MAJORIZATION: ITS EXTENSIONS AND THE PRESERVATION THEOREMS

by

KOON WING CHENG

B.Sc., University of Hong Kong, 1974

A THESIS SUBMITTED IN PARTIAL FULFILLMENT OF

THE REQUIREMENTS FOR THE DEGREE OF

MASTER OF SCIENCE

in

THE FACULTY OF GRADUATE STUDIES

DEPARTMENT OF MATHEMATICS

and the

INSTITUTE OF APPLIED MATHEMATICS AND STATISTICS

We accept this thesis as conforming

to the required standard

THE UNIVERSITY OF BRITISH COLUMBIA

April, 1977

In presenting this thesis in partial fulfilment of the requirements for an advanced degree at the University of British Columbia, I agree that the Library shall make it freely available for reference and study. I further agree that permission for extensive copying of this thesis for scholarly purposes may be granted by the Head of my Department or by his representatives. It is understood that copying or publication of this thesis for financial gain shall not be allowed without my written permission.

Department of Mathematics

The University of British Columbia
2075 Wesbrook Place
Vancouver, Canada
V6T 1W5

Date April 21, 1977

Abstract

This thesis deals mainly with the orderings induced by majorization, the two weak majorizations and their associated inequalities. One of these weak majorizations has received some attention in the literature. However, the other one, being dual to the former, is totally overlooked. Schur functions which preserve these orderings are shown to have applications in statistics.

In Chapter 1, we discuss briefly the majorizations and their relations with Schur functions, bringing out the extensions to the weak majorizations where possible. In Chapter 2, we generalize the ordinary majorizations to those parametrized by a vector $\underline{p}$ of positive components. We discuss the properties of these new majorizations in a direction parallel to that of ordinary majorizations. The Schur functions are likewise generalized. In Chapter 3, we discuss the stochastic extensions of majorizations and the preservation theorem of Schur convexity due to Proschan and Sethuraman (1977). With a preservation theorem on monotonicity, we study the stochastic extensions of the weak majorizations. Some inequalities arising in some multivariate distributions are found to be direct consequences of the two preservation theorems. Finally, we consider the unbiasedness of a certain class of tests of significance.

Table of Contents

List of Figure

Acknowledgement

I would like to thank my supervisor, Dr. A.W. Marshall for his constant help and encouragement. Also, I am grateful to the University of British Columbia for the financial assistance. Finally, I wish to express my gratitude to Mrs. Evelyn Yang for her typing of this thesis.

Chapter 0

Notation and Conventions

For easier reading of this thesis, the following notation and conventions are employed throughout. Those not listed below will be specified within the context.

(1) Real numbers and vectors are denoted by small letters, usually by u,v,x,y,z; random variables and vectors are denoted by capital letters, usually by X,Y; vectors are underlined with components indicated by subscripts.

(2) Univariate functions are usually denoted by f,g,h and multivariate functions are usually denoted by F,G,H. Densities for both univariate and multivariate distributions are denoted by ϕ or $\phi(\cdot|\underline{\lambda})$ if parametrized by $\underline{\lambda}$.

(3) The decreasing order arrangement $x_{[1]} \geq \cdots \geq x_{[n]}$ is a permutation of the components $x_1,\ldots,x_n$ of $\underline{x}$. Reverse order statistics $X_{[1]} \geq \cdots \geq X_{[n]}$ are defined in like manner.

(4) If π is a permutation of $\{1,\ldots,n\}$, we denote by $\underline{x}^{\pi}$ the vector $(x_{\pi_1},\ldots,x_{\pi_n})$.

(5) Ordinary ordering of vectors and matrices is componentwise.

(6) Monotonicity refers to both univariate and multivariate functions, the latter with respect to the ordering defined in (5). Also, by increasing (decreasing) we shall mean non-decreasing (non-increasing).

(7) The abbreviations "iff" for "if and only if" and "i.i.d." for "independent and identically distributed" are used.

(8) Limits of integration will be from $-\infty$ to ∞ unless otherwise stated.

Chapter 1

A. Introduction.

In this chapter, a brief survey of the concept of majorization and Schur convex functions is discussed. But some properties of majorization will be generalized in Chapter 2 . The main theorem introduced here gives equivalent conditions of majorizations. These conditions will be extended to their stochastic versions in Chapter 3 .

B. Majorization and weak majorizations.

Majorization is a partial ordering defined for pairs of vectors both lying on the same hyperplane $\{ \underline{x} : x_1+\ldots+x_n = K \}$ in R^n.

Definition 1.1. $\underline{x}$ is said to be majorized by $\underline{y}$, in notation $\underline{x} \prec \underline{y}$, iff

$$\sum_{i=1}^{k} x_{[i]} \leq \sum_{i=1}^{k} y_{[i]} \quad \text{for} \quad k=1,\ldots,n-1$$

and

$$\sum_{i=1}^{n} x_{[i]} = \sum_{i=1}^{n} y_{[i]} \quad . \tag{1.1}$$

Note that we have the following relations:

(1) the last equality in (1.1) is the same as $\sum_{i=1}^{n} x_i = \sum_{i=1}^{n} y_i$;

(2) if $\underline{x} \prec \underline{y}$ and $\underline{y} \prec \underline{x}$, then $\underline{x}$ and $\underline{y}$ differ by a permutation of their components;

(3) if $\underline{x} \prec \underline{y}$, then $\underline{x}^{\pi} \prec \underline{y}^{\pi'}$ for any permutations π and π'.

If we extend the inequalities to the last equality as well, then we have one version of weak majorization.

<u>Definition 1.2</u>. $\underline{x}$ is said to be weakly majorized by $\underline{y}$, in notation $\underline{x} \underset{w}{\prec} \underline{y}$, iff $\sum_{i=1}^{k} x_{[i]} \leq \sum_{i=1}^{k} y_{[i]}$ for $k=1,\ldots,n$.

We can rewrite the system (1.1) in another manner, as:

$$\sum_{i=k}^{n} x_{[i]} \geq \sum_{i=k}^{n} y_{[i]} \quad \text{for } k=2,\ldots,n$$

and

$$\sum_{i=1}^{n} x_{[i]} = \sum_{i=1}^{n} y_{[i]} .$$

This suggests another extension to weak majorization of a second type:

<u>Definition 1.3</u>. $\underline{x} \overset{w}{\prec} \underline{y}$ iff $\sum_{i=k}^{n} x_{[i]} \geq \sum_{i=k}^{n} y_{[i]}$ for $k=1,\ldots,n$.

Each of these weak majorizations, $\underset{w}{\prec}$ and $\overset{w}{\prec}$, also defines a partial ordering on R^n. Unlike majorization, vectors related by these orderings need not be confined to a hyperplane.

Trivial results following from these definitions are:

<u>Proposition 1.4</u>. (1) $\underline{x} \prec \underline{y}$ iff $-\underline{x} \prec -\underline{y}$;

(2) $\underline{x} \underset{w}{\prec} \underline{y}$ iff $-\underline{x} \overset{w}{\prec} -\underline{y}$.

Marshall and Olkin (to appear) use (2) as the definition of $\overset{w}{\prec}$. But the properties of $\overset{w}{\prec}$, being dual to those of $\underset{w}{\prec}$, seldom appear in the literature.

Next we give two relations between majorization and weak majorizations.

<u>Proposition 1.5</u>. (1) $\underline{x} \underset{w}{\prec} (\overset{w}{\prec}) \underline{y}$ iff there exists $\underline{u}$ satisfying $\underline{x} \prec \underline{u}$ and $\underline{u} \leq (\geq) \underline{y}$,

(2) $\underline{x} \prec \underline{y}$ iff $\underline{x} \underset{w}{\prec} \underline{y}$ and $\underline{x} \overset{w}{\prec} \underline{y}$.

<u>Proof</u> : (1) If $\underline{x} \underset{w}{\prec} \underline{y}$, let k be the index for which y_k is the smallest component of $\underline{y}$. Take $d=(y_1 + \ldots + y_n) - (x_1 + \ldots + x_n) \geq 0$. Let $\underline{u}$ be the vector obtained from $\underline{y}$ by replacing y_k by y_k-d. Then it is easy to verify that $\underline{u}$ satisfies $\underline{x} \prec \underline{u}$ and $\underline{u} \leq \underline{y}$. The converse of (1), and (2) are trivial.

