. Wynn (1962)). P. Noordhoff Ltd., Groningen.

LEHMANN, E. L. (1955). Ordered families of distributions. *Ann. Math. Statist.* *26* 399–419.

LEHMANN, E. L. (1959). *Testing Statistical Hypotheses*. Wiley, New York.

LEVIN, B. and ROBBINS, H. (1981). Selecting the highest probability in binomial or multinomial trials. *Proc. Nat. Acad. Sci. U.S.A.* *78* 4663–4666.

MAHAMUNULU, D. M. (1967). Some fixed-sample ranking and selection problems. *Ann. Math. Statist.* *38* 1079–1091.

MARSHALL, A. W. and OLKIN, I. (1979). *Inequalities: Theory of Majorization and Its Applications*. Academic Press, New York.

MCDONALD, G. C. (1969). A note on montonicity in the gamma distributions. *Res. Publ. GMR-930*. General Motors Research Laboratories, Warren, Michigan.

MITRINOVIĆ, D. S. (1964). *Elementary Inequalities*. P. Noordhoff, Groningen.

MITRINOVIĆ, D. S. (1970). *Analytic Inequalities*. Springer-Verlag, Berlin and New York.

OLKIN, I. (1972). Monotonicity properties of Dirichlet integrals with applications to the multinomial distribution and the analysis of variance. *Biometrika* *59* 303–307.

PANCHAPAKESAN, S. (1969). Some Contributions to Multiple Decision (Selection and Ranking) Procedures. Ph.D. Thesis. (Mimeo. Ser. No. 192). Dept. of Statistics, Purdue University, West Lafayette.

PANCHAPAKESAN, S. (1978). On a monotonicity property relating to the gamma distributions. *J. Chinese Statist. Assoc.* *18* 6003–6005.

PÓLYA, G. and SZEGÖ, G. (1951). *Isoperimetric Inequalities in Mathematical Physics*. Ann. of Math. Studies, No. 27, Princeton University Press, Princeton.

RINOTT, Y. and SANTNER, T. J. (1977). An inequality for multivariate normal probabilities with application to a design problem. *Ann. Statist.* *5* 1228–1234.

SAUNDERS, I. W. and MORAN, P. A. P. (1978). On the quantiles of the gamma and F distributions. *J. Appl. Prob.* *15* 426–432.

SHISHA, O. (Ed.) (1967). *Inequalities*. Academic Press, New York.

SLEPIAN, D. (1962). The one-sided barrier problem for Gaussian noise. *Bell System Tech. J.* *41* 463–501.

TONG, Y. L. (1977a). An ordering theorem for conditionally independent and identically distributed random variables. *Ann. Statist.* *5* 274–277.

TONG, Y. L. (1977b). Applications of a probability inequality to ranking and selection and other related problems. *Commun. Statist.-Theor. Meth.* *A6* 1105–1120.

TONG, Y. L. (1980). *Probability Inequalities in Multivariate Distributions*. Academic Press, New York.

VAN ZWET, W. R. (1964). *Convex Transformations of Random Variables*. Mathematical Center, Amsterdam.

Inequalities in Statistics and Probability
IMS Lecture Notes–Monograph Series Vol. 5 (1984), 228–235

DUAL CONVEX CONES OF ORDER RESTRCITIONS WITH APPLICATIONS[1]

BY RICHARD L. DYKSTRA
University of Iowa

The concept of closed convex cones in finite dimensional Euclidian space and their duals has proven to be a useful construct. Here dual cones are exhibited for specific closed, convex cones including those pertaining to starshaped orderings and concave (convex) functions.

Applications include finding projections involving starshaped orderings, generalizations of Chebyshev's (Kimball's) inequality, an inequality for concave (convex) functions and a characterization of certain kinds of positive dependence.

1. Introduction. Several authors have made extensive use of the concept of convex cones and their duals in $\mathcal{R}^n$. Among these are Rockafellar (1970), Robertson and Wright (1981), and Barlow and Brunk (1972). Here we wish to specifically exhibit certain convex cones and their duals and discuss the implications.

To be precise, we call $K \subset R^n$ a convex cone if (a) $\mathbf{x},\mathbf{y} \in K \Rightarrow \mathbf{x} + \mathbf{Y} \in K$, and (b) $\mathbf{x} \in K$, $a \geqslant 0 \Rightarrow a\,\mathbf{x} \in K$. Of course if K is a convex cone, so is $-K = \{\mathbf{x}: -\mathbf{x} \in K\}$ which we will call the "negative" of K.

Another important convex cone induced by K is the "dual" of K. For a fixed positive vector $\mathbf{w}$, the dual of K is given by

$$K^{\mathbf{w}*} = \{\mathbf{y}:(\mathbf{x},\mathbf{y}) = \Sigma_{i=1}^n x_i y_i w_i \leqslant 0 \text{ for all } \mathbf{x} \in K\}.$$

(Some authors prefer the term "polar" to "dual." Some also define the dual as the negative of our dual.) Of course if K is closed, then $(K^{\mathbf{w}*})^{\mathbf{w}*} = K$. It is evident that if $K_1 \subset K_2$, then $K_1^{\mathbf{w}*} \supset K_2^{\mathbf{w}*}$.

New convex cones can be formed from existing cones in several ways. Two important methods are through intersections and direct sums.

If the closed, convex cones $K_1, \ldots, K_n$ are sufficiently nice (say finitely generated), the direct sum $\Sigma_{i=1}^n K_i = \{\Sigma_{i=1}^n \mathbf{x}_i | \mathbf{x}_i \epsilon K_i,\ i=1, \ldots, n\}$ is also a closed, convex cone. However, in general the closure property is not guaranteed (see Hestenes (1975), pp. 196–198). Nevertheless, intersections and direct sums of closed, convex cones are closely related because it is always true that $(\Sigma_{i=1}^n K_i)^{\mathbf{w}*} = \bigcap_{i=1}^n K_i^{\mathbf{w}*}$ and

$$(1.1) \qquad (\textstyle\bigcap_{i=1}^n K_i)^{\mathbf{w}*} = \Sigma_{i=1}^n K_i^{\mathbf{w}*}$$

if the latter cone is closed. This is guaranteed if the relative interiors of the K_i have a point in common (see Rockafellar (1970), p. 146) or, as we said, if the $K_i^{\mathbf{w}*}$ are finitely generated. (1.1) is equivalent to the well-known Farkas' Lemma if the K_i are generated by a single vector.

An important cone, especially in the area of isotone regression, is the cone of vectors which are nondecreasing, i.e.

$$(1.2) \qquad K_I = \{\mathbf{x} | x_1 \leqslant x_2 \leqslant \ldots \leqslant x_n\}.$$

The dual cone here, as discussed in Barlow and Brunk (1972), is

[1] This work was partially supported by ONR Contract N00017–78–C–0655.

AMS 1980 subject classifications. Primary 52A40, secondary 52A20.

Key words and phrases: Convex cones, dual cones, polars, projections, starshaped orderings, concave (convex) functions, Chebyshev's (Kimball's) Inequality, positively dependent random variables, order restrictions.

(1.3) $$K_I^{\mathbf{w}*} = \{\mathbf{y}|\Sigma_{j=1}^{i} y_j w_j \geq 0, i = 1, \ldots, n-1, \Sigma_{j=1}^{n} y_j w_j = 0\}.$$

We note in passing that the important concept of majorization as discussed extensively in Marshall and Olkin (1979) is closely connected with the cone in (1.3). If the vectors $\mathbf{x}$ and $\mathbf{y}$ are each ordered from largest to smallest to form $\tilde{\mathbf{x}}$ and $\tilde{\mathbf{y}}$, $\mathbf{x}$ majorizes $\mathbf{y}$ iff, $\tilde{\mathbf{x}} - \tilde{\mathbf{y}} \in K_I^{\mathbf{1}*}$. (We let $\mathbf{1}$ denote a vector containing all 1's.) Further discussion of such cone orderings is given in Marshall, Walkup, and Wets (1967).

If the cone specified in (1.2) is modified to require that it contain only nonnegative vectors, i.e., $K = \{\mathbf{x}|0 \leq x_1 \leq x_2 \leq \ldots \leq x_n\}$, the dual is equivalent to that given in (1.3) with a modification of the last equality. In this case,

$$K^{\mathbf{w}*} = \{\mathbf{y}|\Sigma_{i=1}^{n} y_j w_j \geq 0, i = 1, \ldots, n\}.$$

Much of our interest in dual cones hinges on a duality result discussed in Barlow and Brunk (1972). In particular if g^* solves the problem

(1.4) $$\underset{\mathbf{x}\in K}{\text{Minimize}}\ \Sigma_{j=1}^{n} (g_i - x_i)^2 w_i$$

where K is a closed convex cone, then $\mathbf{g} - \tilde{\mathbf{g}}^*$ solves

(1.5) $$\underset{\mathbf{x}\in K^{\mathbf{w}*}}{\text{Minimize}}\ \Sigma_{i=1}^{n} (g_i - x_i)^2 w_i.$$

Robertson and Wright (1980) make extensive use of this duality in dealing with stochastic ordering restrictions for multinomial parameters. This duality is also important in deriving distributional theory, i.e., see Robertson and Wegman (1978).

2. The Starshaped Ordering. An interesting order restriction is that a vector be starshaped. Shaked (1979) defines a vector $\mathbf{x}$ to be lower (upper) starshaped with respect to the positive weights $\mathbf{w}$ if $\bar{x}_1 \geq \bar{x}_2 \geq \ldots \geq \bar{x}_n \geq 0$ $(0 \leq \bar{x}_1 \leq x_2 \ldots \leq \bar{x}_n)$ where

(2.1) $$\bar{x}_i = \Sigma_{j=1}^{i} x_j w_j / \Sigma_{j=1}^{i} w_j.$$

Shaked is concerned with finding maximum likelihood estimates of Poisson and normal means which must satisfy starshaped restrictions.

Dykstra and Robertson (1982) use the term "decreasing (increasing) on the average" when the nonnegativity restrictions in (2.1) are omitted, and are concerned with such restrictions when testing for trend.

Surprisingly the dual cone of "increasing on the average" vectors is closely associated with the cone of "decreasing on the average" vectors.

THEOREM 2.1. *If* $K_{IA} = \{\mathbf{x}|\bar{x}_1 \leq \bar{x}_2 \leq \ldots \leq \bar{x}_n\}$, *then* $K_{IA}^{\mathbf{w}*} = \{\mathbf{y}|\bar{y}_1 \geq \bar{y}_2 \geq \ldots \geq \bar{y}_n = 0\}$.

Proof. Note that we can write

$$K_{IA} = \{\mathbf{x}|\bar{x}_i - \bar{x}_{i+1} \leq 0, i = 1, \ldots, n-1\} = \bigcap_1^{n-1} K_i$$

where

(2.2) $$K_i = \{\mathbf{x}|\bar{x}_1 - \bar{x}_{i+1} \leq 0\}.$$

Now we claim that

(2.3) $$H_i = \{\mathbf{y}|0 \leq y_1 = y_2 = \ldots = y_i, \Sigma_1^{i+1} y_j w_j = 0, y_j = 0, j > i+1\}$$

is actually $K_i^{\mathbf{w}*}$. If $\mathbf{y} \in H_i$, then $y_{i+1} = -y_1 W_i w_{i+1}^{-1}$ where $W_i = \Sigma_1^i w_j$.

If $\mathbf{x} \in K_i$ and $\mathbf{y} \in H_i$, then

$$(\mathbf{x},\mathbf{y}) = \Sigma_1^n x_j y_j w_j = y_1[\Sigma_1^n x_j w_j - x_{i+1} W_i] \leq 0$$

by (2.2) and (2.3). Since $H_i^{\mathbf{w}*}$ is clearly K_i, we have that $H_i = K_i^{\mathbf{w}*}$.

Since from (1.1)

$$(\bigcap_{i=1}^{n-1} K_i)^{\mathbf{w}*} = \Sigma_{i=1}^{n-1} K_i^{\mathbf{w}*},$$

we need to show that $\Sigma_{i=1}^{n-1} K_i^{\mathbf{w}*} = \{\mathbf{y}|\bar{y}_1 \geq \bar{y}_2 \geq \ldots \geq \bar{y}_n = 0\}$.

First assume $\mathbf{x}_i \in K_i^{\mathbf{w}*}$, $i = 1, \ldots, n-1$. Then we may write

$$\mathbf{x}_1 = (x_1, -x_1 W_1 w_2^{-1}, 0, \ldots, 0) \qquad (x_1 \geq 0)$$
$$\mathbf{x}_2 = (x_2, x_2, -x_2 W_2 w_3^{-1}, 0, \ldots, 0) \qquad (x_2 \geq 0)$$
$$\ldots$$
$$\mathbf{x}_{n-1} = (x_{n-1}, x_{n-1}, \ldots, x_{n-1}, -x_{n-1} W_{n-1} w_n^{-1}) \qquad (x_{n-1} \geq 0)$$

After adding coordinates we see that

$$(\Sigma_{i=1}^{n-1}\mathbf{x}_j)_i - (\Sigma_{i=1}^{n-1}\mathbf{x}_j)_{i+1} = x_i \geq 0, i = 1, \ldots, n-1$$

and $(\Sigma_{i=1}^{n-1}\mathbf{x}_j)_n = 0$. Thus $\Sigma_{i=1}^{n-1} K_i^{\mathbf{w}} \subset \{\mathbf{y}: \bar{y}_1 \geq \bar{y}_2 \geq \ldots \geq 0\}$.

Conversely, consider $\mathbf{y} = (y_1, \ldots, y_n)$ such that $\bar{y}_1 \geq \bar{y}_2 \geq \ldots \geq \bar{y}_n = 0$. Recalling that $w_i = \Sigma_1^i W_j$, we partition $\mathbf{y}$ as follows:

$$\mathbf{x}_1 = (-w_2 W_1^{-1} z_1, z_1, 0, \ldots, 0)$$
$$\mathbf{x}_2 = (-w_3 W_2^{-1} z_2, -w_3 W_2^{-1} z_2, z_1, 0, \ldots, 0)$$
$$\ldots$$
$$\mathbf{x}_{n-1} = (-w_n W_{n-1}^{-1} z_{n-1}, \ldots, -w_n W_{n-1}^{-1} z_{n-1}, z_{n-1})$$

where

$$\mathbf{z}_{i-1} = y_i + W_i^{-1}\Sigma_{i+1}^n y_j w_j.$$

It can be verified that the i-th column of the above array sums to y_i and that each row is such that $\Sigma_1^n x_{ij} w_j = 0$.

