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. $a_i^p = c b_i^q$ 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 $a_{(1)} \geq a_{(2)} \geq \dots \geq a_{(n)} > 0$ denote the ordered values of $a_1, a_2, \dots, a_n$, $m \leq n$, $b_{[j]} = b_{(j)} + b_{(j+1)} + \dots + b_{(n)}$, the tail sum of the smallest b's and $p \geq 1$. Then Mudholkar, Freimer, and Subbaiah (1983) consider the norm $\phi(\mathbf{a}) = (\Sigma_{i=1}^m a_{(i)}^p)^{1/p}$ and show that

$$\Sigma_{i=1}^n a_i b_i \leq \{\Sigma_{i=1}^n a_{(i)}^p\}^{1/p} \{\Sigma_{i=1}^n b_{(i)}^q + (m-k)(b_{[k+1]}/(m-k))^q\}^{1/q}, \tag{1.4}$$

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 $\phi_m(\mathbf{a}) = \phi(a_{(1)}, a_{(2)}, \dots, a_{(m)})$, where $a_{(1)} \geq a_{(2)} \geq \dots \geq a_{(n)} \geq 0$ 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 $a_{(1)} \geq a_{(2)} \geq \dots \geq a_{(n)} \geq 0$ 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,

$$a_1 \geq a_2 \geq \dots \geq a_n \geq 0; b_1 \geq b_2 \geq \dots \geq b_n \geq 0, \tag{2.1}$$

and for $p \geq 1$ consider the problem:

$$\text{Maximize } \Sigma_{i=1}^n a_i b_i \text{ subject to } \Sigma_{i=1}^m a_i^p = 1. \tag{2.2}$$

Clearly the solution to (2.2) must satisfy $a_m = a_{m+1} = \dots = a_n$. 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) Minimize $f(x_1, x_2, \ldots, x_m)$ 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) $$\partial L/\partial x_i = 0, \quad i = 1,2, \ldots, m,$$
$$\lambda_i \geq 0 \quad \text{and } \lambda_i g_i(\mathbf{x}) = 0, \ i = 1,2, \ldots, s.$$

In the standard format our problem (2.3) can be written down as:

(2.7) Minimize $\Sigma_{i=1}^{m}(-c_j)\, x_j$ subject to $x_j - x_{j-1} \leqslant 0, j = 2,3, \ldots, m, -x_m \leqslant 0$,
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 $a_1 \geq a_2 \geq \ldots \geq a_n \geq 0$ and $b_1 \geq b_2 \geq \ldots \geq b_n \geq 0$ 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) $\phi(\epsilon_1 x_{i_1}, \epsilon_2 x_{i_2}, \ldots, \epsilon_n x_{i_n}) = \phi(\mathbf{x})$ for any permutation $(i_1, i_2, \ldots, i_n)$ of $(1, 2, \ldots, n)$ and $\epsilon_i = \pm 1$, $i=1, 2 \ldots, n$. 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 $x_{(1)} \geq x_{(2)} \geq \ldots x_{(n)} \geq 0$ 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 $\phi_m(\mathbf{x}) = \phi(x_{(1)}, x_{(2)}, \ldots, x_{(n)})$ 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*

$$\phi_m^\circ(\mathbf{y}) = \phi^\circ(y_{(1)}, y_{(2)}, \ldots, y_{(k)}, \bar{y}, \bar{y}, \ldots, \bar{y}), \tag{3.1}$$

where $\phi^\circ \in \phi_m$ is the polar of ϕ. $y_{(1)} \geq y_{(2)} \geq \ldots \geq y_{(n)} \geq 0$ 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

$$\int_0^N a(t)b(t)\,dt \leq \{\int_0^M a(t)^p\,dt\}^{1/p}\{\int_0^M \hat{b}(t)^q dt\}^{1/q}, \tag{4.2}$$

where

$$\begin{aligned} \hat{b}(t) &= b(t), \quad 0 \leq t \leq K \\ &= \{\int_K^N b(t)\,dt\}/(M-K), K \leq t \leq M. \end{aligned} \tag{4.3}$$

The defining equation for K is analogous to that in the discrete case, namely

$$b(K) = \int_K^N b(t)dt/(M-K). \tag{4.4}$$

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

$$\{\int_0^1 (1-t)\,g(t)\,dt\}^2 \leq \tfrac{1}{2}\int_0^{1/2} g^2(t)\,dt, \tag{4.5}$$

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

$$H_o: \theta_1 = \theta_2 = \ldots = \theta_k, \tag{5.1}$$

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)):

$$\max_{\mathbf{c}'\mathbf{1}=0} \mathbf{c}'\,\mathbf{x}/||\mathbf{c}||_p = \min_{\eta} ||\mathbf{x}-\eta\mathbf{1}||_q, \tag{5.2}$$

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

$$|\mathbf{c}'\mathbf{x}| \leq s\Sigma|c_i|\,\{\max_{i,j}(x_i - x_j)/s\}, \tag{5.3}$$

$$\text{and} \quad |\mathbf{c}'\,\mathbf{x}| \leq s(\Sigma_{i=1}^k c_i^2)^{1/2}\ \{(\Sigma_{i=1}^k x_i^2)^{1/2}/s\}, \tag{5.4}$$

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

$$\Sigma_{j=1}^{n}(1 + b_j') \leq \Sigma_{j=1}^{n}(N_j I_j)V_j. \tag{5.9}$$

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 $\mathscr{R}^n$ 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.