Also, $\underline{x} \overset{w}{\prec} \underline{y}$ iff $-\underline{x} \underset{w}{\prec} -\underline{y}$ iff $-\underline{x} \prec -\underline{u}$, $-\underline{u} \leq -\underline{y}$ for some $\underline{u}$ iff $\underline{x} \prec \underline{v}$ ($\equiv -\underline{u}$), $\underline{v} \geq \underline{y}$ for some $\underline{v}$.

We can also show that $\underline{x} \underset{w}{\prec} \underline{y}$ iff there exists $\underline{u}$ satisfying $\underline{x} \leq \underline{u}$ and $\underline{u} \prec \underline{y}$. But in this case, we obtain u_i by adding d_i to x_i. The proof is more involved because of the presence of the inequality constraints not to be violated.

<u>Definition 1.6</u>. A matrix with non-negative entries is called <u>doubly stochastic</u> (<u>doubly substochastic</u>) if each row sum and column sum is equal to (less than or equal to) 1.

The following theorem characterizes $\prec$, $\underset{w}{\prec}$ and $\overset{w}{\prec}$ in terms of doubly stochastic and substochastic matrices.

<u>Theorem 1.7</u>. (1) $\underline{x} \prec \underline{y}$ iff there exists a doubly stochastic matrix P such that $\underline{x} = \underline{y}P$.

(2) $\underline{x} \underset{w}{\prec} \underline{y}$ in R_+^n ($\underline{x} \overset{w}{\prec} \underline{y}$ in R_-^n) iff there exists a doubly substochastic matrix P such that $\underline{x} = \underline{y}P$, where $R_+(R_-)$ is the set of non-negative (non-positive) numbers.

The proof is omitted here because this theorem is not related to the following discussion. Those interested are referred to Marshall and Olkin (to appear).

The next theorem is a powerful tool in majorization, as it reduces many proofs concerning majorization from R^n to R^2.

<u>Theorem 1.8</u>. $\underline{x} \prec \underline{y}$ iff there exist $\underline{u}^1, \ldots, \underline{u}^h$ such that $\underline{x} \equiv \underline{u}^0 \prec \underline{u}^1 \prec \cdots \prec \underline{u}^h \prec \underline{u}^{h+1} \equiv \underline{y}$ where for $i=0, 1, \ldots, h$, $\underline{u}^i$ and $\underline{u}^{i+1}$ differ in only two components.

<u>Proof</u>: Notice that for any vector $\underline{z}$, $\underline{z}$ and $\underline{z}^1 = (z_{[1]}, \ldots, z_{[n]})$ differ by a permutation which can be represented as products of transpositions. So we can assume that $x_1 \geq \cdots \geq x_n$, and $y_1 \geq \cdots \geq y_n$.

Now we proceed by induction on n. If any of the inequalities in (1.1) is actually an equality, then for some k,

$$\underline{x}^a \equiv (x_1,\ldots,x_k) \prec (y_1,\ldots,y_k) \equiv \underline{y}^a ,$$

$$\underline{x}^b \equiv (x_{k+1},\ldots,x_n) \prec (y_{k+1},\ldots,y_n) \equiv \underline{y}^b .$$

By the induction assumption, $\underline{x}^a \prec \underline{v}^1 \prec \ldots \prec \underline{v}^r \prec \underline{y}^a$ and $\underline{x}^b \prec \underline{w}^1 \prec \ldots \prec \underline{w}^s \prec \underline{y}^b$, so $\underline{x} = (\underline{x}^a,\underline{x}^b) \prec (\underline{v}^1,\underline{x}^b) \prec \ldots \prec (\underline{v}^r,\underline{x}^b) \prec (\underline{y}^a,\underline{x}^b) \prec (\underline{y}^a,\underline{w}^1) \prec \ldots \prec (\underline{y}^a,\underline{w}^s) \prec (\underline{y}^a,\underline{y}^b) = \underline{y}$, where two adjacent vectors differ in only two components. If all the inequalities are strict inequalities, we take

$$\delta = \min_{k=1,\ldots n-1} \{ (y_1+\ldots+y_k)-(x_1+\ldots+x_k) \} > 0$$

and $$\underline{x}^\delta = (x_1+\delta,x_2,\ldots,x_{n-1},x_n-\delta) .$$

Then $\underline{x} \prec \underline{x}^\delta$ and we can apply the preceding arguments to $\underline{x}^\delta$ and $\underline{y}$, because $\underline{x}^\delta \prec \underline{y}$ and at least one of the inequalities of their components is an equality.

It is interesting to note that if $\underline{x}$ and $\underline{y}$ are similarly ordered, then we can choose the $\underline{u}^i$ in such a way that they are all ordered in the same manner as $\underline{x}$ and $\underline{y}$.

The proof presented here is different from the one originally given by Hardy, Littlewood and Pólya (1952), who considered the discrepancies (number of different components) between $\underline{x}$ and $\underline{y}$. But it will be seen in Chapter 2 that the above proof can be employed to prove a more general result.

Because $\underline{u}^i \prec \underline{u}^{i+1}$ and they differ in only two components, we can relate them by a doubly stochastic matrix as :

$$\underline{u}^{i} = \underline{u}^{i+1}\begin{pmatrix} 1 & & & & & \\ & \ddots & & & & \\ & & \alpha & & \bar{\alpha} & \\ & & & \ddots & & \\ & & \bar{\alpha} & & \alpha & \\ & & & & & \ddots \\ & & & & & & 1 \end{pmatrix}$$

for some $\alpha, \bar{\alpha}$ satisfying $\alpha>0$, $\bar{\alpha}>0$ and $\alpha+\bar{\alpha}=1$. A matrix of this form, with 1 in the diagonal entries except the two α, and 0 in all the other entries except the two $\bar{\alpha}$, corresponds to a "T-transform". The main idea of Theorem 1.8 is that: If $\underline{x} \prec \underline{y}$, then $\underline{x}$ can be derived from $\underline{y}$ by a finite number of T-transforms.

C. Schur functions.

Now we consider functions that preserve majorization.

Definition 1.9. Let S be an arbitrary subset of R^n. A function $F:R^n\ (S) \longrightarrow R$ is said to be a Schur convex function (on S) iff for any $\underline{x}, \underline{y}\ (\varepsilon S)$, $\underline{x} \prec \underline{y} \Longrightarrow F(\underline{x}) \leq F(\underline{y})$, and F is Schur concave iff $-F$ is Schur convex.

It follows from this definition that a Schur convex (or Schur concave) function (on S) is necessarily symmetric, because if $\underline{x}'$ is obtained from $\underline{x}$ by a permutation of its components (and if $\underline{x}, \underline{x}' \varepsilon S$), then $\underline{x} \prec \underline{x}' \prec \underline{x}$, implying $F(\underline{x}) \leq F(\underline{x}') \leq F(\underline{x})$, so that $F(\underline{x}) = F(\underline{x}')$.

The extension to weak majorization is the following:

Proposition 1.10. F is Schur convex and increasing (decreasing) iff for any $\underline{x}, \underline{y}$, $\underline{x} \underset{w}{\prec} (\overset{w}{\prec}) \underline{y} \Longrightarrow F(\underline{x}) \leq F(\underline{y})$.

Proof : Necessity. Suppose that F is Schur convex and increasing. Then for $\underline{x} \prec_w \underline{y}$, there exists $\underline{u}$ satisfying $\underline{x} \prec \underline{u}$, $\underline{u} \leq \underline{y}$. So it follows that $F(\underline{x}) \leq F(\underline{u}) \leq F(\underline{y})$.
For sufficiency, notice that $\underline{x} \prec_w \underline{y}$ whenever $\underline{x} \prec \underline{y}$ or $\underline{x} \leq \underline{y}$, hence $F(\underline{x}) \leq F(\underline{y})$ whenever $\underline{x} \prec \underline{y}$ or $\underline{x} \leq \underline{y}$. Therefore F is Schur convex and increasing.
The proof for the other version of weak majorization is similar.

The following theorem characterizes differentiable Schur functions in terms of their derivatives.