Finally we note that

$$\bar{y}_{i-1} \geq \bar{y}_i \Longleftrightarrow W_i \Sigma_1^{i-1} y_j w_j \geq w_{i-1}\Sigma_1^i y_j w_j \Longleftrightarrow \Sigma_1^{i-1} y_j w_j \geq W_{i-1} y_i.$$

Therefore

$$0 = W_i^{-1}[\Sigma_1^{i-1} y_j w_j + \Sigma_i^n y_j w_j] \geq W_i^{-1}[W_{i-1} y_i + \Sigma_i^n y_j w_j]$$
$$= y_i + W_i^{-1}\Sigma_{i+1}^n y_j w_j = z_{i-1},$$

so that $-w_{i+1} W_i^{-1} z_i \geq 0$, and hence $\mathbf{x}_i \in K_i^{\mathbf{w}*}$. Thus we have that

$$\{\mathbf{y}|\bar{y}_1 \geq \bar{y}_2 \geq \ldots \geq \bar{y}_n = 0\} \subset K_1^{\mathbf{w}*} + \ldots + K_{n-1}^{\mathbf{w}*}$$

so that equality holds. □

The dual cones of lower and upper starshaped vectors discussed by Shaked (1979) can also be found. First we handle the lower starshaped vector.

COROLLARY 2.2. *If* $K_{LS} = \{\mathbf{x}|\bar{x}_1 \geq \bar{x}_2 \geq \ldots \geq \bar{x}_n \geq 0\}$, *then*

$$K_{LS}^{\mathbf{w}*} = \{\mathbf{y}|\bar{y}_1 \leq \bar{y}_2 \leq \ldots \leq \bar{y}_n \leq 0\}.$$

(Note that this dual also has the property that $K_{LS}^{\mathbf{w}*} = -K_{LS}$).

Proof. Note that

$$K_{LS} = K_{DA} \quad \{\mathbf{x}|\Sigma_{i=1}^n x_j w_j \geq 0\}.$$

Since the dual of this last cone is

$$\{\mathbf{y}|y_1 = y_2 = \ldots = y_n \leq 0\}, \tag{2.4}$$

the identity in (1.1) implies that $K^{\mathbf{w}*}_{LS}$ is the direct sum of $K^{\mathbf{w}*}_{DA}$ and the cone in (2.4). This can be shown to be the desired cone. □

The dual cone of the upper starshaped vectors is not quite as elegant.

COROLLARY 2.3. *If* $K_{US} = \{\mathbf{x}|0 \leqslant \bar{x}_1 \leqslant \bar{x}_2 \leqslant \ldots \leqslant \bar{x}_n\}$, *then*
$K^{\mathbf{w}*}_{US} = \{\mathbf{y}|y_{i+1} - \bar{y}_i \leqslant (\Sigma_1^i w_j)^{-1}\Sigma_{j=1}^n y_j w_j \leqslant 0 \textit{ for } i = 1, \ldots, n-1\}$.

Proof. The proof follows by writing

$$K_{US} = K_{IA} \cap \{\mathbf{x}: x_1 \geqslant 0\},$$

recognizing that

$$\{\mathbf{x}|x_1 \geqslant 0\}^{\mathbf{w}*} = \{\mathbf{y}|y_1 \leqslant 0, y_2 = y_3 = \ldots = y_n = 0\}$$

and using (1.1) and Theorem 2.1. □

3. The Concave Ordering. A frequently occurring closed convex cone in $\mathcal{R}^n$ is the class of concave (convex) functions $K_{CC}(K_{CV})$ defined on the set of real numbers $\{x_1, \ldots, x_n\}$. Thus a point $\mathbf{y} = (y_1, \ldots, y_n) \in \mathcal{R}^n$ is interpreted as the function whose image of x_i is y_i. If we let $\Delta y_i = y_{i+1} - y_i$ and $\Delta x_i = x_{i+1} - x_i$, we can write $K_{CC} = \bigcap_{i=1}^{n-2} H_i$ where $H_i = \{\mathbf{y}|\Delta y_i/\Delta x_i \geqslant \Delta y_{i+1}/\Delta x_{i+1}\}$. The dual cone of $K_{CC}(K_{CV})$ is surprisingly tractable.

THEOREM 3.1. *The dual cone of the set of concave functions on* $\{x_1, \ldots, x_n\}$ *is given by*

$$K^{\mathbf{w}*}_{CC} = \left\{\mathbf{z}|\Sigma_{i=1}^n z_i w_i = 0, \Sigma_{i=1}^{n-r-1}(x_{n-r} - x_i) z_i w_i \begin{cases} = 0, r = 0 \\ \geqslant 0, r = 1, 2, \cdots, n-2 \end{cases}\right\}.$$

Our proof proceeds similarly to Theorem 2.1 and is not given. The theorem is closely related to a result of Brunk (1956).

4. Applications. Of course by their very definitions, a convex cone K and its dual $K^{\mathbf{w}*}$ give rise to natural inequalities. In particular, if $\mathbf{x} \in K$ and $\mathbf{y} - \mathbf{z} \in K^{\mathbf{w}*}$, then

$$\Sigma_1^n x_j(y_j - z_j) w_j \leqslant 0. \tag{4.1}$$

This has some straightforward implications in terms of sample covariances by taking $\mathbf{w} = \mathbf{1}$.

COROLLARY 4.1. *Suppose* $\mathbf{x}, \mathbf{y}$ *and* $\mathbf{z}$ *are vectors in* $\mathcal{R}^n$. *If*

$$i^{-1}\Sigma_{j=1}^i x_j \geqslant (i+1)^{-1}\Sigma_{j=1}^{i+1} x_j,\ i = 1, \ldots, n-1, \tag{4.2}$$

and

$$i^{-1}\Sigma_{j=1}^i (y_j - z_j) \geqslant (i+1)^{-1}\Sigma_{j=1}^{i+1} y_j - z_j,\ i = 1, \ldots, n-1, \tag{4.3}$$

then the sample covariance of $(\mathbf{x},\mathbf{y})$ *is at least as large as the sample covariance of* $(\mathbf{x},\mathbf{z})$.

Proof. Condition (4.2) states that $\mathbf{x} \in K_{DA}$. Condition (4.3) implies that $\mathbf{z} - \mathbf{y} \in K_{IA}$ which is equivalent to saying $(\mathbf{z} - \mathbf{y}) - (\bar{\mathbf{z}} - \bar{\mathbf{y}}) \in K^{\mathbf{1}*}_{DA}$ (where $\bar{\mathbf{a}} = (\bar{a}, \bar{a}, \ldots, \bar{a})$). Thus

$$\Sigma_1^n (x_i - \bar{x})(z_i - \bar{z}) = \Sigma_1^n x_i(z_i - \bar{z}) \leqslant \Sigma_1^n x_i(y_i - \bar{y}) = \Sigma_1^n (x_i - \bar{x})(y_i - \bar{y}).$$ □

Of course if $\mathbf{z} = \mathbf{0}$, this result is equivalent to saying that if $\mathbf{x}, \mathbf{y} \in K_{DA}(K_{IA})$ then $(\mathbf{x}, \mathbf{y}) \geqslant n\bar{x}\bar{y}$. Of course since $K_{DA} = -K_{IA}$, if $x \in K_{IA}$ and $y \in K_{DA}$ (or vice versa) $(\mathbf{x}, \mathbf{y}) \leqslant n\bar{x}\bar{y}$. These inequalities are as strong as possible in the sense if $\mathbf{x} \notin K_{DA}(K_{IA})$, one can find a $\mathbf{y} \in K_{DA}(K_{IA})$ such that $(\mathbf{x},\mathbf{y}) < (>) n\bar{x}\bar{y}$. Note that the above inequality generalizes the well known result for nondecreasing (nonincreasing) vectors.

Another application concerns Shaked's paper (1979). In this paper Shaked wants to find a weighted least squares projection of say $\mathbf{g}$ onto the cone K_{LS}. However Shaked actually finds the projection, say $\mathbf{g}^*$, onto the cone K_{DA} and hopes that $\mathbf{g}^*$ is in K_{LS} (in which case $\mathbf{g}^*$ is also the projection onto K_{LS}). However, if $\mathbf{g}^*$ is *not* in K_{LS}, i.e., $\Sigma_1^n g_j^* w_j < 0$, one can say that the true projection $\hat{\mathbf{g}}$ has the property that $\Sigma_1^n \hat{g}_j w_j = 0$ (see page 89, Barlow, Bartholomew, Bremner and Brunk (1972)). In this case, we know that $\hat{\mathbf{g}}$ must be the projection onto the dual of K_{IA}.

In this event (see (1.5)), $\hat{\mathbf{g}} = \mathbf{g} - \bar{\mathbf{g}}$ where $\bar{\mathbf{g}}$ is the projection of g onto K_{IA} which is a problem that Shaked also solves. From Shaked's solution we can verify that $\mathbf{g} - \bar{\mathbf{g}} = \mathbf{g}^* \cdot \bar{\mathbf{g}}$. Thus the projection onto K_{LS} is given by

$$\hat{\mathbf{g}} = \begin{cases} \mathbf{g}^*, & \text{if } \Sigma_1^n g_j^* w_j \geq 0 \\ \mathbf{g}^* - \bar{\mathbf{g}}, & \text{if } \Sigma_1^n g_j^* w_j < 0. \end{cases}$$

A useful inequality attributed to Chebyshev (see Hardy, Littlewood and Pólya (1959, p. 43)) and discussed and generalized in various places such as Horn (1979), Kimball (1951), and Dykstra, Hewett and Thompson (1973) concerns the expected value of a product of monotone functions of a random variable. Thus, for example, if f, g are nondecreasing (nonincreasing) functions,

$$Ef(X) \cdot g(X) \geq Ef(X) \cdot Eg(X) \tag{4.4}$$

assuming the expectations are defined. We can develop similar types of inequalities based upon closed convex cones and their duals.

COROLLARY 4.2. *If f, g are real valued functions in the class*

$$A_X = \{f: f(X) \text{ is integrable, } E[f(X)I_{[X \leq x]}]/P(X \leq x) \text{ is nondecreasing over } \{x: P(X \leq x) > 0\}\},$$

then

$$Ef(X)g(X) \geq Ef(X) \cdot Eg(X).$$

Proof. Suppose first that X is finitely discrete on the set $\{x_1, \ldots, x_n\}$. If we let $\mathbf{w} = (w_1, \ldots, w_n)$ where $w_i = P(X = x_i)$, then the condition that $f \in A_X$ is equivalent to saying

$$(f(x_1), f(x_2), \ldots, f(x_n)) \in K_{IA}.$$

If $g \in A_X$, $Eg(X) - g$ must belong to $K^{\mathbf{w}*}_{IA}$ and the result follows.

In the general case, we let $x_{n,j}$, $j = 0, \ldots, k(n)$ be a series of nested partitions covering the support of X which generate the Borel sets in the support of X. We define

$$f_n(X) = \Sigma_{i=1}^{k(n)} E(f(X)I_{A_{n,i}}(X)) \cdot I_{A_{n,i}}(x)/w_{n,i}$$
$$g_n(X) = \Sigma_{i=1}^{k(n)} E(g(X)I_{A_{n,i}}(x))\, I_{A_{n,i}}(X)/w_{n,i}$$

where $A_{n,i} = (x_{n,i-1}, x_{i,i}]$ and $w_{n,i} = P(X \in A_{n,i})$. (We take $x_{n,0} = -\infty$.)

Viewing $f_n(X)$ and $g_n(X)$ as conditional expectations, we can use Theorem 5.21 of Breiman (1968) to argue that $f_n(X) \xrightarrow[\text{a.s.}]{L_1} f(X)$ and $g_n(X) \xrightarrow[\text{a.s.}]{L_1} g(X)$.

We have from the first part of the proof that

$$Ef(X)\, Eg(X) = Ef_n(X)\, \mathrm{E}g_n(X) \leq Ef_n(X)g_n(X) \quad \text{for all } n.$$

Therefore if f is bounded above, by Fatou's lemma,

$$\begin{aligned} Ef(X)Eg(X) &\leq \limsup Ef_n(X) \cdot g_n(X) \\ &\leq E \limsup f_n(X)g_n(X) = Ef(X)g(X). \end{aligned} \tag{4.5}$$

Finally, noting that if $h \in A_X$, so does $\min\{h, c\}$ for any positive constant c, we have the desired result for $\min\{f, c\}$ and $\min\{g, c\}$. Note that (4.5) guarantees that $E[f(X)g(X)^-] < \infty$. If $E[f(X)g(X)^+] = \infty$, the desired result clearly holds, so we may assume that $f(X)g(X)$

is integrable. Finally, letting $c \to \infty$ and using the Dominated Convergence Theorem on each side concludes the proof. □

We can obtain similar type inequalities by working with other cones and their duals. For example, we can establish the following corollary which is closely related to the basic lemma of Marshall and Proschan (1970).

COROLLARY 4.3. *If f is a real-valued nondecreasing function with $f(X)$ integrable and g is a real-valued function in the class*

$$B_x = \{g\colon g(X) \textit{ is integrable}, E[g(X)I_{(X \le x)}] \le Eg(X) \quad \textit{for all } x\},$$

then

$$Ef(X)g(X) \ge Ef(X)g(X).$$

The proof follows the lines of Corollary 4.2 and is not given.

Note that if we define the class of real-valued functions

$$C_X = \{g\colon g(X) \quad \text{is integrable and } g \text{ is nondecreasing}\},$$

then $C_X \subset A_X \subset B_X$. Thus both Corollary 4.2 and Corollary 4.3 generalize the Chebyshev inequality (4.4). The results of this section enable us to obtain some insight into certain types of positive dependence as discussed in Lehmann (1966) and elsewhere.

Let us say that the random variables (X, Y) satisfy the following kinds of positive dependence: (1) Type I if $P[X \le x,\ Y \le y] \ge P[X \le x]P[Y \le y]$ for all x, y, (2) Type II if $P[Y \ge y | X \le x]$ is nondecreasing in x for all y, and (3) Type III if $P[Y \ge y | X = x]$ is nondecreasing in x for all y. Assuming that all quantities are defined, each of the above types of dependence can be characterized by the inequality

$$Ef(X) \cdot g(X) \ge Ef(X) \cdot Eg(X) \tag{4.6}$$

as shown in the following Theorem.