Theorem 1.11 (Ostrowski, 1952). A differentiable function F is Schur convex iff it satisfies :

(1) F is symmetric,

(2) for all i, j, $(x_i - x_j)(\frac{\partial F}{\partial x_i} - \frac{\partial F}{\partial x_j}) \geq 0$.

The proof in a slightly different setting will be given in Chapter 2 .

This theorem is important in identifying Schur convex functions because conditions (1) and (2) are usually easier to verify than the direct definition in the case of differentiable functions. It should be noted that the symmetry property (1) is necessary as can be illustrated by the following example.

Example 1.12. Let $F: R^2 \longrightarrow R$ be defined by

$$F(x_1, x_2) = (x_1 - x_2)^2 \quad \text{if} \quad x_1 \geq x_2$$
$$= 0 \qquad\qquad \text{otherwise} \quad ,$$

then F is differentiable with derivatives satisfying (2). But F is not Schur convex because $(3/4,1/4) \prec (0,1)$, but $F(0,1) = 0 < 1/4 = F(3/4,1/4)$.

<u>Corollary 1.13</u>. If f is differentiable, then $F(\underline{x}) = \sum_{i=1}^{n} f(x_i)$ is Schur convex iff f is convex.

A relation between Schur convex functions and convexity is:

<u>Proposition 1.14</u>. If F is symmetric and convex, then F is Schur convex.

<u>Proof</u> : By Theorem 1.8, we can reduce the proof to R^2. So let $x_1 \geq x_2$, $y_1 \geq y_2$. By the remark following Theorem 1.8, $\underline{x} \prec \underline{y} \Longrightarrow (x_1,x_2) = \lambda(y_1,y_2)+\bar{\lambda}(y_2,y_1)$ for some $\lambda,\bar{\lambda} \geq 0$, $\lambda+\bar{\lambda}=1$.

Hence
$$F(x_1,x_2) \leq \lambda F(y_1,y_2) + \bar{\lambda}F(y_2,y_1)$$
$$= \lambda F(y_1,y_2) + \bar{\lambda}F(y_1,y_2) = F(y_1,y_2) .$$

<u>Proposition 1.15</u>. If f is convex, then $F(\underline{x}) = \sum_{i=1}^{n} f(x_i)$ is symmetric and convex.

The proof is simple and is omitted.

Now we state the theorem which will be extended to the stochastic versions in Chapter 3 .

<u>Theorem 1.16</u>. The following conditions are equivalent:

(1) $\underline{x} \prec \underline{y}$,

(2) $F(\underline{x}) \leq F(\underline{y})$ for every Schur convex function F on R^n,

(3) $F(\underline{x}) \leq F(\underline{y})$ for every symmetric, convex function F on R^n,

(4) $\sum_{i=1}^{n} f(x_i) \leq \sum_{i=1}^{n} f(y_i)$ for every convex function f.

<u>Proof</u> : (1) $\Longrightarrow$ (2), (2) $\Longrightarrow$ (3), (3) $\Longrightarrow$ (4) follow respectively from Definition 1.9, Propositions 1.14 and 1.15.

For (4) $\Longrightarrow$ (1), we may assume $x_1 \geq x_2 \geq \dots \geq x_n$ and $y_1 \geq y_2 \geq \dots \geq y_n$. First notice that the functions $f(x) = x$ and $f(x) = -x$ are convex, putting them in (4) yield

$$\sum_{i=1}^{n} x_i \leq \sum_{i=1}^{n} y_i \quad \text{and} \quad -\sum_{i=1}^{n} x_i \leq -\sum_{i=1}^{n} y_i \text{ , so that } \sum_{i=1}^{n} x_i = \sum_{i=1}^{n} y_i .$$

Next we see that the functions $f(x) = (x-y_k)^+ \equiv \max\{0, x-y_k\}$, $k=1,\dots,n$ are convex, therefore

$$x_1 + \dots + x_k - ky_k = \sum_{i=1}^{k} f(x_i) \leq \sum_{i=1}^{n} f(x_i)$$

$$\leq \sum_{i=1}^{n} f(y_i) = \sum_{i=1}^{k} f(y_i) = y_1 + \dots + y_k - ky_k ,$$

so that $$\sum_{i=1}^{k} x_i \leq \sum_{i=1}^{k} y_i \quad \text{for } k=1,\dots n-1 .$$

Hence $\underline{x} \prec \underline{y}$.

The equivalence of (1) and (4) is proved by Hardy, Littlewood and Pólya (1929), and that of (1) and (3) is proved by Mirsky (1959). The extension of this theorem to the two versions of weak majorization is direct and simple.

<u>Theorem 1.17</u>. The following are equivalent:

(1) $\underline{x} \underset{w}{\prec} (\overset{w}{\prec}) \underline{y}$,

(2) $F(\underline{x}) \leq F(\underline{y})$ for every Schur convex and increasing (decreasing) function F ,

(3) $F(\underline{x}) \leq F(y)$ for every symmetric, convex and increasing (decreasing) function F ,

(4) $\sum_{i=1}^{n} f(x_i) \leq \sum_{i=1}^{n} f(y_i)$ for every convex and increasing (decreasing) function f .

The proof is contained in the previous discussion and is not repeated. The equivalence of (1) and (4) is proved by Tomić (1949).

Finally a geometric picture of majorization in R^3 is illustrated. In general, for a vector $\underline{z} \in R^3$ with different components, there are six permutations of the components. These six vectors are represented by the points A,B,C,D,E and F in the figure shown below. They all lie on the plane $x_1+x_2+x_3 = K$. If we let S be the set of vectors majorized by $\underline{z}$, and T the set of vectors majorizing $\underline{z}$, then in the figure, S is the convex hexagon ABCDEF, and T is that part of the plane outside the dotted lines. Furthermore

$$S = \{\underline{x}:\underline{x} \prec \underline{z}\} = \{\underline{x}:F(\underline{x}) \leq F(\underline{z}) \text{ for every Schur convex } F\} ,$$

$$T = \{\underline{x}:\underline{x} \succ \underline{z}\} = \{\underline{x}:F(\underline{x}) \geq F(\underline{z}) \text{ for every Schur convex } F\} .$$

If 1_A is the indicator function of the set A defined by

$$1_A(\underline{x}) = 1 \quad \text{if } \underline{x} \in A$$
$$= 0 \quad \text{if } \underline{x} \notin A \quad ,$$

then it is easy to verify that 1_S is Schur concave and 1_T is Schur convex.

For any Schur convex function F restricted to the plane $x_1+x_2+x_3 = K$, F increases along any ray on the plane from the centre $(K/3,K/3,K/3)$.

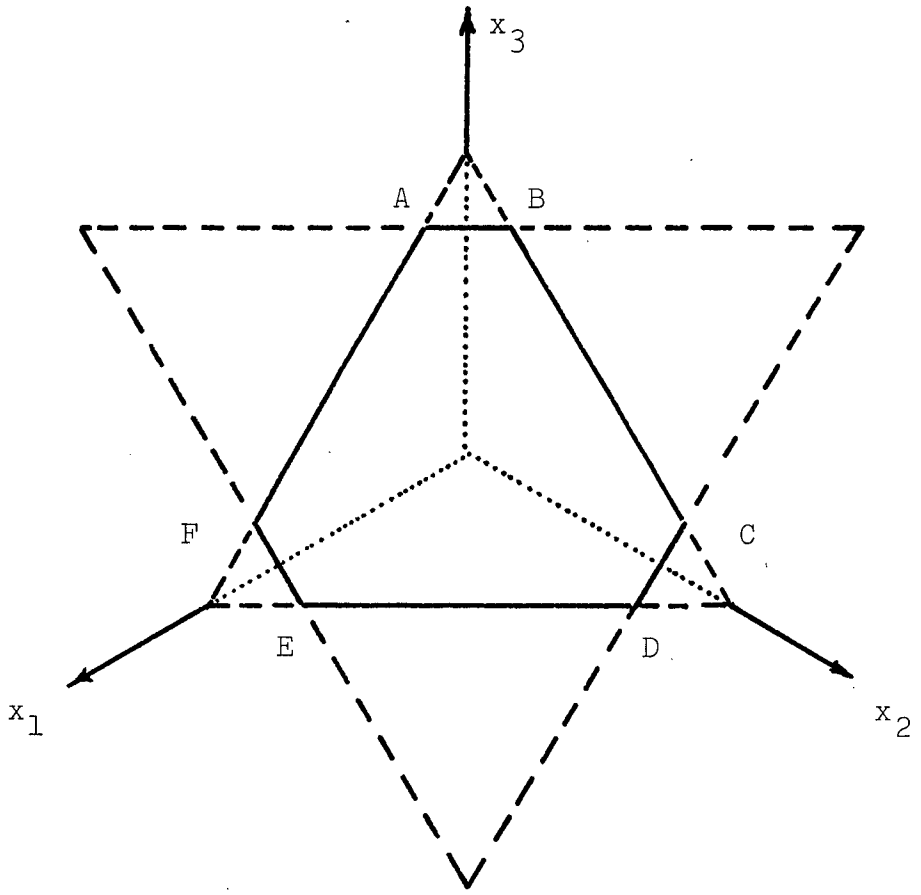

Figure. An illustration of majorization in R^3.