THEOREM 4.1. *Assume $g \in C_y$. Then (X, Y) exhibits Type I, II, or III dependence iff (4.6) holds for all $f \in C_X$, A_X, or B_X respectively.*

Proof. The result for Type I dependence is handled in Lehmann (1966). For Type II, let $h(t) = P[Y \ge y | X = t]$. Then $h \in A_X$ iff

$$E[P\{Y \ge y | X\} I_{(X \le x)}]/P[X \le x] = P[Y \ge y | X \le x]$$

is nondecreasing in x. Thus if f also belongs to A_X, we have by Corollary 4.2

$$\begin{aligned} E[f(X)h(X)] &= Ef(X) \cdot I_{(Y \ge y)} \\ &\ge Ef(X) \cdot P(Y \ge y), \quad \text{for all } y. \end{aligned} \tag{4.7}$$

Thus

$$Ef(X)\Sigma a_i I_{(Y \ge y)} \ge Ef(X)\Sigma a_i P(Y \ge y_i)$$

for all nonnegative a_i. A passage to the limit will imply the desired result for a nonnegative, nondecreasing g in C_Y from which the result follows. If $P[Y \ge y | X \le x]$ is *not* nondecreasing in x, then $h \notin A_X$ which implies there is an $f \in A_X$ such that (4.7) does not hold.

The case of Type III dependence is handled similarly. □

We note that while Type I dependence is symmetric in X and Y, Types II and III are not as is evident from our characterizations. In some sense, the size of the sets C_X, A_X, and B_X is a measure of the relative strengths of the dependence relations.

We can use the dual cones derived in section 3 to obtain inequalities for concave (convex) functions somewhat similar to those given in Corollary 4.2. To set some notation, we note that if the random variables X and $f(X)$ are square integrable, then the linear function of X which is closest to $f(X)$ in the sense of minimizing $E(f(X)-(aX+b))^2$ is given by $l_f(X) = a_f X + b_f$ where

(4.8) $$a_f = E(Xf(X)) - E(X)Ef(X)/\sigma_X^2,\ b_f = Ef(X) - a_f E(X).$$

It is well known that $Ef(X) = E\,l_f(X)$ and $EXf(X) = EX\,l_f(X)$. Interestingly, if f and g are both concave (convex) functions such that $f(X)$ and $g(X)$ are integrable, then replacing $f(X)$ and/or $g(X)$ by their linear approximations can only decrease the expected value of the product. We begin with a more general result for discrete random variables.

COROLLARY 4.4. *If the random variable X is finitely discrete (on the values $x_1 < x_2 < \ldots < x_n$), f is concave on the range of X and g is such that* (1) $Eg(X) = 0$, (2) $EXg(X) = 0$, (3) $E(x-X)g(X)I_{(X<x)} \geqslant 0$ *for all x in the support of X, then $Ef(X)g(X) \leqslant 0$.*

Proof. The proof follows directly from Theorem 3.1 by letting $\mathbf{w_i} = P(X = x_i)$. □

An important class of functions which satisfies the above conditions is given in the following theorem.

THEOREM 4.2. *If $g(x)$ is convex then $g(x) - (a_g x + b_g)$ (as defined in 4.8) satisfies conditions 1), 2) and 3) of Corollary 4.4.*

Proof. The proof is trivial if g is linear so assume that it is not. It is easily shown that conditions 1) and 2) hold so we consider condition 3). Now by the convexity assumption, $g(x) - (a_g x + b_g)$ must be positive, negative and positive again. Thus $\Sigma_{j=1}^{i} g(x_j) - (a_g x_j + b_g)$ must first be nonnegative and then nonpositive as i increases from 1 to n. Thus $g(x) - (a_g x + b_g)$ is in the cone K_I^{w*} (see 1.3) for the weights $w_i = P(X = x_i)$. Since for each $i, h(x_j) = \sup\{x_i - x_j, 0\}$ is in $-K_I$ (see (1.2)), condition 3) must hold by the definition of dual convex cones. □

This leads to the following corollary which also holds for the continuous case. Note that b) is similar to the Chebyshev inequality (4.4) with monotonicity replaced by concavity (convexity).

COROLLARY 4.5. *If f and g are both concave (convex) functions such that $X, f(X)$ and $g(X)$ are all square integrable, then* (a)

$$Ef(X)g(X) \geqslant Ef(X)(a_g X + b_g) = E(a_f X + b_f)(a_g X + b_g).$$

Moreover, if $EXf(X) - EXEf(X)$ and $EXg(X) - EXEg(X)$ have the same sign, then (b) $Ef(X)g(X) \geqslant Ef(X)\,Eg(X)$.

Proof. The first inequality follows by considering finer and finer partitions of the support of X, noting that f and g are concave on the partition points, and employing Theorem 4.2 and Corollary 4.4 together with limiting arguments. The equality in (a) follows from $a_g x + b_g$ being both concave and convex. Inequality (b) then follows from Chebyshev's inequality on the last part of a). □

Acknowledgement. The author would like to thank Tim Robertson for helpful discussions on the subject of convex cones.

REFERENCES

BARLOW, R. E., BARTHOLOMEW, D. J., BREMNER, J. M., and BRUNK, H. D. (1972). *Statistical Inference Under Order Restrictions*. Wiley, New York.

BARLOW, R. E. and BRUNK, H. D. (1972). The isotonic regression problem and its dual. *J. Amer. Statist. Assoc.* *67* 140–147.

BREIMAN, LEO (1968). *Probability*. Addison Wesley, Reading, Mass.

BRUNK, H. D. (1956). On an inequality for convex functions. *Proc. Amer. Math. Soc.* *7* 817–824.

DYKSTRA, RICHARD L. and ROBERTSON, TIM (1982). Order restricted statistical tests on multinomial and Poisson Parameters: the starshaped restriction. *Ann. Statist.* *10* 1246–1252.

HARDY, G. H., LITTLEWOOD, J. E., and PÓLYA, G. (1959). *Inequalities*, 2nd ed. Cambridge Univ. Press, London and New York.

HESTENES, MAGNUS R. (1975). *Optimization Theory*, Wiley, New York.

HORN, M. (1979). Some inequalities for the expectation of a product of functions of a random variable and for the multivariate distribution function at a random point. *Biometrical J.* *21* 243–245.

KIMBALL, A. W. (1951). On dependent tests of significance in the analysis of variance. *Ann. Math. Statist.* *22* 600–602.

LEHMANN, E. L. (1966). Some concepts of dependence. *Ann. Math. Statist.* *37* 1137–1153.

MARSHALL, ALBERT W. and OLKIN, INGRAM (1979). *Inequalities: Theory of Majorization and Its Applications*. Academic Press, New York.

MARSHALL, ALBERT W. and PROSCHAN, FRANK (1970). Mean life of series and parallel systems. *J. Appl. Probab.* *7* 165–174.

MARSHALL, A. W., WALKUP, D. W., and WETS, R. J. B. (1967). Order preserving functions: Applications to majorization and order statistics. *Pacific J. Math.* *23* 569–584.

ROBERTSON, TIM and WEGMAN, E. J. (1978). Likelihood ratio tests for order restrictions in exponential families. *Ann. Statist.* *6* 485–505.

ROBERTSON, TIM and WRIGHT, F. T. (1981). Likelihood ratio tests for and against a stochastic ordering between multinomial populations. *Ann. Statist.* *9* 1248–1257.

ROCKAFELLAR, A. TYRELL (1970). *Convex Analysis*. Princeton University Press, Princeton, New Jersey.

SHAKED, MOSHE (1979). Estimation of starshaped sequences of Poisson and normal means. *Ann. Statist.* *7* 729–741.

Inequalities in Statistics and Probability
IMS Lecture Notes–Monograph Series Vol. 5 (1984), 236–243

TESTS FOR AND AGAINST TRENDS AMONG POISSON INTENSITIES[1]

BY RHONDA MAGEL[2] and F. T. WRIGHT
North Dakota State University and University of Missouri-Rolla

Suppose one observes independent Poisson processes with unknown intensities λ_i, $i = 1, \ldots, k$, and that apriori it is believed that these intensities satisfy a known ordering. For preliminary analysis, it might be desirable to test for homogeneity among the intensities and, of course, one would want a test that utilizes the information in the ordering. Let t_i denote the length of time for which the ith process was observed. The case in which the t_i are equal has been studied in the literature. We develop the conditional likelihood ratio test for arbitrary t_i. This test is equivalent to the unconditional likelihood ratio test, but leads to an interesting multinomial testing situation, ie. testing for homogeneity of p_i/t_i versus a trend among the p_i/t_i, *where the* p_i are the cell probabilities. If the number of trials in the multinomial setting, or the total number of occurrences in the Poisson processes, is large, then the test statistic has an approximate chi-bar-squared distribution which has been studied in the literature. Results of a Monte Carlo study comparing this test with the maximin test developed by Lee (1980) are discussed. Similar results are also obtained for testing the null hypothesis that the intensities satisfy the prescribed ordering.

1. Introduction. Barlow, Bartholomew, Bremner and Brunk (1972) discuss the problem of estimating a finite sequence of Poisson intensities which are assumed to be nonincreasing. For instance, consider a system which is observed for t_1 units of time with X_1 failures, is then modified in an attempt to improve its performance, is observed for t_2 units of time with X_2 failures, is modified again, and this is repeated until it is observed for the kth time for t_k units of time with X_k failures. If it is believed that the modifications will not harm the system's performance, then one might wish to estimate the vector of intensities, $\lambda = (\lambda_1, \ldots, \lambda_k)$, subject to $\lambda_1 \geqslant \ldots \geqslant \lambda_k$. It would also be of interest to test for homogeneity among the intensities with the alternative $\lambda_1 \geqslant \ldots \geqslant \lambda_k$ and $\lambda_1 > \lambda_k$, or if the assumption concerning the modification were in question, one could test $\lambda_1 \geqslant \ldots \geqslant \lambda_k$ against $\lambda_i < \lambda_{i+1}$ for some i.

Suppose $X_1, \ldots, X_k$ are independent Poisson variables with means $\mu_i = \lambda_i t_i$, let $<<$ be a partial order on $\{1,2, \ldots, k\}$, let $\lambda^{(0)}$ be a fixed vector, let a be an unknown scale parameter and let H_0: $\lambda = a\lambda^{(0)}$, H_1: $\lambda_i \leqslant \lambda_j$ whenever $i << j$ and H_2:$\sim H_1$ (that is, $\lambda_i > \lambda_j$ for some $i << j$). The hypothesis H_1 stipulates that $\lambda = (\lambda_1, \ldots, \lambda_k)$ is isotonic (with respect to $<<$) and we suppose that $\lambda^{(0)}$ is isotonic. We consider the likelihood ratio test (lrt) for H_0 versus $H_1 - H_0$ and H_1 versus H_2 conditional on $\Sigma_{i=1}^{k} X_i = n$. While it will be shown that the conditional test is equivalent to the unconditional lrt, it does lead to an interesting multinomial testing situation. We also know that for $k = 2$ it is UMP unbiased. (See Ferguson (1967, p. 228)).

Robertson and Wegman (1978) consider order restricted tests for members of the exponential family, but their work requires that the sample sizes be equal. Their results can be applied in the testing situation considered here only if the t_i are all equal. Boswell (1966)

---

[1] This research was sponsored by the Office of Naval Research under ONR Contract N00014–80–C–0322.
[2] Parts of this work are taken from this author's doctoral dissertation written at the University of Missouri-Rolla.

AMS 1980 subject classification. Primary 62F03, secondary 62E20.
Key words and phrases: Order restricted inferences, trends in Poisson intensities, trends in multinomial parameters.

considers a closely related problem and develops the conditional lrt for testing that the intensity of a nonhomogeneous Poisson process is constant versus it is nondecreasing.

Lee (1980) developed a maximin test for the multinomial setting arising in the conditional framework. It could be used to provide a test of H_0 versus $H_1 - H_0$. The results of a Monte Carlo study comparing the lrt and the maximin test are given in Section 4. It is found that for small k ($k = 3,4,5$), the two perform similarly with the power of the maximin test larger for "regular" alternatives and the power of the lrt larger for "nonregular" alternatives. Furthermore, the differences in power are not too large. However, for larger k ($k = 10$) the differences are more pronounced and if nonregular alternatives cannot be ruled out the lrt should be used.

2. Estimation. The maximum likelihood estimate (mle) of λ subject to H_1 can be expressed as a projection onto a cone of isotonic functions. We introduce some notation. With $<<$ a fixed partial order on $\{1,2, \ldots, k\}$, let $\mathcal{R}_k$ denote the k-dimensional reals, let

$$C = \{x \in \mathcal{R}_k : x \text{ is isotonic with respect to } <<\},$$

let $\mathbf{w} = (w_1, w_2, \ldots, w_k)$ be a vector of positive weights, for $\mathbf{x},\mathbf{y} \in \mathcal{R}_k$ let $(x,y)_w$ denote the inner product $\Sigma_{i=1}^k w_i x_i y_i$, and for $\mathbf{y} \in \mathcal{R}_k$ let $E_\mathbf{w}(\mathbf{y}|C)$ denote the projection of $\mathbf{y}$ onto C, that is $E_\mathbf{w}(\mathbf{y}|C)$ minimizes

$$\Sigma_{i=1}^k w_i (y_i - x_i)^2 \quad \text{for } \mathbf{x} \in C.$$

Theorems 1.4 and 1.5 of Barlow, Bartholomew, Bremner and Brunk (1972) state that

$$\Sigma_{i=1}^k w_i E_w(\mathbf{y}|C)_i = \Sigma_{i=1}^k w_i y_i \tag{2.1}$$

and

$$\Sigma_{i=1}^k w_i (y_i - E_w(\mathbf{y}|C)_i) E_\mathbf{w}(\mathbf{y}|C)_i = 0 \tag{2.2}$$

The mle of λ subject to H_1 is $\bar{\lambda} = E_t(\mathbf{X}/\mathbf{t}|C)$ where $\mathbf{t} = (t_1, \ldots, t_k)$, $\mathbf{X} = (X_1, \ldots, X_k)$ and for $\mathbf{x},\mathbf{y} \in \mathcal{R}_k$, $\mathbf{x}/\mathbf{y} = (x_1/y_1, \ldots, x_k/y_k)$ and $\mathbf{xy} = (x_1 y_1, \ldots, x_k y_k)$. (See Barlow, Bartholomew, Bremner and Brunk (1972, p. 44)).

Conditioning on $Y = \Sigma_{i=1}^k X_i = n$, the density of $\mathbf{X}$ is that of a multinomial with parameters n and

$$p_i = \lambda_i t_i / \Sigma_{j=1}^k \lambda_j t_j \quad \text{for } i = 1, \ldots, k. \tag{2.3}$$

So H_0: $\lambda = a\lambda^{(0)}$ is equivalent to H'_0: $p_i = p_i^{(0)} = \lambda_i^{(0)} t_i / \Sigma_{j=1}^k \lambda_j^{(0)} t_j$, $1 \le i \le k$,

H_1: λ is isotonic is equivalent to H'_1: $\mathbf{p}/\mathbf{t}$ is isotonic, and

H_2: λ is not isotonic is equivalent to H'_2: $\mathbf{p}/\mathbf{t}$ is not isotonic.

The mle of $\mathbf{p}$ subject to H'_1 is also of interest.

THEOREM 1. *The mle of* $\mathbf{p}$ *subject to* H'_1 *is given by* $\bar{\mathbf{p}} = \mathbf{t}E_\mathbf{t}(\mathbf{X}/\mathbf{t}|C)/n$. *These* $\bar{p}_i$ *satisfy* $\bar{p}_i \geqslant 0$ *and* $\Sigma_{i=1}^k \bar{p}_i = 1$, *and* $\bar{\mathbf{p}} \to \mathbf{p}$ *almost surely provided* $\mathbf{p}/\mathbf{t}$ *is isotonic.*

Proof. The mle of $\mathbf{p}$ under H'_1 maximizes $\Pi_{i=1}^k (p_i/t_i)^{x_i}$ subject to $\mathbf{p}/\mathbf{t}$ is isotonic, $p_i \geqslant 0$ and $\Sigma_{i=1}^k p_i = 1$. Applying the result in Barlow, Bartholomew, Bremner and Brunk (1972, p. 46), we see that $\bar{\mathbf{p}} = \mathbf{t}E_\mathbf{t}(\mathbf{X}/\mathbf{t}|C)/n$. Since the projection $E_\mathbf{w}(\cdot|C)$ is continuous for fixed C and $\mathbf{w}$, $\bar{\mathbf{p}} = \mathbf{t}E_\mathbf{t}(\mathbf{X}/(n\mathbf{t}|C) \to \mathbf{t}E_\mathbf{t}(\mathbf{p}/\mathbf{t}|C)$ almost surely. If $\mathbf{p}/\mathbf{t}$ is isotonic the right hand side (rhs) is $\mathbf{p}$ and the proof is completed. □

If $<<$ is a total order, then the pool-adjacent-violators algorithm can be used to compute the projection in the formula for $\bar{\mathbf{p}}$ and the lower sets algorithm can be used for an arbitrary partial order. (See Chapter 2 of Barlow, Bartholomew, Bremner and Brunk (1972)).

3. Tests of Hypotheses. As was mentioned in the Introduction, we consider the conditional lrts of H_0 versus H_1–H_0 and of H_1 versus H_2. However, these lead to the lrts of H'_0 versus H'_1–H'_0 and of H'_1 versus H'_2. Chacko (1966) developed the lrt and an asymptotically equivalent modified χ^2 test for H'_0 versus H'_1–H'_0 in the totally ordered case with $\mathbf{p}^{(0)}$ and $\mathbf{t}$ constant vectors. Robertson (1978) developed both lrts for partial orders with $\mathbf{t}$ a constant vector.

Denoting the conditional likelihood ratios by λ'_{01} and λ'_{12}, we see that

$$T'_{01} = -2\, ln\, \lambda'_{01} = 2\Sigma_{i=1}^{k} X_i\{lnE_{\mathbf{t}}(\mathbf{X}/\mathbf{t}|C)_i - ln(p_i^{(0)}/t_i\} - 2n\, ln\, n \tag{3.1}$$

and

$$T'_{12} = -2ln\lambda'_{12} = 2\Sigma_{i=1}^{k} X_i\{ln(X_i/t_i) - lnE_{\mathbf{t}}(\mathbf{X}/\mathbf{t}|C)_i\}. \tag{3.2}$$

Hence, the test statistic for H_0 versus $H_1 - H_0$ is obtained by replacing $p_i^{(0)}$ with $\lambda_i^{(0)}t_i/\Sigma_{j=1}^{k}\lambda_j^{(0)}t_j$ in (3.1) and for testing H_1 versus H_2, $T_{12} = T'_{12}$:

Remark 2. The conditional lrts, T_{01} and T_{12}, are also the unconditional lrts.

Proof. Under H_0, the mle of a is $n/\Sigma_{j=1}^{k}\lambda_j^{(0)}t_j$. Using (2.1) and straight-forward algebra and denoting the likelihood ratio for testing H_0 versus $H_1 - H_0$ (H_1 versus H_2) by λ_{01} (λ_{12}), one can show that $-2\, ln\, \lambda_{01} = T_{01}$ and $-2\, ln\, \lambda_{12} = T_{12}$. □

Next the large sample distributions of these test statistics are determined. The derivations are like those given in Robertson (1978) and in fact we will make use of several lemmas proved there.

LEMMA 3. In the multinomial setting, let Z_i be independent normal variables with mean 0 and variance $1/p_i$ and let $\bar{Z} = \Sigma_{i=1}^{k} p_i Z_i$. As $n \to \infty$, $\sqrt{n}\,(\hat{p}-p) \xrightarrow{\mathcal{D}} (p_1(Z_1-\bar{Z}), \ldots, p_k Z_k - \bar{Z}))$. Let q denote a fixed vector of probabilities with $\mathbf{q}/\mathbf{t}$ isotonic, let D denote the closed, convex cone

$$D = \{x \in \mathcal{R}_k: x_i \leq x_j \text{ if } i << j \text{ and } q_i/t_i = q_j/t_j\},$$

let $r_1 < \ldots < r_h$ denote the distinct values among t_i, $i = 1, \ldots, k$, and let $M(i) = \{j: q_j/t_j = r_i\}$ for $i = 1, \ldots, h$. Robertson (1978) observed that $E_{\mathbf{w}}(\mathbf{x}|D)$ can be computed by independently computing its values for subscripts in $M(i)$ for $i = 1, \ldots, h$. This implies that if $\mathbf{y} \in \mathcal{R}_k$ has positive entires and is constant on each $M(i)$, then

$$E_{\mathbf{w}}(\mathbf{x}|D) = E_{\mathbf{wy}}(\mathbf{x}|D). \tag{3.3}$$

(This can be easily seen by considering the lower-sets algorithm.)

LEMMA 4. If $\mathbf{x},\mathbf{y} \in \mathcal{R}_k$ with $\mathbf{y}$ constant on each $M(i)$, then $E_{\mathbf{w}}(\mathbf{x}-\mathbf{y}|D) = E_{\mathbf{w}}(\mathbf{x}|D) - \mathbf{y}$. If in addition, $\mathbf{y}$ has nonnegative entries, then $E_{\mathbf{w}}(\mathbf{xy}|D) = \mathbf{y}E_{\mathbf{w}}(\mathbf{x}|D)$.

LEMMA 5. If $\mathbf{x} \in \mathcal{R}_k$ with $\max_{j \in M(i)} x_i < \min_{j \in M(i+1)} x_j$ for $i = 1, \ldots, h-1$, then $E_{\mathbf{w}}(\mathbf{x}|C) = E_{\mathbf{w}}(\mathbf{x}|D)$.

The cone D is determined by $<<$ and q. For a given D, define $P_w(\ell,k)$ to be the probability of exactly ℓ distinct values in $E_{\mathbf{w}}(V|D)$ where $\mathbf{V} = (V_1, \ldots, V_k)$, with the V_i independent normal variables with mean 0 and variance $1/w_i$. *If* $<<$ is the usual total order on $1, \ldots, k$ and q_i/t_i is constant, then $D = \{x: x_1 \leq \ldots \leq x_k\}$ and the $P_w(\ell,k)$ are discussed in detail in Barlow, Bartholomew, Bremner and Brunk (1972). Approximations for nonconstant $\mathbf{w}$ are discussed in Siskind (1976) and Robertson and Wright (1983). If $<<$ is the same total order, but q_i/t_i is not constant, then the results mentioned above can be used to determine these probabilities for each $M(i)$ and the h-fold convolution gives the desired $P_w(\ell,k)$. (Cf.

Barlow, Bartholomew, Bremner and Brunk (1972)). The latter reference also discusses the the $P_w(\ell,k)$ for some partial orders. Let χ^2_ℓ denote a chi-squared variable with ℓ degrees of freedom ($\chi^2_0 \equiv 0$).

THEOREM 6. *If in the Poisson setting, the vector of intensities is a* $\lambda^{(0)}$ *and* $\mathbf{q} = \mathbf{p}^{(0)}$ *(cf. (2.3)), or in the multinomial setting, if the probability vector is* $\mathbf{q} = \mathbf{p}^{(0)}$ *and if D is determined by* $\mathbf{q}$, *then for* $t \geq 0$,

(3.4) $$\lim_{n\to\infty} P[T_{01} \geq t | Y=n] = \lim_{n\to\infty} P[T'_{01} \geq t] = \Sigma^k_{\ell=1} P_q(\ell,k) P[\chi^2_{\ell-1} \geq t].$$

If in the Poisson setting, the vector of intensities is of the form $a\nu$ with ν isotonic and $\mathbf{q}$ is determined by (2.3) with $\lambda^{(0)}$ replaced by ν, or if in the multinomial setting the vector of probabilities, $\mathbf{q}$, is such that $\mathbf{q}/\mathbf{t}$ is isotonic, and if D is determined by $\mathbf{q}$, then for $t \geq 0$,

(3.5) $$\lim_{n\to\infty} P[T_{12} \geq t | Y=n] = \lim_{n\to\infty} P[T'_{12} \geq t] = \Sigma^k_{\ell=1} P_q(\ell,k) P[\chi^2_{k-\ell} \geq t],$$

(3.6) $$\lim_{n\to\infty} P[T_{12} \geq t | Y=n] \leq \lim_{n\to\infty} P_e[T_{12} \geq t | Y=n], \text{ and}$$

(3.7) $$\lim_{n\to\infty} P[T'_{12} \geq t] \leq \lim_{n\to\infty} P_\delta[T'_{12} \geq t],$$

where in the Poisson setting $P_e[\ ]$ denotes the probability under $\lambda = (1,1, \ldots, 1)$ and in the multinomial setting, $P_\delta[\ ]$ denotes the probability under $p_i = t_i/\Sigma^k_{j=1} t_j$, $i = 1,2, \ldots, k$.

Comments. In the Poisson setting, if one wishes to test H_0 versus $H_1 - H_0$, then $\mathbf{q}$ is set equal to $\mathbf{p}^{(0)}$ determined by (2.3). The vector $\mathbf{q}$ determines D which in turn determines the $P_\mathbf{q}(\ell,k)$ and so large sample p-values can be calculated for the conditional test from (3.4). In testing H_1 versus H_2, (3.6 indicates that the asymptotically least favorable configuration in H_1 is $\lambda = (1,1, \ldots, 1)$ and so with $q_i = t_i/\Sigma^k_{j=1} t_j$, $D = C$ and approximate p-values can be computed from (3.5).

In the multinomial setting, in testing H'_0 versus $H'_1 - H'_0$, set $q = p^{(0)}$ and use (3.4) to determine large sample p-values. In testing H'_1 versus H'_2, the asymptotically least favorable configuration in H_1 is $q_i = t_i/\Sigma^k_{j=1} t_j$ in which case $D = C$ and approximate p-values can be computed from (3.5).

In either case, one might not want to use the asymptotically least favorable configuration when testing H_1 versus H_2 or H'_1 versus H'_2, so one could use the restricted mles of λ or $\mathbf{p}$ rather than $\mathbf{e}$ or δ.

Proof. T_{01} is defined to be T'_{01} with $p_i^{(0)}$ replaced by $\lambda_i^{(0)} t_i/\Sigma^k_{j=1}\lambda_j^{(0)} t_j$ and $T_{12} = T'_{12}$. Furthermore, conditional on $Y = n$, $(X_1, \ldots, X_k)$ is multinomial with parameters n and $p_i = \lambda_i t_i/\Sigma^k_{j=1}\lambda_j t_i$ and $H_i \equiv H'_i$, $i = 0,1,2$. So we need only consider the multinomial situation. We first consider the distribution of T'_{01} under H'_0. Setting $\hat{p}_i = X_i/n$, expressing T'_{01} as

$$2n\Sigma^k_{i=1}\hat{p}_i\{ln\, E_\mathbf{t}(\hat{\mathbf{p}}/\mathbf{t}|C)_i - ln(p_i^{(0)}/t_i)\},$$

and expanding $ln\, E_\mathbf{t}(\hat{\mathbf{p}}/\mathbf{t}|C)_i$ and $ln\,(p_i^{(0)}/t_i)$ about $\hat{p}_i/t_i$, we write T'_{01} as

(3.8) $$2n\Sigma^k_{i=1} t_i\{E_\mathbf{t}(\hat{\mathbf{p}}/\mathbf{t}|C)_i - p_i^{(0)}/t_i\}$$
$$+\ n\Sigma^k_{i=1}\hat{p}_i\{((\hat{p}_i - p_i^{(0)})/t_i)^2/\beta_i^2 - \mathrm{E}_\mathbf{t}(\hat{\mathbf{p}}/\mathbf{t}|C)_i - \hat{p}_i/t_i)^2/\alpha_i^2\},$$

where $\alpha_i(\beta_i)$ is between p_i/t_i and $E_\mathbf{t}(\mathbf{p}/\mathbf{t}|C)_i$ $(p_i^{(0)}/t_i)$. Under H'_0, both $E_\mathbf{t}(\mathbf{p}/\mathbf{t}|C)_i$ and p_i/t_i converge almost surely to $p_i^{(0)}/t_i$. Recall, (2.1) implies that the first term in (3.8) vanishes.