Chapter 2

A. Introduction.

In this chapter, we generalize majorization and weak majorizations to the p-majorizations parametrized by a vector $\underline{p}$ of positive components. The Hardy-Littlewood-Pólya inequality is extended to a more general form. The properties of the p-majorizations are discussed, with the characterization by a class of matrices similar to the doubly stochastic matrices. Functions preserving these orderings are considered.

Throughout this chapter, the parameter $\underline{p}$ is a vector with positive components. We shall adopt the following notation: if π is a permutation of $\{1, \ldots, n\}$, $\underline{x} \in D^{\pi}$ means that $x_{\pi_1} \geq \cdots \geq x_{\pi_n}$; in other words, the vector $\underline{x}^{\pi}$ obtained from permuting the components of $\underline{x}$ under π is in decreasing order. Moreover, we denote by $\underline{x} \in D$ if π is the identity mapping.

B. The p-majorizations.

The following is a generalization of majorization.

Definition 2.1. $\underline{x} \prec_p \underline{y}$ on D^{π} iff $\underline{x}, \underline{y} \in D^{\pi}$ (this means that the components of $\underline{x}$ and $\underline{y}$ are similarly ordered),

$$\sum_{i=1}^{k} p_{\pi_i} x_{\pi_i} \leq \sum_{i=1}^{k} p_{\pi_i} y_{\pi_i} \qquad \text{for } k=1, \ldots, n-1$$

and $\sum_{i=1}^{n} p_{\pi_i} x_{\pi_i} = \sum_{i=1}^{n} p_{\pi_i} y_{\pi_i}$ (or equivalently $\sum_{i=1}^{n} p_i x_i = \sum_{i=1}^{n} p_i y_i$).

The extensions to $\underline{x} \underset{w}{\prec}_p \underline{y}$ and $\underline{x} \overset{w}{\prec}_p \underline{y}$ are obtained by replacing the above system by $\sum_{i=1}^{k} p_{\pi_i} x_{\pi_i} \leq \sum_{i=1}^{k} p_{\pi_i} y_{\pi_i}$ and

$\sum_{i=k}^{n} p_{\pi_i} x_{\pi_i} \geq \sum_{i=k}^{n} p_{\pi_i} y_{\pi_i}$ for $k=1, \ldots, n$.

One remark about this p-majorization is that vectors related by this ordering have to by similarly ordered. Indeed, the ordering $\prec_p$ depends on the permutation π. The rigorous way of writing should be $\underline{x} \; \pi\!\prec_p \underline{y}$. However, the requirement $\underline{x}, \underline{y} \in D^{\pi}$ has made the dependence on π clear enough to be omitted.

Theorem 2.2 below is an extension of the Hardy-Littlewood-Pólya inequality which corresponds to the case in which all the p_i are equal.

<u>Theorem 2.2</u>. If $\sum_{i=1}^{n} p_i f(x_i) \leq \sum_{i=1}^{n} p_i f(y_i)$ for every convex f, then for any such that $\underline{x}, \underline{y}$
then for any π such that $\underline{x}, \underline{y} \in D^{\pi}$, we have $\underline{x} \prec_p \underline{y}$ on D^{π}.
Conversely, if for some π, $\underline{x} \prec_p \underline{y}$, on D^{π}, then

$\sum_{i=1}^{n} p_i f(x_i) \leq \sum_{i=1}^{n} p_i f(y_i)$ for any convex f.

<u>Proof</u> : We only prove the case that π is the identity mapping. For the proof of the general case, just replace all the subscripts i by π_i. First suppose that $\underline{x}, \underline{y} \in D$ and

$\sum_{i=1}^{n} p_i f(x_i) \leq \sum_{i=1}^{n} p_i f(y_i)$ for any convex f. Choosing $f(x)=x$ and $f(x)=-x$, we get $\sum_{i=1}^{n} p_i x_i = \sum_{i=1}^{n} p_i y_i$. Next choosing $f(x)=(x-y_k)^+$ for $k=1, \ldots, n-1$, we get

$$\sum_{i=1}^{k} p_i x_i - \sum_{i=1}^{k} p_i y_k \leq \sum_{i=1}^{k} p_i f(x_i) \leq \sum_{i=1}^{n} p_i f(x_i)$$

$$\leq \sum_{i=1}^{n} p_i f(y_i) = \sum_{i=1}^{k} p_i f(y_i) = \sum_{i=1}^{k} p_i y_i - \sum_{i=1}^{k} p_i y_k,$$

so $$\sum_{i=1}^{k} p_i x_i \leq \sum_{i=1}^{k} p_i y_i .$$

Conversely, suppose $\underline{x} \not\prec \underline{y}$ on D. We may assume that $x_k \neq y_k$ for $k=1, \ldots, n$. Define $A_0=0$, $B_0=0$, $A_k = \sum_{i=1}^{k} p_i x_i$,

$B_k = \sum_{i=1}^{k} p_i y_i$ and $Q_k = [f(x_k)-f(y_k)] \,/\, [x_k-y_k]$ for $k=1, \ldots, n$.

Because f is convex, we have $Q_k \geq Q_{k+1}$.

Thus $$\sum_{k=1}^{n-1} (A_k-B_k)(Q_k-Q_{k+1}) + (A_n-B_n)Q_n \leq 0 ,$$

therefore $$\sum_{k=1}^{n-1} A_k(Q_k-Q_{k+1}) + A_n Q_n \leq \sum_{k=1}^{n-1} B_k(Q_k-Q_{k+1}) + B_n Q_n ,$$

therefore $$\sum_{k=1}^{n} (A_k-A_{k-1})Q_k \leq \sum_{k=1}^{n} (B_k-B_{k-1})Q_k ,$$

so that $$\sum_{k=1}^{n} p_k[f(x_k)-f(y_k)] = \sum_{k=1}^{n} [(A_k-A_{k-1})-(B_k-B_{k-1})]Q_k \leq 0 .$$

It can be seen from the proof that the converse is true for

any p_i. This part is proved by Fuchs (1947). But the first part of the proof requires that all the p_i are positive (or negative). However, this condition is left out by Mitronović (1970), who quoted the sufficient condition of Fuchs as necessary and sufficient. It can be proved that for n=2, the theorem holds for arbitrary p_1 and p_2. Yet from the following example we see that it does not hold in general if the p_i are of different sign.

<u>Example 2.3</u>. Take $p_1=p_2=1$, $p_3=-1$, $y_1=x_1=x_2=x_3 > y_2=y_3$, then $\underline{x}, \underline{y} \in D$; and for any convex f , $p_1f(x_1)+p_2f(x_2)+p_3f(x_3) \leq p_1f(y_1)+p_2f(y_2)+p_3f(y_3)$, but $p_1x_1+p_2x_2 > p_1y_1+p_2y_2$.

Finally, we remark that if $\underline{x}$ and $\underline{y}$ are not similarly ordered, then the inequality $\sum_{i=1}^{n} p_i f(x_i) \leq \sum_{i=1}^{n} p_i f(y_i)$ can either be true or false. Take for instance, $f(5)+f(3)+2f(1) \leq f(6)+f(0)+2f(2)$ for any convex f because $(5,3,1,1) \prec (6,2,2,0)$. But $f(5)+f(5)+2f(1) \leq$ (or $\geq$) $f(6)+f(0)+2f(2)$ for any convex f is false because $(5,5,1,1)$ and $(6,3,3,0)$ cannot be ordered by $\prec$.

Just like the Hardy-Littlewood-Pólya inequality, it is possible to extend theorem 2.2 to convex functions that are increasing (or decreasing) by replacing $\prec_p$ by $\prec_{wp}$ (or $\overset{w}{\prec}_p$).

The ordering $\prec_p$ is quite similar to majorization. It can be derived by a finite number of transformations similar to T-transforms. Also, it can be characterized by matrices similar

to the doubly stochastic matrices.

Because of the property that $\underline{x} \prec \underline{y}$ on D^{π} iff $\underline{x}^{\pi} \prec^{\pi} \underline{y}^{\pi}$ on D, we only have to consider $\underline{x} \prec \underline{y}$ on D and then extend to D^{π} in the obvious way.

<u>Theorem 2.4</u>. If $\underline{x} \prec \underline{y}$ on D , then there exist $\underline{u}^1, \ldots, \underline{u}^h$ such that $\underline{x} \equiv \underline{u}^0 \prec \underline{u}^1 \prec \ldots \prec \underline{u}^h \prec \underline{u}^{h+1} \equiv \underline{y}$ on D, where for $i=0, 1, \ldots, h$, $\underline{u}^i$ and $\underline{u}^{i+1}$ differ in only two components.

<u>Proof</u> : By induction on n . We consider the following two cases:

(1) If any of the inequalities $\sum_{i=1}^{k} p_i x_i \leq \sum_{i=1}^{k} p_i y_i$, $k=1, \ldots, n-1$ is actually an equality, then we can write $\underline{x} \prec \underline{y}$ on D as:

$$\underline{x}^a \prec^a \underline{y}^a \text{ on } D, \quad \underline{x}^b \prec^b \underline{y}^b \text{ on } D ,$$

where $\underline{x} = (\underline{x}^a, \underline{x}^b)$, $\underline{y} = (\underline{y}^a, \underline{y}^b)$, $\underline{p} = (\underline{p}^a, \underline{p}^b)$.

By induction assumption, there exist $\underline{v}^i$, $i=1, \ldots, r$ and $\underline{w}^j$, $j=1, \ldots, s$ such that

$$\underline{x}^a \prec^a \underline{v}^1 \prec^a \ldots \prec^a \underline{v}^r \prec^a \underline{y}^a \quad \text{on } D ,$$
$$\underline{x}^b \prec^b \underline{w}^1 \prec^b \ldots \prec^b \underline{w}^s \prec^b \underline{y}^b \quad \text{on } D ,$$

and such that each two adjacent vectors above differ in only two components. Combining these vectors, we get

$$\underline{x} \doteq (\underline{x}^a, \underline{x}^b) \prec (\underline{v}^1, \underline{x}^b) \prec \ldots \prec (\underline{v}^r, \underline{x}^b) \prec (\underline{y}^a, \underline{x}^b)$$
$$\prec (\underline{y}^a, \underline{w}^1) \prec \ldots \prec (\underline{y}^a, \underline{w}^s) \prec (\underline{y}^a, \underline{y}^b) = \underline{y} \text{ on } D.$$

It is easy to verify that all these vectors are in decreasing order and that any two adjacent vectors differ in only two

components.

(2) If all inequalities are strict, take

$$\delta \overset{=}{=} \min_{k=1,\ \ldots,\ n-1} \{ \sum_{i=1}^{k} p_i(y_i - x_i) \} > 0$$

and $\quad \underline{x}^{\delta} = (x_1 + \delta/p_1,\ x_2,\ \ldots,\ x_{n-1},\ x_n - \delta/p_n)$.

Then $\quad \underline{x} \lessdot \underline{x}^{\delta} \lessdot \underline{y}$ on D .

For the n-1 component inequalities of $\underline{x}^{\delta} \lessdot \underline{y}$, at least one must be an inequality. Now application of (1) to $\underline{x}^{\delta}$ and $\underline{y}$ completes the proof.

C. Characterization by matrices.

Let $\mathcal{A}_p$ be a set of n×n matrices defined by

$\mathcal{A}_p = \{A \geq 0$: for any $\underline{y} \in D$, $\underline{y}A \in D$ and $\underline{y}A \lessdot \underline{y}$ on D$\}$.

The following theorem characterizes the matrices A of $\mathcal{A}_p$.

Theorem 2.5. If $A \geq 0$, then $A \in \mathcal{A}_p$ iff A satisfies:

(1) $\underline{e}A = \underline{e}$, where $\underline{e} = (1, \ldots, 1)$; i.e. columns of A sum to 1 ,

(2) $\sum_{j=1}^{k} a_{ij}$ is decreasing in i for k=1, ..., n-1,

(3) $A\underline{p}^t = \underline{p}^t$, where $\underline{p}^t$ is the transpose of $\underline{p}$.

Proof : Necessity. $\underline{e}, -\underline{e} \in D \Longrightarrow \underline{e}A \lessdot \underline{e}, -\underline{e}A \lessdot -e$ on D $\Longrightarrow$ $\underline{e}A = \underline{e}$. For k=1, ..., n-1, let $\underline{z}^k$ be the vector with the first k components equal to 1 and the rest equal to 0 . Then $\underline{z}^k \in D$ $\xRightarrow{\underline{z}^k \in D}$ $\sum_{j=1}^{k} a_{ij}$ is decreasing in i . Finally, $\underline{z}^k A \lessdot \underline{z}^k$ on D

$D \Longrightarrow \underline{z}^k A \underline{p}^t = \underline{z}^k \underline{p}^t \Longrightarrow \sum_{i=1}^{k} \sum_{j=1}^{n} a_{ij} p_j = \sum_{i=1}^{k} p_i$ for k=1, ..., n. But

this set of equalities is equivalent to $A\underline{p}^t=\underline{p}^t$.

Sufficiency. For $\underline{y} \in D$, let $\underline{x}=\underline{y}A$. For indices $h<k$,

$$x_h-x_k = \sum_{j=1}^{n} a_{hj}y_j-\sum_{j=1}^{n} a_{kj}y_j$$

$$= \sum_{m=1}^{n-1}[\sum_{j=1}^{m} a_{hj}-\sum_{j=1}^{m} a_{kj}](y_m-y_{m+1})+[\sum_{j=1}^{n} a_{hj}-\sum_{j=1}^{n} a_{kj}]y_n \geq 0 .$$

This implies $\underline{x} \in D$. Also,

$$\underline{yp}^t=\underline{x}A\underline{p}^t=\underline{xp}^t \Longrightarrow \sum_{i=1}^{n} p_i x_i=\sum_{i=1}^{n} p p_i y_i \quad .$$

For k=1, ..., n-1, applying conditions (1), (2) and (3), we obtain

$$\sum_{j=1}^{k} p_j x_j = \sum_{j=1}^{k} \sum_{i=1}^{n} y_i a_{ij} p_j$$

$$= \sum_{j=1}^{k} \sum_{i=1}^{k} y_i a_{ij} p_j + \sum_{j=1}^{k} \sum_{i=k+1}^{n} y_i a_{ij} p_j$$

$$\leq \sum_{j=1}^{k} \sum_{i=1}^{k} y_i a_{ij} p_j + \sum_{j=1}^{k} \sum_{i=k+1}^{n} y_k a_{ij} p_j$$

$$= \sum_{j=1}^{k} \sum_{i=1}^{k} (y_i-y_k) a_{ij} p_j + y_k \sum_{j=1}^{k} \sum_{i=1}^{n} a_{ij} p_j$$

$$= \sum_{j=1}^{k} \sum_{i=1}^{k} (y_i-y_k) a_{ij} p_j + y_k \sum_{j=1}^{k} p_j$$

$$= \sum_{j=1}^{k} \sum_{i=1}^{k} (y_i-y_k) a_{ij} p_j + y_k \sum_{i=1}^{k} \sum_{j=1}^{n} a_{ij} p_j$$

$$= \sum_{i=1}^{k} \sum_{j=1}^{k} y_i a_{ij} p_j + \sum_{i=1}^{k} \sum_{j=k+1}^{n} y_k a_{ij} p_j$$

$$\leq \sum_{i=1}^{k} \sum_{j=i}^{k} y_i a_{ij} p_j + \sum_{i=1}^{k} \sum_{j=k+1}^{n} y_i a_{ij} p_j$$

$$= \sum_{i=1}^{k} \sum_{j=1}^{n} y_i a_{ij} p_j = \sum_{i=1}^{k} p_i y_i \ .$$

Hence $\underline{x} \prec_{\mathcal{P}} \underline{y}$ on D .