Since $\hat{\mathbf{p}}/\mathbf{t}$ is consistent for $\mathbf{p}^{(0)}/\mathbf{t}$ under H'_0, there is for almost all ω in the underlying probability space an N, possibly depending on ω, for which $\hat{\mathbf{p}}/\mathbf{t}$ satisfies the hypothesis of Lemma 5 for $n \geq N$. Hence, for such ω and n, Lemmas 4 and 5 can be applied to the second term in (3.8) to obtain

$$\Sigma_{i=1}^{k}\hat{p}_i\{[\sqrt{n}\ (\hat{p}_i-p_i^{(0)})/t_i\beta_i]^2-[(t_iE_t(\sqrt{n}\ (\hat{\mathbf{p}}-\mathbf{p}^{(0)})/\mathbf{t}|D)_i-\sqrt{n}\ (\hat{p}_i-p_i^{(0)})/t_i\alpha_i]^2\}$$

which converges in distribution to

$$\Sigma_{i=1}^{k}(t_i^2/p_i^{(0)})\{((p_i^{(0)}/t_i)(Z_i-\bar{Z}))^2-(E_t((p^{(0)}/\mathbf{t})(\mathbf{Z}-\bar{\mathbf{Z}})|D)_i-(p_i^{(0)}/t_i)(Z_i-\bar{Z}))^2\}.$$

Applying Lemma 4 and (3.3), this can be written as

$$\begin{aligned}(3.9)\qquad & \Sigma_{i=1}^{k}p_i^{(0)}(\mathbf{Z}_i-\bar{\mathbf{Z}})^2-\Sigma_{i=1}^{k}p_i^{(0)}(\mathbf{Z}_i-E_{p(0)}(\mathbf{Z}|D)_i)^2\\ &=\Sigma_{i=1}^{k}p_i^{(0)}(E_{p(0)}(\mathbf{Z}|D)_i-\bar{\mathbf{Z}})^2+2\Sigma_{i=1}^{k}p_i^{(0)}(\mathbf{Z}_i-E_{p(0)}(\mathbf{Z}|D)_i)(E_{p(0)}(\mathbf{Z}|D)_i-\bar{\mathbf{Z}}).\end{aligned}$$

The second term in the rhs of (3.9) can be shown to be zero using (2.1) and (2.2), and Theorem 3.1 of Barlow, Bartholomew, Bremner and Brunk (1972) shows that the first term on the rhs has the desired distribution.

Next we consider the distribution of T'_{12} with q a fixed probability vector for which $\mathbf{q}/\mathbf{t}$ is isotonic. Writing T'_{12} as $2n\Sigma_{i=1}^{k}\hat{p}_i\ ln(\hat{p}_i/t_i)-ln\ E_t(\hat{\mathbf{p}}/\mathbf{t}|C)_i$, expanding $lnE_t(\mathbf{p}/\mathbf{t}|C)_i$ about $\hat{p}_i/t_i$ and applying (2.1), T'_{12} can be written as

$$\Sigma_{i=1}^{k}\hat{p}_i(\sqrt{n}\ (E_t(\hat{\mathbf{p}}/\mathbf{t}|C)_i-\hat{p}_i/t_i))^2/\gamma_i^2$$

where γ_i is between $E_t(\hat{\mathbf{p}}/\mathbf{t}|C)_i$ and $\hat{p}_i/t_i$ and hence converges almost surely to q_i/t_i. As in the proof of the first part of the theorem, for almost all ω and n sufficiently large $\sqrt{n}\ (E_t(\hat{\mathbf{p}}/\mathbf{t}|C)-\mathbf{q}/\mathbf{t})=E_t(\sqrt{n}\ (\hat{\mathbf{p}}-\mathbf{q})/\mathbf{t}|D)$.

Hence, T'_{12} converges in distribution to

$$(3.10)\qquad \Sigma_{i=1}^{k}(t_i^2/q_i)(E_t((\mathbf{q}/\mathbf{t})(\mathbf{Z}-\bar{\mathbf{Z}})|D)_i-(q_i/t_i)(Z_i-\bar{Z}))^2=\Sigma_{i=1}^{k}q_i(E_t(\mathbf{Z}|D)_i-\bar{Z}_i)^2.$$

Applying (3.3) this becomes

$$\Sigma_{i=1}^{k}q_i(E_q(Z|D)_i-Z_i)^2$$

and this has the desired distribution (cf. Theorem 2.5 of Robertson and Wegman (1978)).

We establish (3.7) to conclude the proof. The variables $U_i=\sqrt{q_i/t_i}\ Z_i$ are independent normal variables with means zero and $\mathrm{var}(U_i)=1/t_i$. Using (3.3), the rhs of (3.10) can be written as $\Sigma_{i=1}^{k}t_i(E_t(\mathbf{U}|D)-U_i)^2$, which is the distance from $E_t(\mathbf{U}|D)$ to $\mathbf{U}$. Since $C\subset D$, by the definition of projection this is maximized for $D=C$, which occurs if $\mathbf{q}/\mathbf{t}$ is constant, ie. $q_i=t_i/\Sigma_{i=1}^{k}t_j$. □

In the multinomial setting, Lee (1980) developed a maximin test for $p_1/t_1=p_2/t_2=\ldots=p_k/t_k$ versus $p_{i+1}/t_{i+1}\geq dp_i/t_i$, $i=1,2,\ldots,k-1$, with $d>1$. The test statistic is $S_{01}=\Sigma_{i=1}^{k}iX_i$, which has an approximate normal distribution and under this null hypothesis its mean and variance are

$$n\Sigma_{i=1}^{k}it_i/\Sigma_{i=1}^{k}t_i \quad \text{and}\ n\{\Sigma_{i=1}^{k}i^2t_i/\Sigma_{i=1}^{k}t_i-(\Sigma_{i=1}^{k}it_i/\Sigma_{i=1}^{k}t_i)^2\},$$

respectively. The tests S_{01} and T'_{01} are compared in the next section.

4. Comparison of the lrt and Maximin Tests. The maximin test is a contrast test and Section 4.2 of Barlow, Bartholomew, Bremner and Brunk (1972) contains a discussion of the use of the likelihood ratio and contrast approaches in testing for trends among normal means. They concluded that, while the contrast test is typically much easier to use, the lrt provides the most satisfactory general way of incorporating prior information about ordering. In the case of a total order and small k, the contrast statistic provides a suitable alternative. If additional information is available about the spacing of the parameters, then a contrast test based on this additional information may be preferred.

To give some idea of the differences in the power for the two tests a Monte Carlo study was conducted. In particular, if one were testing

$$p_i = t_i/\Sigma_{j=1}^{k} t_j,\ 1 \leqslant i \leqslant k, \text{ versus } p_1/t_1 \leqslant p_2/t_2 \leqslant \ldots \leqslant p_k/t_k \text{ with } p_1/t_1 < p_k/t_k,$$

then T'_{01}, with $p_i^{(0)}$ replaced by $t_i/\Sigma_{j=1}^{k} t_j$, and S_{01} could be used. Because the distribution of T'_{01} under H'_1 is quite complex, Monte Carlo experiments were conducted. With $k = 3, 4, 5, 10$, $n = 25, 80$, nominal levels of .1, .05, and various choices of **t** and **p**, the powers of T'_{01} and S_{01} were approximated based on 5000 repetitions. Some of these values are given in Tables 1 and 2. To assess the accuracy of the approximations for the distributions under the null hypothesis, the estimates of the power under the null hypothesis are included. For $n = 80$ and the cases presented in these tables the largest discrepancy in the α level for $T'_{01}(S_{01})$ and a nominal level of .1 is .016 (.009) and for $\alpha = .05$ it is .005 (.007). For $n = 25$ the maximum discrepancies were larger but both approximations seem to be useful for k in this range. However, for $k = 10$ and $n = 25$ the approximation for the distribution of T'_{01} seems to give a test that is quite conservative. Its estimated α level is .065 (.031) when the nominal level was .1 (.05). The approximation for S_{01} seemed quite adequate even with $n = 25$ and $k = 10$. Its estimated α levels are .100 (.046), respectively.

It is clear from Tables 1 and 2 that neither test is uniformly better than the other. In fact, when the p_i/t_i increase regularly, such as in the cases **p/t** = (.25, .30, .45), **p/t** = (.20, .25, .35), **p/t** = (.15, .20, .30, .35), **p/t** = (.10, .15, .20, .25, .30), etc., then the maximin test outperforms T'_{01}, but for irregular increases in **p/t**, such as in the cases **p/t** = (.25, .25, .50), **p/t** = (.2, .3, .3), **p/t** = (.052, .052, .120), **p/t** = (.04, .04, .04, .07$\bar{3}$), **p/t** =(.03, .03, .07, .07), **p/t** = (.15, .2, .2, .25), etc., T'_{01} has greater power than S_{01}. For $k = 3,4,5$ the differences in power are not too large and the magnitudes are similar in both directions. So for small k one could use S_{01} if the alternative were believed to be "regular" in the sense described above or T'_{01} could be used if it is desirable to protect againt nonregular alternatives.

It is interesting to note that for $k = 3$ and 4 the above conclusions held whether the vector t was constant or not. (Several other choices of p and t, not given in Tables 1 and 2, were considered and these conclusions were substantiated in those cases, also.) For this reason only constant t's were considered for $k = 5$ and 10. Recall that for $k \geq 5$, the $P_t(\ell,k)$ are intractable for nonconstant **t**.

Power comparisons were made for $k = 10$, but with $n = 25$ they were not very meaningful because of the conservative nature of the approximation to the null distribution of T'_{01}. For $n = 80$ both approximations were very reasonable and so power comparisons could be made in that case. Linear alternatives were considered. For $p_i = i/55$ both tests have powers that are essentially one. So $p_i = i/110 + .05$ was considered. The tests with nominal level .1 had powers .822 and .861, and the tests with nominal level .05 had powers .712 and .759. Of course, the maximin test performed better for such an alternative. The alternative $p_i = .09$, $1 \leqslant i \leqslant 9$, and $p_{10} = .19$ was considered. The estimated powers for the tests with $\alpha = .1$ are .700 and .556 and for the tests with $\alpha = .05$ they are .578 and .409. Finally the alternative $p_1 = .07$, $p_2 = \ldots = p_9 = .1$, and $p_{10} = .13$ was considered, and the approximate powers for the $\alpha = .1$ tests are .377 and .332. For the $\alpha = .05$ tests they are .252 and .199. Of course, the last two alternatives are nonregular and the lrt has the larger power. In these cases, the increase in power may be as large as 40 percent and so the lrt should definitely be considered to guard against nonregular alternatives for larger k.

TABLE 1. Estimated Powers of the Maximin and LRTs, $k=3$

	$k=3$							
	$n=25$		$n=80$		$n=25$		$n=80$	
	Nominal Level							
	.10	.05	.10	.05	.10	.05	.10	.05
	t=(1,1,1), **p/t**=(1/3,1/3,1/3)				**t**=(2,1,1), **p/t**=(.25,.25,.25)			
T'_{01}	.084	.059	.108	.047	.101	.047	.085	.052
S_{01}	.088	.056	.095	.043	.081	.049	.100	.044
	t=(1,1,1), **p/t**=(.25,.30,.45)				**t**=(2,1,1), **p/t**=(.20,.25,.35)			
T'_{01}	.389	.227	.718	.565	.459	.315	.771	.664
S_{01}	.451	.350	.812	.689	.423	.337	.802	.677
	t=(1,1,1), **p/t**=(.25,.25,.50)				**t**=(2,1,1), **p/t**=(.2,.3,.3)			
T'_{01}	.559	.445	.937	.871	.346	.220	.624	.509
S_{01}	.572	.475	.917	.843	.304	.230	.633	.473
	t=(2,3,5), **p/t**=(.10,.10,.10)				**t**=(2,3,5), **p/t**=(.05,.10,.12)			
T'_{01}	.126	.046	.093	.054	.571	.332	.891	.803
S_{01}	.095	.056	.106	.048	.504	.386	.889	.778
	t=(8,2,8), **p/t**=(1/18,1/18,1/18)				**t**=(8,2,8), **p/t**=(.04,.06,.07)			
T'_{01}	.112	.051	.084	.052	.527	.351	.832	.754
S_{01}	.080	.053	.100	.051	.452	.368	.855	.763
	t=(4,6,4), **p/t**=(1/14,1/14,1/14)				**t**=(4,6,4), **p/t**=(.052,.052,.120)			
T'_{01}	.112	.067	.101	.048	.758	.609	.981	.947
S_{01}	.117	.040	.100	.046	.723	.535	.965	.921

TABLE 2. Estimated Powers of the Maximin and LRTs, $k=4,5$

	$k=4$							
	$n=25$		$n=80$		$n=25$		$n=80$	
	Nominal Level							
	.10	.05	.10	.05	.10	.05	.10	.05
	t=(1,1,1,1),**p/t**=(.25,.25,,25,.25)				**t**=(4,6,4,6), **p/t**=(.05,.05,.05,.05)			
T'_{01}	.089	.055	.097	.051	.099	.054	.095	.055
S_{01}	.105	.050	.101	.050	.086	.042	.103	.049
	t=(1,1,1,1), **p/t**=(.15,.20,.30,.35)				**t**=(4,6,4,6), **p/t**=(.034,.042,.0495,.069)			
T'_{01}	.578	.446	.930	.860	.487	.338	.826	.729
S_{01}	.634	.482	.945	.883	.480	.342	.860	.757
	t=(1,1,1,1), **p/t**=(.10,.25,.25,.40)				**t**=(4,6,4,6), **p/t**=(.04,.04,.04,.07/3)			
T'_{01}	.752	.665	.995	.985	.501	.345	.844	.766
S_{01}	.804	.668	.993	.981	.464	.333	.834	.727
	t=(1,1,1,1), **p/t**=(.23,.23,.23,.31)				**t**=(4,6,4,6), **p/t**=(.03,.03,.07,.07)			
T'_{01}	.208	.129	.376	.256	.730	.562	.983	.955
S_{01}	.240	.143	.388	.245	.684	.544	.981	.946
	$k=5$ and **t**=(1,1,1,1)							
	p/t=(.2,.2,.2,.2,.2)				**p/t**=(.10,.15,.20,.25,.30)			
T'_{01}	.094	.050	.099	.049	.649	.489	.966	.930
S_{01}	.086	.051	.091	.050	.679	.567	.974	.946
	p/t=(.15,..15,.15,.15,.40)				**p/t**=(.15,.2,.2,.2,.25)			
T'_{01}	.711	.613	.985	.969	.264	.172	.496	.356
S_{01}	.654	.555	.955	.924	.252	.176	.488	.361

REFERENCES

BARLOW, R. E. BARTHOLOMEW, D. J., BREMNER, J. M. and BRUNK, H. D. (1972). *Statistical Inferences Under Order Restrictions*. Wiley, New York.