From the definition of $\mathcal{A}_p$ and the equivalent conditions given in the above theorem, we see that $\mathcal{A}_p$ possess the following properties :

(1) $\mathcal{A}_p$ is closed under multiplication.

(2) $A' \in \mathcal{A}_{p'} \Longrightarrow \begin{pmatrix} I_1 & 0 & \\ & A' & \\ 0 & & I_2 \end{pmatrix} \in \mathcal{A}_p$,

where $\underline{p}'$ is the vector obtained from $\underline{p}$ by truncating some left and/or right end components, I_1, I_2 are identity matrices of appropiate dimensions.

(3) If $\underline{x}^a \prec_{\mathcal{P}^a} \underline{y}^a$, $\underline{x}^b \prec_{\mathcal{P}^b} \underline{y}^b$ on D are such that $\underline{x}^a = \underline{y}^a A^a$, $\underline{x}^b = \underline{y}^b A^b$ where $A^a \in \mathcal{A}_{p^a}$, $A^b \in \mathcal{A}_{p^b}$, then

$$(\underline{x}^a, \underline{x}^b) = (\underline{y}^a, \underline{y}^b) \begin{pmatrix} A^a & o \\ 0 & A^b \end{pmatrix} \quad \text{and} \quad \begin{pmatrix} A^a & 0 \\ 0 & A^b \end{pmatrix} \in \mathcal{A}_p \ .$$

(4) $\mathcal{A}_p$ is convex, as can be verified using properties (1) to (3) of Theorem 2.5 . However, the extreme points turn out to be quite complicated, unlike the case of doubly stochastic matrices where the extreme points are simply the permutation matrices.

Now we come to the characterization of $\mathcal{L}_p$ by matrices A of $\mathcal{A}_p$.

<u>Theorem 2.6</u>. If $\underline{x}, \underline{y} \in D$, then $\underline{x} \mathcal{L}_p \underline{y}$ on D iff there exists $A \in \mathcal{A}_p$ such that $\underline{x}=\underline{y}A$.

For the proof we need the following lemmas.

<u>Lemma 2.7</u>. Suppose $1 \leq r < n$, $s \geq t$, $\gamma > 0$, and let $a = \sum_{i=1}^{r} p_i$,

$b = \sum_{i=r+1}^{n} p_i$, $s' = s + \frac{\gamma}{a}$, $t' = t - \frac{\gamma}{b}$, $\xi = \frac{\gamma}{a}[(s-t)+\gamma(\frac{1}{a}+\frac{1}{b})]^{-1}$,

$\eta = \frac{\gamma}{b}[(s-t)+\gamma(\frac{1}{a}+\frac{1}{b})]^{-1}$, then

(1) $(s, \ldots, s, t, \ldots, t) = (s', \ldots, s', t', \ldots, t')\begin{pmatrix} A_{11} & A_{12} \\ A_{21} & A_{22} \end{pmatrix}$

and $\begin{pmatrix} A_{11} & A_{12} \\ A_{21} & A_{22} \end{pmatrix} \in \mathcal{A}_p$,

where A_{11} is $r \times r$, A_{12}, A_{21}, A_{22} have appropriate dimensions; each composed of identical columns, with entries given by:

ik_1^{th} entry of A_{11} is $p_i(1-\xi)/a$
ik_2^{th} entry of A_{12} is $p_i\eta/a$ — for $i=1,\ldots,r$,

jk_1^{th} entry of A_{21} is $p_j\xi/b$
jk_2^{th} entry of A_{22} is $p_j(1-\eta)/b$ — for $j=r+1,\ldots,n$;

(2) for $c = \sum_{i=1}^{n} p_i$, $(s+\frac{\gamma}{c},\ldots,s+\frac{\gamma}{c}) = (s+\frac{\gamma}{p_1}, s, \ldots, s)A$,

where A, composed of identical columns, with p_i/c as the

ik^{th} entry, satisfies $A \in \mathcal{A}_p$.

The proof of this lemma is just a direct algebraic verification and is omitted. The idea is that we can distribute some positive quantity from the smaller to components to the larger s components evenly, or wholly to the first component, by some sort of transformation involving the multiplication by an element of $\mathcal{A}_p$.

<u>Lemma 2.8</u>. If $\underline{x} \in D$, then for any $\delta>0$, there exists $A \in \mathcal{A}_p$ such that $\underline{x}=\underline{x}^{\delta}A$ for $\underline{x}^{\delta}=(x_1+\frac{\delta}{p_1},x_2,\ldots,x_{n-1},x_n-\frac{\delta}{p_n})$.

<u>Proof</u> : Partition $x_1,\ldots,x_{n-1}$ into k groups, each having equal components and such that components in different groups are different, i.e. $\underline{x}=(\underline{u}_1,\ldots,\underline{u}_k,x_n)$. Now distribute a positive quantity γ from x_n to the components of $\underline{u}_k$; then to the components of $\underline{u}_{k-1}$; and so on until it is distributed to the components of $\underline{u}_1$, by repeated applications of (1) of Lemma 2.7; where γ is chosen small enough so that it does not exceed δ and the new vectors formed all belong to D . Finally, apply (2) of Lemma 2.7 to distribute it wholly to the first component x_1. If $m\gamma<\delta\leq(m+1)\gamma$, perform the above procedure m times using γ , then repeat it again using $\delta-m\gamma$ in place of γ . Because of properties (1) and (2) of $\mathcal{A}_p$, the final matrix A we obtained belongs to $\mathcal{A}_p$.

Now we can proceed to prove Theorem 2.6. Sufficiency follows from definition of $\mathcal{A}_p$. For necessity, we prove by induction on

n. If any of the inequalities $\sum_{i=1}^{k} p_i x_i \leq \sum_{i=1}^{k} p_i y_i$, $k=1,\ldots,n-1$ is actually an equality, then the hypothesis of property (3) of $\mathcal{A}_p$ is true from the induction assumption and we are done. Otherwise take $\underline{x}_\delta^\delta$ as in Lemma 2.8 with

$$\delta = \min_{k=1,\ldots,n-1} \{ \sum_{i=1}^{k} p_i (y_i - x_i) \} > 0.$$

Then $\underline{x} \prec_p \underline{x}^\delta \prec_p \underline{y}$ on D. By the preceding argument $\underline{x}^\delta = \underline{y}A'$, $A' \in \mathcal{A}_p$. Also, from Lemma 2.8, $\underline{x} = \underline{x}^\delta A''$, $A'' \in \mathcal{A}_p$. Hence $\underline{x} = \underline{y}A$ where $A = A'A'' \in \mathcal{A}_p$.

D. <u>Functions preserving the p-majorizations</u>.

The following characterizes the functions that preserve the orderings $\prec_p$: For each $\underline{p}$, let $F(\cdot;\underline{p}) : D \longrightarrow R$ be such that $F(\underline{x};\underline{p}) \leq F(\underline{y};\underline{p})$ whenever $\underline{x} \prec_p \underline{y}$ on D. Then we extend $F(\cdot;\underline{p})$ to R^n by $F(\underline{x};\underline{p}) = F(\underline{x}^\pi;\underline{p}^\pi)$ for $\underline{x} \in D^\pi$. Consequently, we see that because $\underline{x} \prec_p \underline{y}$ on $D^\pi \Longrightarrow \underline{x}^\pi \prec_p^{\pi\pi} \underline{y}^\pi$ on $D \Longrightarrow F(\underline{x};\underline{p}) = F(\underline{x}^\pi;\underline{p}^\pi) \leq F(\underline{y}^\pi;\underline{p}^\pi) = F(\underline{y},\underline{p})$, so that F preserves the orderings $\prec_p$, independently of the permutation involved. If we denote by $\mathcal{F}_p$ the class of all such functions, $\mathcal{I}$ the class of all increasing functions and $\mathcal{D}$ the class of all decreasing functions, then we have the following:

<u>Theorem 2.9</u>. F preserves all $\prec_{wp}$ ($\overset{w}{\prec}_p$) iff $F \in \mathcal{F}_p \cap \mathcal{I}$ ($\mathcal{F}_p \cap \mathcal{D}$).

The proof is similar to that of Schur convex function.

In case of differentiable functions, we can characterize

functions of $\mathcal{F}_p$ by their derivatives.

<u>Theorem 2.10</u>. If F is differentiable, then $F \in \mathcal{F}_p$ iff F satisfied : (1) $F(\underline{x};\underline{p}) = F(\underline{x}^\pi;\underline{p}^\pi)$ for every permutation π , (2) $(x_i-x_j)(\frac{1}{p_i}\frac{\partial F}{\partial x_i} - \frac{1}{p_j}\frac{\partial F}{\partial x_j}) \geq 0$ for all $i,j=1,\ldots,n$.

<u>Proof</u> : First suppose $F \in \mathcal{F}_p$, then (1) follows from the definition. To prove (2), assume without loss of generality that $\underline{x} \in D$. For $\delta>0$ and $i<j$, let

$$\underline{x}^a=(x_1+\frac{\delta}{p_i}+\frac{\delta}{p_j},\ldots,x_i+\frac{\delta}{p_i}+\frac{\delta}{p_j},x_{i+1}+\frac{\delta}{p_j},\ldots,x_{j-1}+\frac{\delta}{p_j},x_j,\ldots,x_n) ,$$

$$\underline{x}^b=(x_1+\frac{\delta}{p_i}+\frac{\delta}{p_j},\ldots,x_{i-1}+\frac{\delta}{p_i}+\frac{\delta}{p_j},x_i+\frac{\delta}{p_j},\ldots,x_j+\frac{\delta}{p_j},x_{j+1},\ldots,x_n) ,$$

$$\underline{x}'=(x_1+\frac{\delta}{p_i}+\frac{\delta}{p_j},\ldots,x_{i-1}+\frac{\delta}{p_i}+\frac{\delta}{p_j},x_i+\frac{\delta}{p_j},\ldots,x_{j-1}+\frac{\delta}{p_j},x_j,\ldots,x_n) .$$

Taking $\frac{\partial F}{\partial x_i} = \lim_{\delta\to 0^+}\frac{F(\underline{x}^a;\underline{p})-F(\underline{x}';\underline{p})}{\delta/p_i}$, $\frac{\partial F}{\partial x_j}= \lim_{\delta\to 0^+}\frac{F(\underline{x}^b;\underline{p})-F(\underline{x}';\underline{p})}{\delta/p_j}$,

we have $$\frac{1}{p_i}\frac{\partial F}{\partial x_i} - \frac{1}{p_j}\frac{\partial F}{\partial x_j} = \lim_{\delta\to 0^+}\frac{F(\underline{x}^a;\underline{p})-F(\underline{x}^b;\underline{p})}{\delta} \geq 0 .$$

Next suppose (1) and (2) hold. We only need to prove that F preserves $\preccurlyeq$ on D . Because of Theorem 2.5 , we may take $\underline{x}=(x_1,x_2)$, $\underline{y}=(y_1,y_2)$, $\underline{x} \preccurlyeq \underline{y}$ on D and $\underline{x}\neq\underline{y}$, then we have $y_1>x_1\geq x_2>y_2$ and $p_1x_1+p_2x_2=p_1y_1+p_2y_2$. Therefore we have $p_1(y_1-x_1) = p_2(y_2-x_2) > 0$. For some θ , $0<\theta<1$,

$$F(\underline{x};\underline{p})-F(\underline{y};\underline{p}) = -\left\{(y_1-x_1)\frac{\partial F}{\partial x_1}\Big|_\theta + (y_2-x_2)\frac{\partial F}{\partial x_2}\Big|_\theta\right\},$$

where "$|_\theta$" indicates evaluation at the point $(1-\theta)\underline{x}+\theta\underline{y}$.

Hence $\quad F(\underline{x};\underline{p})-F(\underline{y};\underline{p}) = -p_1(y_1-x_1)\left\{\frac{\partial F}{\partial x_1}\Big|_\theta - \frac{\partial F}{\partial x_2}\Big|_\theta\right\} \leq 0.$

<u>Corollary 2.11</u>. If f is differentiable, then

$F(\underline{x};\underline{p})=\sum_{i=1}^{n} p_i f(x_i)$ belongs to $\mathfrak{F}p$ iff f is convex.

Just like Schur convex functions, $F \in \mathfrak{F}p$ restricted to the hyperplane $p_1x_1+\ldots+p_nx_n=K$ attains its minimum at the point $\underline{z}=(K/\sum_{i=1}^{n} p_i,\ldots,K/\sum_{i=1}^{n} p_i)$ where all components are equal. In order to see this, it is sufficient to show that any point $\underline{y} \in D^\pi$ lying on this hyperplane satisfies $\underline{z} \prec \underline{y}$ on D^π. Without loss of generality, we assume $\underline{y} \in D$. If for some k, $k=1,\ldots,n-1$,

$$p_1y_1+\ldots+p_ky_k < p_1\frac{K}{\sum_{i=1}^{n} p_i}+\ldots+p_k\frac{K}{\sum_{i=1}^{n} p_i},$$

then $\quad p_1y_k+\ldots+p_ky_k \leq p_1y_1+\ldots+p_ky_k < p_1\frac{K}{\sum_{i=1}^{n} p_i}+\ldots+p_k\frac{K}{\sum_{i=1}^{n} p_i}.$

So we have $\qquad y_n\leq\ldots\leq y_k<K/\sum_{i=1}^{n} p_i.$

Therefore $\qquad p_{k+1}y_{k+1}+\ldots+p_ny_n < p_{k+1}\frac{K}{\sum_{i=1}^{n} p_i}+\ldots+p_n\frac{K}{\sum_{i=1}^{n} p_i}.$

Summing the first and the last inequalities, we arrive at the contradiction $\sum_{i=1}^{n} p_i y_i < K$. Hence we conclude that $\underline{z} \not\prec_p \underline{y}$ on D . Therefore, the function F increases from $\underline{z}$ along any ray lying on the hyperplane $p_1x_1+\ldots+p_nx_n = K$. The class $\mathfrak{J}_p$ includes the class of Schur convex functions corresponding to the case $p_1 = \ldots = p_n$.

As an illustration, let us consider stratified sampling with r strata, and with $N_1,\ldots N_r$ as the sizes of the subpopulations. Suppose that the variance within each stratum is known (say, estimated from past experience) as $S_1^2,\ldots,S_r^2$. If we draw samples $n_1,\ldots,n_r$ from the corresponding strata with sample means $\bar{y}_1,\ldots,\bar{y}_r$, then the estimate of the population mean is $\bar{y} = \sum_{i=1}^{r} N_i\bar{y}_i/N$, where $N = \sum_{i=1}^{r} N_i$; with variance

$$V_{\underline{n}} = (1/N^2)\sum_{i=1}^{r} N_i(N_i-n_i)S_i^2/n_i = (1/N^2)\sum_{i=1}^{r} N_i^2S_i^2/n_i - \text{constant} .$$

For fixed sampling cost $c_1n_1+\ldots+c_rn_r = K$, the objective is to choose sampling units $\underline{n}$ so as to minimize the variance, or equivalently the sum $\sum_{i=1}^{r} N_i^2S_i^2/n_i$: With $p_i = N_iS_i\sqrt{c_i}$, $x_i = n_i\sqrt{c_i}/N_iS_i$ and $f(x) = 1/x$, we have $N_i^2S_i^2/n_i = p_if(x_i)$. Since f is convex, the minimum of $\sum_{i=1}^{r} p_if(x_i)$ is attained

where $x_1 = \ldots = x_r$, i.e. $n_1\sqrt{c_1}/N_1S_1 = \ldots = n_r\sqrt{c_r}/N_rS_r$,

agreeing with the result obtained using the Lagrange multiplier technique.