BOSWELL, M. T. (1966). Estimating and testing trend in a stochastic process of Poisson type. *Ann. Math. Statist. 37* 1564–1573.

CHACKO, V. J. (1966). Modified chi-square tests for ordered alternatives. *Sankyā Ser. B 28,* 185–190.

FERGUSON, T. S. (1967). *Mathematical Statistics: A Decision Theoretic Approach*. Academic Press, New York.

LEE, Y. J. (1980). Tests of trend in count data: multinomial distribution case. *J. Amer. Statist. Assoc. 72* 673–675.

SISKIND, V. (1976). Approximate probability integrals and critical values for Bartholomew's test for ordered means. *Biometrika 63* 647–654.

ROBERTSON, Tim (1978). Testing for and against an order restriction on multinomial parameters. *J. Amer. Statist. Assoc. 73* 197–202.

ROBERTSON, Tim and WEGMAN, E. J. (1978). Likelihood ratio rests for order restrictions in exponential families. *Ann. Math. Statist. 6* 485–505.

ROBERTSON, Tim and WRIGHT, F. T. (1983). On approximation of the level probabilities and associated distributions in order restricted inference. *Biometrika 70* 597–606.

Inequalities in Statistics and Probability
IMS Lecture Notes–Monograph Series Vol. 5 (1984), 244–250

A MEASURE OF THE CONFORMITY OF A PARAMETER SET TO A TREND: THE PARTIALLY ORDERED CASE[1]

BY TIM ROBERTSON and F. T. WRIGHT
University of Iowa and University of Missouri-Rolla

Inferences concerning order restrictions on a collection of parameters, $\theta_1, \theta_2, \ldots, \theta_k$, are considered with the order restrinctions of the form, $\theta_i \leq \theta_j$ for $i \preceq j$ where $\preceq$ is a partial order on 1, 2, ... , k . Clearly, some parameter sets conform more closely to these order restrinctions than others. We are interested in measures of the degree of conformity. Some of the measures available in the literature for the totally ordered case are generalized to the partially ordered case and the theory developed is applied in several tests of order restricted hypotheses.

1. Introduction. In various situations, one is interested in a collection of parameters $\theta_1, \theta_2, \ldots, \theta_k$ which are believed to satisfy certain known order restrictions and inference procedures which make use of this ordering information are preferred. We consider order restrictions that are induced by partial orders on $\Omega = \{1, 2, \ldots, k\}$. That is, suppose that $\preceq$ is a partial order on Ω and that the order restrictions are $\theta_i \leq \theta_j$ when $i \preceq j$. Such a vector $\theta = (\theta_1, \theta_2, \ldots, \theta_k)$ is said to be isotone (with respect to $\preceq$). In studying such inference procedures it is helpful to have a measure of the degree of conformity to the order restrictions. For instance, a test of H_0: θ is constant versus H_1: θ is isotone, but not constant should have power that increases with the degree of conformity. For a non-simple null hypothesis such a concept could be useful in identifying a least favorable configuration. In a Bayesian approach, priors which assign larger probabilities to parameters conforming more closely to the order restrictions would be sought.

Barlow, Bartholomew, Bremner and Brunk (1972) contains a thorough discussion of order restricted inference. Robertson and Wright (1982) develop several measures of conformity for the totally ordered case, ie. $\theta_1 \leq \theta_2 \leq \ldots \leq \theta_k (1 \preceq 2 \preceq \ldots \preceq k)$. In considering unimodal structures, partial orders of the type $1 \preceq 2 \preceq \ldots \preceq r \succeq r + 1 \succeq \ldots \succeq k$ arise and when making one-sided comparisons of several treatments with a common control, the partial order $1 \preceq i$ for $i = 2, 3, \ldots, k$ occurs. (See Bartholomew (1959) and Robertson and Wright (1981).) Suppose that a dependent variable has mean $\theta(i,j)$ when the first independent variable is fixed at level i, $1 \leq i \leq r$, and the second independent variable is fixed at level j, $1 \leq j \leq c$. If the levels are increasing and if $\theta(\cdot,\cdot)$ increases with each independent variable as the other is held fixed, then the order restrictions are $\theta(i,j) \leq \theta(s,t)$ for $i \leq s$ and $j \leq t$. This is another example of a partial order that is not total. We extend the measures of conformity in Robertson and Wright (1982) to the partially ordered case.

A set $L \subset \Omega$ is a lower layer provided $i \in L$ whenever $i \preceq j$ and $j \in L$. We denote the collection of lower layers by $\mathcal{L}$. To allow for different weights on the parameters, let $\mathbf{w}$ be a positive weight function defined on Ω, ie. $\mathbf{w} = (w_1, w_2, \ldots, w_k)$. For situations in which the degree of conformity should be translation invariant, we consider the relationship

[1] This research was supported by the Office of Naval Research under the Contracts ONR N00014–80–C–0321 and ONR N00014–80–C–0322.

AMS 1980 subject classifications. Primary, 62F03; secondary, 62E15.

Key words and phrases: Order restricted inference, tests for and against a trend, isotonic inference, chi-bar-squared distributions, E-bar-squared distributions.

$>>$, defined on Euclidean space $\mathcal{R}^k$, by $\mathbf{x} = (x_1, x_2, \dots, x_k) >> \mathbf{y} = (y_1, y_2, \dots, y_k)$ if and only if

$$\Sigma_{i \in L} w_i(x_i - m(\mathbf{x})) \leq \Sigma_{i \in L} w_i(y_i - m(\mathbf{y})) \quad \text{for each } L \in \mathcal{L},$$

with $m(\mathbf{x}) = \Sigma_{i=1}^k w_i x_i / \Sigma_{i=1}^k w_i$. Robertson and Wright (1982) argue that $>>$ is appropriate for normal means, but for Poisson means a more appropriate measure is the following: $\mathbf{x} >> {}^*\mathbf{y}$ if and only if

$$\Sigma_{i \in L} w_i x_i \leq \Sigma_{i \in L} w_i y_i \text{ for each } L \in \mathcal{L} \text{ and } \Sigma_{i=1}^k w_i x_i = \Sigma_{i=1}^k w_i y_i.$$

Remark 1.1. The relationship $>>$ *and* $>>^*$ *are transitive and symmetric,* $>>^*$ *is reflexive, and* $\mathbf{x} << \mathbf{y}$ *and* $\mathbf{x} >> \mathbf{y}$ *imply that* $\mathbf{x}-\mathbf{y}$ *is a constant vector.*

Proof. The first conclusion is obvious and because $\mathbf{x} >> \mathbf{y}$ is equivalent to $\mathbf{x}-m(\mathbf{x}) >>^* \mathbf{y}-m(\mathbf{y})$, it suffices to show that $>>^*$ is reflexive.

Suppose $\mathbf{x} << {}^*\mathbf{y}$ and $\mathbf{x} >> {}^*\mathbf{y}$. Let $L_0 = \phi$ and inductively define L_α to consist of those $j \in \Omega$ for which $i \preceq j$ and $i \neq j$ *imply that* $i \in L_{\alpha-1}$. Observe that $L_{\alpha-1} \subset L_\alpha$, $L_\alpha - L_{\alpha-1} \neq \phi$, and because Ω is finite, there is an integer h for which $\phi = L_0 \subset L_1 \subset \dots \subset L_h = \Omega$. For each $j \in L_1$, $\{j\} \in \mathcal{L}$ and so $x_j = y_j$. Next, for $j \in L_2$, $L(j) = \{i \in \Omega : i \preceq j\} \in \mathcal{L}$, $L(j) - L_1 = \{j\}$ and so $x_j = y_j$. Continuing we see that $\mathbf{x}=\mathbf{y}$ and the proof is completed. □

If one identifies vectors $\mathbf{x}$ and $\mathbf{y}$ which differ by a constant vector, then $>>$ induces a partial order on the equivalence classes which is essentially $>>^*$.

Let $C = \{\mathbf{x} \in \mathcal{R}^k : \mathbf{x} \text{ is isotone with respect to } \preceq\}$ and note that the apriori belief concerning θ is that $\theta \in C$. Typically, estimates of θ are obtained by projecting initial estimates onto C, and test statistics are related to the distance from the initial estimates to the projections. The above measures of conformity can be characterized in terms of the Fenchel dual of C, which is defined by

$$C^{*\mathbf{w}} = \{\mathbf{y} \in \mathcal{R}^k : \Sigma_{i=1}^k w_i x_i y_i \leq 0 \text{ for all } \mathbf{x} \in C\}.$$

(If $\mathbf{w}$ is constant we denote the dual cone by C^*.) Barlow and Brunk (1972) and Dykstra (1981) discuss some of the implications of duality theory in order restricted inference. The following result is proved in the former reference (cf. Section 4).

Remark 1.2. With $\mathbf{x},\mathbf{y} \in \mathcal{R}^k$, *the following are equivalent:*

(A) $\mathbf{x} >> \mathbf{y}$ $(\mathbf{x} >>^* \mathbf{y})$;

(B) $\mathbf{y}-m(\mathbf{y})-\mathbf{x}+m(\mathbf{x}) \in C^{*\mathbf{w}}$ $(\mathbf{y}-\mathbf{x} \in C^{*\mathbf{w}})$; *and*

(C) $\Sigma_{i=1}^k w_i(y_i - m(\mathbf{y}) - x_i + m(\mathbf{x}))z_i \leq 0$ $(\Sigma_{i=1}^k w_i(y_i - x_i)z_i \leq 0)$
for each $\mathbf{z} \in C$.

Real valued functions which are nondecreasing with respect to these orderings are of interest. If $f: \mathcal{R}^k \to \mathcal{R}$ and $f(\mathbf{x}) \geq f(\mathbf{y})$ for all $\mathbf{x},\mathbf{y} \in \mathcal{R}^k$ with $\mathbf{x} >> \mathbf{y}$ $(\mathbf{x} >>^* \mathbf{y})$, then f is said to be ISO(ISO*). The next result is immediate.

Remark 1.3. A function $f: \mathcal{R}^k \to R$ *is ISO if and only if it is ISO* and* $f(\mathbf{x} + c\mathbf{e}_k) = f(\mathbf{x})$ *for all* $\mathbf{x} \in \mathcal{R}^k$ *and* $c \in R$, *where* $\mathbf{e}_k$ *is a k-dimensional vector of ones.*

Remark 1.4. Let $\mathbf{x},\mathbf{y} \in \mathcal{R}^k$. $\mathbf{x} >> \mathbf{y}$ $(\mathbf{x} >>^* \mathbf{y})$ *if and only if* $f(\mathbf{x}) \geq f(\mathbf{y})$ *for all* f *which are ISO(ISO*).*

Proof. The result is an easy consequence of the definitions of ISO and ISO* and the following facts: $f_L(\mathbf{x}) = -\Sigma_{i \in L} w_i(x_i - m(\mathbf{x}))$ is ISO for each $L \in \mathcal{L}$, $g_L(\mathbf{x}) = -\Sigma_{i \in L} w_i x_i$ is ISO* for each $L \in \mathcal{L}$ and $\Sigma_{i=1}^k w_i x_i$ is ISO*. □

The partial ordering $>>^*$ is a cone ordering as discussed in Marshall, Walkup and Wets

(1967) and the following result is contained in their work. However, its proof is so simple it is included here.

THEOREM 1.5. *Let $f: \mathcal{R}^k \to R$ be differentiable and let $f_i(\mathbf{x}) = (\partial/\partial x_i)f(\mathbf{x})$ for $i = 1,2, \ldots, k$. If $f_i(\mathbf{x})/w_i \leq f_j(\mathbf{x})/w_j$ for all i and j with $i \preceq j$ and all $\mathbf{x} \in \mathcal{R}^k$, then f is ISO*.*

Proof. Suppose $\mathbf{x} >>^* \mathbf{y}$. Using the mean value theorem there is a point $\mathbf{z}$ on the line segment joining $\mathbf{x}$ and $\mathbf{y}$ for which

$$f(\mathbf{y}) - f(\mathbf{x}) = \Sigma_{i=1}^k (y_i - x_i) f_i(\mathbf{z}) = \Sigma_{i=1}^k w_i (y_i - x_i)(f_{ij}(\mathbf{z})/w_i)$$

and the latter sum is non-positive since $\mathbf{y} - \mathbf{x} \in C^{*\mathbf{w}}$ and $(f_1(\mathbf{z})/w_1, \ldots, f_k(\mathbf{z})/w_k) \in C$ by hypothesis. □

2. Preservation Theorems. In this section, we establish results which say that if $\mathbf{X}$ is a set of observations, $f(\mathbf{X})$ is a statistic with f ISO(ISO*) and $h(\theta) = E_\theta f(\mathbf{X})$, then h is ISO(ISO*). The first result deals with a multinomial setting. Let $\mathbf{w} = \mathbf{e}_k$, let $A_n = \{\mathbf{x} \in \mathcal{R}^k$: each x_i is a nonnegative integer and $\Sigma_{i=1}^k x_i = n\}$, let $B = \{\mathbf{p} \in \mathcal{R}^k$: each $p_i \geq 0$ and $\Sigma_{i=1}^k p_i = 1\}$ and let $\mathbf{X} = (X_1, X_2, \ldots, X_k)$ be a multinomial vector with parameters n and $\mathbf{p} = (p_1, p_2, \ldots, p_k)$.

THEOREM 2.1. *If $f: A_n \to R$ is ISO, then $h(\mathbf{p}) = Ef(\mathbf{X})$ is ISO on B.*

Proof. As in Robertson and Wright (1982), $h_i(\mathbf{p}) - h_j(\mathbf{p}) =$

$$\Sigma_{\mathbf{y} \in A_{n-1}} (f(\mathbf{y}+\delta_i) - f(\mathbf{y}+\delta_j)) n! \Pi_{i=1}^k (p_i^{y_i}/y_i!),$$

where δ_r is a k-dimensional vector with sth coordinate zero unless $s = r$ and the rth coordinate is one. Suppose $i \preceq j$ and let $L \in \mathcal{L}$. If $i \notin L$ then $\Sigma_{r\in L}(\mathbf{y}+\delta_i)_r = \Sigma_{r\in L}(\mathbf{y}+\delta_j)_r$; if $i \in L$ and $j \notin L$, *then* $\Sigma_{r\in L}(\mathbf{y}+\delta_i)_r \geq \Sigma_{r\in L}(\mathbf{y}+\delta_j)_r$; and if $i,j \in L$, *then* $\Sigma_{r\in L}(\mathbf{y}+\delta_i)_r = \Sigma_{r\in L}(\mathbf{y}+\delta_j)_r$. The proof is completed by applying Theorem 1.5. □

Chacko (1966) and Robertson (1978) considered testing H_0: $\mathbf{p} = k^{-1}\mathbf{e}_k$ with the alternative restricted by the trend, H_1: $\mathbf{p}$ is isotone with respect to $\preceq$. Chacko considered the totally ordered case and Robertson the partially ordered case. The likelihood ratio test statistic is $T_{01} = -2 \ln \lambda = 2\Sigma_{i=1}^k X_i \ln (P(\mathbf{X}|C)_i) - 2n \ln n + 2n \ln k$ where $P(\mathbf{X}|C)$ is the projection of $\mathbf{X}$ onto C, which is characterized by

$$\Sigma_{i=1}^k (X_i - P(\mathbf{X}|C)_i) P(\mathbf{X}|C)_i = 0 \text{ and } \Sigma_{i=1}^k (X_i - P(\mathbf{X}|C)_i) z_i \leq 0$$

for all $\mathbf{z} \in C$. (See Barlow, Bartholomew, Bremner, and Brunk (1972, p. 28). Computation algorithms for $P(\mathbf{X}|C)$ are also discussed in their Chapter 2.) We first show that $f(\mathbf{x}) = \Sigma_{i=1}^k x_i \ln(P(\mathbf{x}|C)_i)$ is ISO on A_n, then note that this implies that $I_{[T_{01} \geq t]}$ is, for fixed t, ISO on A_n and applying Theorem 2.1, we see that the power function of T_{01}, $EI_{[T_{01} \geq t]}$, is ISO on B.

Suppose $\mathbf{x} >> \mathbf{y}$ with $x,y \in A_n$, then $\mathbf{y} - \mathbf{x} \in C^*$ (we omit the superscript w since it is constant) and so

$$\Sigma_{i=1}^k y_i \ln(P(\mathbf{y}|C)_i) = \Sigma_{i=1}^k x_i \ln(P(\mathbf{y}|C)_i) + \Sigma_{i=1}^k (y_i - x_i) \ln(P(\mathbf{y}|C)_i).$$

The second term on the r.h.s. is nonpositive since $y - x \in C^*$ and $P(\mathbf{y}|C) \in C$. Furthermore, $P(\mathbf{x}|C)/n$ maximizes $\Sigma_{i=1}^k x_i \ln p_i$ with $\mathbf{p} \in C$ and so $\Sigma_{i=1}^k x_i \ln(P(\mathbf{y}|C)_i) \leq \Sigma_{i=1}^k x_i \ln(P(\mathbf{x}|C)_i)$. Hence, $\Sigma_{i=1}^k y_i \ln(P(\mathbf{y}|C)_i) \leq \Sigma_{i=1}^k x_i \ln(P(\mathbf{x}|C)_i)$, or f is ISO on A_n.

The next result is an adaptation of Theorem 1.1 of Proschan and Sethuraman (1977). Let $\phi(\theta,x)$ be a nonnegative function defined on $(0,\infty) \times [0,\infty)$ satisfying the semigroup property,

$$\phi(\theta_1+\theta_2,x) = \int_0^\infty \phi(\theta_1,x-y)\phi(\theta_2,y) d\mu(y),$$

with μ either Lebesgue measure on $[0,\infty)$ or counting measure on the nonnegative integers.

THEOREM 2.2. *Let ϕ be as above, let f: $\mathcal{R}^k \to R$ be ISO* and let h be defined on* $(0,\infty)^k$ *by*

$$h(\theta) = \int_{[0,\infty)}\int_{[0,\infty)} \cdots \int_{[0,\infty)} f(\mathbf{x})\Pi_{i=1}^{k}\phi(\theta_i,x_i)d\mu(x_1) \ \ldots \ d\mu(x_k),$$

where the integral is assumed finite. Then h is ISO.*

LEMMA. *For $i,j \in \Omega$, set $\delta_{ij} = \delta_i/w_i - \delta_j/w_j$. $C^{*\mathbf{w}}$, the dual of the cone of isotone vectors, and K, the collection of vectors $\mathbf{x} = \Sigma_{\{(i,j)\in\Omega^2:\ i\neq j,\ i\preceq j\}} c_{ij}\delta_{ij}$ with the $c_{ij} \geq 0$ are equal.*

Proof. A proof similar to that given for the Remark on p. 49 of Barlow, Bartholomew, Bremner and Brunk (1972) shows that

$$C^{*\mathbf{w}} = \{\mathbf{y}: \Sigma_{i\in L} w_i y_i \geq 0 \text{ for every } L\in\mathcal{L} \text{ and } \Sigma_{i=1}^{k} w_i y_i = 0\}.$$

For $L\in\mathcal{L}$, $\alpha,\beta \in \Omega$ with $\alpha \preceq b$ and $\alpha \neq \beta$,

$$\Sigma_{i\in L}(\delta_{\alpha,\beta})_i w_i = \begin{cases} 0 & \text{if } \alpha \notin L \\ 1 & \text{if } \alpha\in L \text{ but } \beta\notin L \\ 0 & \text{if } \alpha,\beta \in L. \end{cases}$$

So $K \subset C^{*\mathbf{w}}$ and hence $K^{*\mathbf{w}} \supset (C^{*\mathbf{w}})^{*\mathbf{w}}$. As Dykstra (1981) observed, $(C^{*\mathbf{w}})^{*\mathbf{w}} = C$ if C is a closed convex cone. This can also be shown using the following: the result holds when $\mathbf{w} = \mathbf{e}_k$, ie. $(C^*)^* = C$ for C closed, (cf. Rockafeller (1970, p. 121)) and $C^{*\mathbf{w}} = \{(y_1/w_1, \ \ldots \ , y_k/w_k): \mathbf{y} \in C^*\}$ (cf. Barlow and Brunk (1972)). Suppose that $\mathbf{z} \in K^{*\mathbf{w}} - C$, that is $\mathbf{z}$ is not isotone and $\Sigma_{i=1}^{k} w_i z_i x_i \leq 0$ for each $\mathbf{x} \in K$. Now if $\mathbf{z}$ is not isotone there exist $\alpha,b \in \Omega$ with $\alpha \preceq \beta$, $\alpha \neq \beta$ and $z_\alpha > z_\beta$ and so $\Sigma_{i=1}^{k} w_i z_i(\delta_{\alpha,\beta})_i = z_\alpha - z_\beta > 0$. This contradiction implies that $K^{*\mathbf{w}} = C$ or $C^{*\mathbf{w}} = (K^{*\mathbf{w}})^{*\mathbf{w}} = K$. □

Proof. (Theorem 2.2) Let $\mathbf{w} = \mathbf{e}_k$ and consider $\theta'' >>^* \theta'$, then $\theta' - \theta'' \in C^*$. Hence, $\theta' = \theta'' + \Sigma_{\{i\leq j, i\neq j\}} c_{ij}\delta_{ij}$ with $c_{ij} \geq 0$. So it suffices to show that for arbitrary θ, $h(\theta + c_{ij}\delta_{ij}) \leq h(\theta)$, but this can be shown using the proof of Theorem 3.3 of Robertson and Wright (1982). □

Suppose that k independent Poisson processes are each observed for T units of time and that the intensity of the ith process is θ_i. The likelihood ratio test of $\theta_1 = \theta_2 = \ \ldots \ = \theta_k$ when the alternative is restricted by the trend, θ is isotone, rejects for large values of

$$T_{01} = -2 \ln \lambda = 2\{\Sigma_{i=1}^{k} X_i \ln(P(\mathbf{X}|C)_i) - (\Sigma_{i=1}^{k} X_i)\ln(\Sigma_{i=1}^{k} X_i/k)\}$$

where λ is the likelihood ratio and $\mathbf{X} = (X_1, X_2, \ \ldots \ , X_k)$ with the X_i independent Poisson variables and $E(X_i) = \theta_i T$. The family of Poisson densities satisfies the semigroup property with μ counting measure on $\{0, 1, \ldots\}$, $-(\Sigma_{i=1}^{k} X_i)\ln(\Sigma_{i=1}^{k} x_i/k)$ is ISO* and we have seen earlier that $\Sigma_{i=1}^{k} x_i \ln(P(\mathbf{x}|C)_i)$ is ISO*. Hence, Theorem 2.2 shows that this test has power function that is ISO*. This result could also have been obtained from Theorem 2.1 since conditioning on the total number of occurrences, $\Sigma_{i=1}^{k} X_i$, leads to a multinomial testing situation. However, this approach is more direct.

THEOREM 2.3. *Suppose $\{P_\theta: \theta \in \Theta\}$ is a family of probability measures on the Borel subsets of $\mathcal{R}^k$ with $\Theta \subset \mathcal{R}^k$ and suppose that if $\mathbf{X}$ has distribution P_θ then $\mathbf{X}-\theta$ has the distribution Q which is independent of θ. If f: $R^k \to R$ is ISO and h: $\Theta \to R$ is defined by $h(\theta) = \int f(\mathbf{x})dP_\theta(\mathbf{x})$ (which is assumed finite for each $\theta \in \Theta$), then h is ISO on Θ.*

The proof of Theorem 2.3 is just like that given for the totally ordered case (cf. Robertson and Wright (1982)) and in fact, the result holds for any cone ordering (cf. Marshall, Walkup and Wets (1967)).

Suppose X_{ij}, $j = 1,2, \ldots, n$ and $i = 1,2, \ldots, k$, are independent normal variables with mean θ_i and common variance σ^2. The estimator $\hat{\sigma}^2 = \Sigma_{i=1}^k \Sigma_{j=1}^n (X_{ij} - \bar{X}_i)^2/(k(n-1))$ for σ^2 is independent of $\hat{\theta}_i = \bar{X}_i = \Sigma_{j=1}^n X_{ij}/n$. To test $\theta_1 = \theta_2 = \ldots = \theta_k$ with the alternative restricted by, θ is isotone, one could use $T = \Sigma_{\{i \preceq j, i \neq j\}}(\hat{\theta}_j - \hat{\theta}_i)/\hat{\sigma}$, or more generally

$$T_c = \sqrt{n}\ \Sigma_{i=1}^k c_i\hat{\theta}_i/[(\Sigma_{i=1}^k c_i^2)^{1/2}\hat{\sigma}] \text{ with } \Sigma_{i=1}^k c_i = 0.$$

Of course, this test rejects for $T_c \geq t$ where t is the $100(1-\alpha)$ percentile of the T distribution with $k(n-1)$ degrees of freedom. The power function is translation invariant, ie. the power is the same at θ and $\theta + c\mathbf{e}_k$, and so it is ISO if it is ISO*. The distribution of $\hat{\sigma}$ is independent of θ and the power at θ is given by

$$E(P_\theta[\Sigma_{i=1}^k c_i\hat{\theta}_i \geq t(\Sigma_{i=1}^k c_i^2)^{1/2}\hat{\sigma}/\sqrt{n}\,|\hat{\sigma}]).$$

So it suffices to show that for each positive a, $P_\theta[\Sigma_{i=1}^k c_i\theta_i \geq at]$ is ISO*, but $\hat{\theta}' >>^* \hat{\theta}$ implies that $\hat{\theta} - \hat{\theta}' \in C^*$ and so $\Sigma_{i=1}^k c_i(\hat{\theta}_i' - \hat{\theta}_i) \geq 0$ if $\mathbf{c} \in C$. Hence if the vector $\mathbf{c}$ is isotone with respect to $\precsim$, then the power function is ISO.

In the case of T, c_i equals card. $\{\ell : \ell \preceq i\}$-card. $\{\ell : \ell \succeq i\}$ which is easily seen to be isotone. For the simple tree ordering, $1 \preceq i$, $i = 2, \ldots, k$, this choice of $\mathbf{c}$ is $(-k+1,1,1, \ldots ,1)$ and for the loop ordering, ie. $1 \preceq i \preceq k$ for $i = 2, \ldots, k-1$, this choice of $\mathbf{c}$ is $(-k+1,0, \ldots, 0,k-1)$. The test for the simple tree case is discussed in Barlow, Bartholomew, Bremner and Brunk (1972, p. 188) and it is argued there that this choice of $\mathbf{c}$ provides the optimum set of scores.

Robertson and Wright (1982) consider the likelihood ratio test for this testing problem with a total order, unequal sample sizes and known variances which are not necessarily equal. The arguments given there also show that the likelihood ratio test in the partially ordered case has power that is ISO.

Robertson and Wegman (1978) developed the likelihood ratio test for H_1: θ is isotone with respect to $\preceq$ versus H_2: $\sim H_1$ for exponential families. In the normal means case with known variances and $w_i = n_i/\sigma_i^2$, the test statistic is $T_{12} = \Sigma_{i=1}^k w_i(\hat{\theta}_i - P_w(\hat{\theta}|C)_i)^2$ where $P_w(\cdot|C)$ denotes the projection with respect to the distance function $d^2(\mathbf{x},\mathbf{y}) = \Sigma_{i=1}^k w_i(x_i - y_i)^2$. It is easy to show that neither T_{12} nor its negative is ISO*. As in Robertson and Wright (1982), we define another measure of conformity $\mathbf{x} \gtrsim \mathbf{y}$ provided $\mathbf{x} - \mathbf{y} \in C$. In the totally ordered case, $\mathbf{x} \gtrsim \mathbf{y}$ implies $\mathbf{x} >> \mathbf{y}$, but the converse is not true. However, in the partially ordered case this implication is not valid in general (For an example, consider $k=3$, the only order restriction is $2 \succeq 1$, $\mathbf{x} = (0,0,0)$, $\mathbf{y} = (1,1,-2)$ and $L = \{3\}$.) A function f: $\mathcal{R}^k \to R$ is ISO** provided $f(\mathbf{x}) \geq f(\mathbf{y})$ for all $\mathbf{x},\mathbf{y} \in \mathcal{R}^k$ with $\mathbf{x} \gtrsim \mathbf{y}$. The analogue of Remark 1.4, $x \gtrsim y$ if and only if $f(\mathbf{x}) \geq f(\mathbf{y})$ for all f which are ISO**, is easy to establish. (Note that $f(\mathbf{x}) = x_j - x_i$ is ISO** if $i \preceq j$.) Furthermore, since $\gtrsim$ is a cone ordering, Theorem 2.3 remains valid if ISO is changed to ISO**. Theorem 2.1 of Robertson and Wegman (1978) shows that the negative of $t_{12}(\mathbf{x}) = \Sigma_{i=1}^k w_i(x_i - P_w(\mathbf{x}|C)_i)^2$ is ISO**. So the modification of Theorem 2.3 which applies to $\gtrsim$ shows that if $\theta \gtrsim \theta'$, then the power of T_{12} at θ' is at least as large as at θ. Furthermore, $\theta \in C$ and θ' imply that $\theta - \theta \in C$ or $\theta \gtrsim \theta'$. Hence, H_0: θ is constant, is least favorable within H_1 and Robertson and Wegman (1978) have shown that under H_0, T_{12} has a chi-bar-squared distribution.

3. Comments. The problem of measuring the degree of conformity to an arbitrary partial order is a very broad one and in particular situations better measures may exist. In fact, we have noticed that none of the measures studied here are applicable to all the situations considered. In studying location parameters which are not related to the scale parameters, as in the normal case, $>>$ is preferred, but for cases such as that of Poisson means, where the location and scale parameters are related, $>>^*$ is more appropriate. We also found that $\gtrsim$ was useful when the null hypothesis stipulates that a collection of normal means satisfies a trend.

Because of the breadth of the problem it should not be surprising that in some special cases one can find a pair of parameter sets for which one of the orderings doesn't agree with our intuition. However, the measures studied here do seem to be useful in a variety of testing situations.

There are a couple of basic results in the totally ordered case which relate projections and the measures of conformity that are not true in the partially ordered case. Theorem 2.2 of Robertson and Wright (1982) states that

$$P_{\mathbf{w}}(\mathbf{y}|C) = \inf\{\mathbf{z} \in C : \mathbf{z} >>^* \mathbf{y}\}$$

and as a corollary $\mathbf{x} >>^* \mathbf{y}$ implies $P_{\mathbf{w}}(\mathbf{x}|C) >>^* P_{\mathbf{w}}(\mathbf{y}|C)$ and $\mathbf{x} >> \mathbf{y}$ implies that $P_{\mathbf{w}}(\mathbf{x}|C) >> P_{\mathbf{w}}(\mathbf{y}|C)$. The same example serves to show that these results are not valid in the general partially ordered case.

Example. Suppose that $k = 3$, $1 \preceq 2 \succeq 3$, $\mathbf{w} = \mathbf{e}_3$, $\mathbf{x} = (0,4.5,4.5)$ *and* $y = (1,3,5)$. *Observe that* $\mathbf{x} >>^* \mathbf{y}$ *(and of course,* $\mathbf{x} >> \mathbf{y}$*),* $P_{\mathbf{w}}(\mathbf{x}|C) = \mathbf{x}$, $P_{\mathbf{w}}(\mathbf{y}|C) = (1,4,4)$ *(one could use the lower sets algorithm discussed in Barlow, Bartholomew, Bremner and Brunk (1972)), but* $P_{\mathbf{w}}(\mathbf{x}|C) >> P_{\mathbf{w}}(\mathbf{y}|C)$ *is not true.*

The Remark on p. 1236 of that paper is also not valid for arbitrary partially ordered situations. It states that if $\phi \neq A \subset \mathcal{R}^k$ and A has a lower bound with respect to $>>(>>^*)$ then A has a greatest lower bound with respect to $>>(>>^*)$ and in the case of $>>^*$ the greatest lower bound is unique. It is not difficult to construct examples with A a set with two elements which has a lower bound with respect to $>>^*$ (and of course then with respect to $>>$) but not a greatest lower bound.

REFERENCES

Barlow, R. E., Bartholomew, D. J., Bremner, J. M. and Brunk, H. D. (1972). *Statistical Inferences Under Order Restrictions*, Wiley, New York.

Barlow, R. E. and Brunk, H. D. (1972). The isotonic regression problem and its dual. *J. Amer. Statist. Assoc. 67* 140–147.

Bartholomew, D. J. (1959). A test of homogeneity for ordered alternatives I, II. *Biometrika 46,* 36–48, 328–335.

Chacko, V. J. (1966). Modified chi-square tests for ordered alternatives. *Sankhyā Ser. B 28* 185–190.

Dykstra, Richard L. (1981). Dual convex cones of order restrictions with applications. *Technical Report No. 111,* Department of Statistics, University of Missouri-Columbia.

Marshall, A. W., Walkup, B. W. and Wets, R. J. B. (1967). Order-preserving functions: applications to majorization and order statistics. *Pacific J. Math. 23* 569-584.

Proschan, F. and Sethuraman, J. (1977). Schur functions in statistics I. The preservation theorem. *Ann. Statist. 5* 256–262.

Robertson, Tim (1978). Testing for and against an order restriction on multinomial parameters. *J. Amer. Statist. Assoc. 73* 197–202.

Robertson, Tim and Wegman, Edward J. (1978). Likelihood ratio tests for order restrictions in exponential families. *Ann. Statist. 6* 485–505.

ROBERTSON, TIM and WRIGHT, F. T. (1981). One-sided comparisons for treatments with a control. *Technical Report No. 78,* Department of Statistics, University of Iowa.

ROBERTSON, TIM and WRIGHT, F. T. (1982). On measuring the conformity of a parameter set to a trend with applications. *Ann. Statist. 10* 1234–1245.

ROCKAFELLER, R. T. (1970). *Convex Analysis.* Princeton University Press, Princeton, New Jersey.

SYMPOSIUM ON INEQUALITIES IN STATISTICS AND PROBABILITY

October 27–30, 1982

The University of Nebraska-Lincoln
Lincoln, Nebraska

Sponsored by
The National Science Foundation
Office of Naval Research
The University of Nebraska-Lincoln

PROGRAM

(Revised October 1982)

WEDNESDAY, OCTOBER 27

7:30–10:00 a.m.	*Registration*: Second floor registration desk, Nebraska Center *Late Registration*: Second floor registration desk, Nebraska Center
8:35 a.m.	*Welcome*, **Martin A. Massengale**, Chancellor, and **Earl J. Freise**, Assistant Vice Chancellor for Research, University of Nebraska-Lincoln
8:50–10:20 a.m.	A. *Matrix Related Inequalities* Chairman: **C.R. Rao**, University of Pittsburgh
8:50	A-1. Eigenvalue Inequalities for Random Evolutions, **Joel E. Cohen**, the Rockefeller University
9:35	A-2. Some Matrix Inequalities and Applications, **Marshall Freimer** and **Govind S. Mudholkar***, University of Rochester
10:20 a.m.	COFFEE
10:40–12:10 p.m.	B. *Multivariate Majorization and Dilations* Chairman: **Frank Proschan**, Florida State University
10:40	B-1. A Review of Multivariate Majorization, **Samuel Karlin**, Stanford University, and **Yosef Rinott***, Stanford University and the Hebrew University of Jerusalem
11:25	B-2. On Approximate Dilations with Applications to Statistics, **J.H.B. Kemperman**, University of Rochester
12:10 p.m.	GROUP LUNCH, The Nebraska Room
1:40–3:10 p.m.	C. *Stochastic Optimization and Rearrangement Inequalities* Chairman: **F.T. Wright**, University of Missouri-Rolla
1:40	C-1. Approximations and Error Bounds in Stochastic Optimization, **Roger J.B. Wets**, University of Kentucky and International Institute for Applied Systems Analysis

2:25 C-2. Stochastic Versions of Rearrangement Inequalities, with Applications to Statistics, **Catherine A. D'Abadie***, Western Electric, and **Frank Proschan**, Florida State University

3:10 p.m. COFFEE

3:30– D. *Probabilities of Geometric Regions*
5:00 p.m. Chairman: **D.R. Jensen**, Virginia Polytechnic Institute and State University

3:30 D-1. Ordering of Polyhedral Scheffé Regions with Respect to Spherically Symmetric Distributions: A Group-Theoretic Approach, **R. Bohrer**, University of Illinois, Urbana-Champaign, and **Henry P. Wynn***, Imperial College

4:15 D-2. Slepian Regions, **S.W. Dharmadhikari**, Southern Illinois University, and **Kumar Joag-Dev***, University of Illinois, Urbana-Champaign and Old Dominion University

8:00 p.m. INFORMAL PARTY, The Nebraska Room

THURSDAY, OCTOBER 28

8:40– E. *Moment and Markov Inequalities*
10:10 a.m. Chairman: **J.H.B. Kemperman**, University of Rochester

8:40 E-1. Moment Inequalities with Applications to Regression and Time Series Models, **T.L. Lai***, Columbia University, and **C.Z. Wei**, University of Maryland

9:25 E-2. Extensions of Markov Inequality, **Albert W. Marshall**, University of British Columbia

10:10 a.m. COFFEE and GROUP PICTURE

10:40– F. *Perspectives on Inequalities and Entropy*
12:10 p.m. *Functions*
Chairman: **Michael D. Perlman**, University of Washington

10:40 F-1. Inequalities—Some Perspectives, **Ingram Olkin**, Stanford University

11:25 F-2. Convexity Properties of Entropy Functions, **C.R. Rao**, University of Pittsburgh

12:10 p.m. GROUP LUNCH, The Nebraska Room

1:40– G. *Association and FKG Inequalities*
3:10 p.m. Chairman; **Ingram Olkin**, Stanford University

1:40 G-1. Using the FKG Inequality for Combinatorial Probability Inequalities, **Lawrence A. Shepp**, Bell Laboratories

2:25 G-2. Asymptotic Independence and Limit Theorems for Associated Variables, **Charles M. Newman**, University of Arizona and New York University

3:10 p.m. COFFEE

3:30– H. *Inequalities for Selecting and Ordering*
5:00 p.m. *Populations*
Chairman: **Milton Sobel**, University of California at Santa Barbara and Columbia University

3:30 H-1. On Some Order Relations and Monotonicity Properties with Applications to Selection and Ranking, **Shanti S. Gupta***, Purdue University, **Deng-Yuan Huang**, National Taiwan Normal University, and **S. Panchapakesan**, Southern Illinois University

4:15 H-2. Monotonicity in Population Selection: A Unified Approach, **Roger L. Berger***, Florida State University and North Carolina State University, and **Frank Proschan**, Florida State University

7:00 p.m. BANQUET, The Nebraska Room

FRIDAY, OCTOBER 29

7:50– I. *Contributed Papers, I*
8:50 a.m. Chairman: **K.M. Lal Saxena**, University of Nebraska-Lincoln

7:50 I-1. A-optimal Design Matrices $\mathbf{X}=(x_{ij})N \times n$ with $x_{ij}=-1, 0, 1$, **C.S. Wong** and **Joseph C. Masaro***, University of Windsor

8:05 I-2. Isotonic Regression and Recursive Partitioning, **Sue E. Leurgrans**, University of Wisconsin-Madison

8:20 I-3. Approximating Multivariate Distributions with Dependence Structures, **Joseph Glaz*** and **Bruce McK. Johnson**, University of Connecticut

8:35 I-4. Stop Rule and Supremum Expectations of i.i.d. Random Variables: A Complete Comparison, **Robert P. Kertz**, Georgia Institute of Technology

8:50– J. *Contributed Papers, II*
9:35 a.m. Chairman: **Daniel P. Mihalko**, University of Nebraska-Lincoln

8:50 J-1. Doob's Inequality Revisited, **David C. Cox**, Battelle's Columbus Laboratories

9:05 J-2. An Inequality from Convex Geometry, **Richard A. Vitale**, Claremont Graduate School

9:20 J-3. Function-Measure Duals via Inequalities, **A.M. Fink**, Iowa State University

9:35 a.m. COFFEE

9:55– K. *Trends and Order Restrictions*
12:10 p.m. Chairman: **Edward J. Wegman**, Office of Naval Research

9:55 K-1. On Measuring the Conformity of a Parameter Set to a Trend, **Tim Robertson***, University of Iowa, and **F.T. Wright**, University of Missouri-Rolla

10:40 K-2. Tests for and against Trends among Poisson Intensities, **Rhonda Magel** and **F.T. Wright***, University of Missouri-Rolla

11:25 K-3. Dual Convex Cones of Order Restrictions with Applications, **Richard L. Dykstra**, University of Missouri- Columbia and University of Iowa

12:10 p.m. GROUP LUNCH, The Nebraska Room

1:40– L. *Inequalities in Multivariate Analysis, I*
3:10 p.m. Chairman: **Shanti S. Gupta**, Purdue University

1:40 L-1. On Group Induced Orderings and Probability Inequalities, **Morris L. Eaton**, University of Minnesotà

2:25 L-2. Invariant Ordering and Order Preservation, **D.R. Jensen**, Virginia Polytechnic Institute and State University

3:10 p.m. COFFEE

3:30– M. *Convex and Stochastic Orderings*
5:00 p.m. Chairman: **Albert W. Marshall**, University of British Columbia

3:30 M-1. Convex-Ordering among Functions, with Applications to Reliability and Mathematical Statistics, **Wai Chan, Frank Proschan**, and **Jayaram Sethuraman***, Florida State University

4:15 M-2. Stochastic Ordering of Spacings from Dependent Random Variables, **Moshe Shaked***, Indiana University and University of Arizona, and **Y.L. Tong**, University of Nebraska-Lincoln

6:30 p.m. GROUP DINNER at Peking Garden

SATURDAY, OCTOBER 30

8:30– N. *Inequalities in Multivariate Analysis, II*
10:45 a.m. Chairman: **Y.L. Tong**, University of Nebraska-Lincoln

8:30 N-1. Unbiasedness of Tests for Variances and Covariance Matrices, **Michael D. Perlman**, University of Washington

9:15 N-2. On TP_2 and Log-Concavity, **Somesh Das Gupta***, University of Minnesota, and **S.K. Sarkar**, University of Pittsburgh

10:00 N-3. Adjusting P-values to Account for Selection over Dichotomies, **Ingram Olkin**, Stanford University, and **Glenn Shafer***, University of Kansas

10:45 a.m. COFFEE

11:00– O. *Inequalities in Reliability*
12:30 p.m. Chairman: **Moshe Shaked**, Indiana University and University of Arizona

11:00 O-1. Inequalities for Coherent Systems, **Henry W. Block***, and **Wagner de Souza Borges**, University of Pittsburgh

11:45 O-2. Reliability Inequalities for General Structure Functions, **Henry W. Block** and **Thomas H. Savits***, University of Pittsburgh

12:30 p.m. GROUP LUNCH, The Nebraska Room