One advantage of using this approach is that we have partial orderings defined on $\underline{x}$ such that $\underline{x} \lessdot \underline{x}'$ on D^{π} $\Longrightarrow V(\underline{x}) \leq V(\underline{x}')$, where $V(\underline{x})$ is the variance corresponding to the sampling units $\underline{n}$ associated with $\underline{x}$. Furthermore, because f is decreasing, we can strengthen the above result by $\underline{x} \overset{W}{\lessdot} \underline{x}'$ on $D^{\pi} \Longrightarrow V(\underline{x}) \leq V(\underline{x}')$.

Chapter 3

A. Introduction.

In this chapter, we extend the majorizations to their stochastic versions and discuss the relations among them. We deal with the preservation theorem of Schur convexity by Proschan and Sethuraman (1977) and of monotonicity, obtaining a class of inequalities that arise in some multivariate distributions. Then we apply the preservation of Schur convexity to construct a class of test functions in one type of hypothesis testing.

B. Stochastic versions of majorizations.

The following notions of stochastic ordering between real random variables are used repeatedly,

Definition 3.1. (1) $\underline{X}$ is stochastically less than $\underline{Y}$, denoted by $\underline{X}$ st$\leq$ $\underline{Y}$, iff for every real t, $Pr\{X>t\}\leq Pr\{Y>t\}$, i.e. $F_{\underline{Y}}(t)\leq F_{\underline{X}}(t)$.

(2) $X \overset{d}{=} Y$ iff X st$\leq Y$ and Y st$\leq X$.

The following is well known.

Proposition 3.2. If X st$\leq Y$, then $EX\leq EY$, provided the expectations are finite.

Proof :
$$EX= \int_{-\infty}^{0} -Pr\{X\leq t\}d\mu(t) + \int_{0}^{\infty} Pr\{X>t\}d\mu(t)$$

and $$EY = \int_{-\infty}^{0} -\Pr\{Y \le t\}d\mu(t) + \int_{0}^{\infty} \Pr\{Y > t\}d\mu(t) ,$$

where μ stands for the Lebesque measure or the counting measure. The inequality $EX \le EY$ now follows from the inequalities of the corresponding integrands.

In the case of degenerate random varibles, the stochastic inequality reduces to ordinary inequality.

Similarly, various stochastic extensions of majorization are possible which, for degenerate random vectors, reduce to ordinary majorization. These extensions, proposed by various authors, are obtained from the four equivalent conditions of Theorem 1.16 by changing the constants to random variables and replacing the inequalities either by those of expectations, or by stochastic inequalities. However, these versions turn out to be quite different except under certain imposed conditions, some or all of them become equivalent. These versions are:

M_1: $\underline{X} \prec \underline{Y}$ with probability 1 .

M_1': $F(\underline{X}) \le F(\underline{Y})$ with probability 1, for every Schur convex F.

M_2: $EF(\underline{X}) \le EF(\underline{Y})$ for every Schur convex F .

M_2': $F(\underline{X})$ st$\le F(\underline{Y})$ for every Schur convex F .

M_3: $EF(\underline{X}) \le EF(\underline{Y})$ for every symmetric convex F .

M_4: $E\{f(X_1)+\ldots+f(X_n)\} \le E\{f(Y_1)+\ldots+f(Y_n)\}$ for every convex f.

M_5: $\sum_{i=1}^{k} X_{[i]}$ st$\le \sum_{i=1}^{k} Y_{[i]}$ for $k=1,\ldots,n-1$ and $\sum_{i=1}^{n} X_{[i]} \overset{d}{=} \sum_{i=1}^{n} Y_{[i]}$.

where $X_{[1]} \ge \ldots \ge X_{[n]}$ is the reversed order statistic.

M_5^*: $\sum_{i=k}^{n} Y_{[i]} \text{ st}\leq \sum_{i=k}^{n} X_{[i]}$ for $k=1,\ldots,n-1$ and $\sum_{i=1}^{n} Y_{[i]} \overset{d}{=} \sum_{i=1}^{n} X_{[i]}$.

M_6: $(EX_{[1]},\ldots,EX_{[n]}) \prec (EY_{[1]},\ldots,EY_{[n]})$.

We can also extend the two weak majorizations $\underset{w}{\prec}$ and $\overset{w}{\prec}$ similarly with slight modification. For st$\underset{w}{\prec}$ (st$\overset{w}{\prec}$), we replace $\prec$ by $\underset{w}{\prec}$ ($\overset{w}{\prec}$) and functions by increasing (decreasing) functions in the above versions except for version 5. We shall denote the resulting versions obtained by $P_1,\ldots,P_6$ ($Q_1,\ldots,Q_6$). For version 5, we have:

P_5: $\sum_{i=1}^{k} X_{[i]} \text{ st}\leq \sum_{i=1}^{k} Y_{[i]}$ for $k=1,\ldots,n$.

Q_5: $\sum_{i=k}^{n} Y_{[i]} \text{ st}\leq \sum_{i=k}^{n} X_{[i]}$ for $k=1,\ldots,n$.

The following two results concerning the relations among these versions are due to Marshall and Olkin (to appear).

<u>Proposition 3.3</u>. M_1 and M_1' are equivalent, so are M_2 and M_2' .

<u>Proof</u> : The equivalence of M_1 and M_1' follows from that of $\underline{x} \prec \underline{y}$ and $F(\underline{x})\leq F(\underline{y})$ for all Schur convex functions F .
$M_2' \Longrightarrow M_2$ follows from Proposition 3.2.
Now suppose M_2 holds. For arbitrary t and Schur convex function F , let $S=\{\underline{z}:F(\underline{z})>t\}$. Then 1_S is Schur convex.
Therefore $\Pr\{F(\underline{X})>t\} = E1_S(\underline{X})\leq E1_S(\underline{Y}) = \Pr\{F(\underline{Y})>t\}$.
Hence $M_2 \Longrightarrow M_2'$.

Theorem 3.4.

$$\begin{tikzcd}
 & & M_3 \arrow[r, Rightarrow] \arrow[dr, Rightarrow] & M_4 \\
M_1 \arrow[r, Rightarrow] & M_2 \arrow[ur, Rightarrow] \arrow[r, Rightarrow] \arrow[dr, Rightarrow] & M_5 \arrow[r, Rightarrow] & M_6 \\
 & & M_5^* \arrow[ur, Rightarrow] &
\end{tikzcd}$$

No further implication is possible.

For the proof and the counterexamples, see Marshall and Olkin (to appear).

Proposition 3.5. If $X_1,\ldots,X_n$ are i.i.d., $Y_1,\ldots,Y_n$ are i.i.d. and $X_1+\ldots+X_n \overset{d}{=} Y_1+\ldots+Y_n$, then $X_i \overset{d}{=} Y_i$. In particular, if $X_1,\ldots,X_n$ are i.i.d., $Y_1,\ldots,Y_n$ are i.i.d., then M_1,M_2,M_5 and M_5^* are equivalent.

Proof : Denote the distributions of X_i,Y_i by Φ_X and Φ_Y.

$$X_1+\ldots+X_n \overset{d}{=} Y_1+\ldots+Y_n \Longrightarrow [\hat{\Phi}_X]^n = [\hat{\Phi}_Y]^n \Longrightarrow \hat{\Phi}_X = \hat{\Phi}_Y \Longrightarrow \Phi_X \overset{d}{=} \Phi_Y ,$$

where $\hat{\Phi}$ is the Fourier transform of Φ.

The equivalence of M_1,M_2,M_5,M_5^* results from :

$$M_1 \Longrightarrow M_2 \Longrightarrow M_5 \text{ or } M_5^* \Longrightarrow X_1+\ldots+X_n \overset{d}{=} Y_1+\ldots+Y_n .$$

It may appear that M_5 and M_5^* are closely related, so their equivalence requiring such strong conditions is a bit surprising.

For stochastic weak majorizations, similar results can be obtained. The proofs are easy extensions of those of the three previous results.

<u>Proposition 3.6</u>. $P_1(Q_1)$ and $P_1'(Q_1')$ are equivalent, so are $P_2(Q_2)$ and $P_2'(Q_2')$.

<u>Theorem 3.7</u>.

$$\begin{tikzcd} & & P_3 \arrow[r, Rightarrow] \arrow[dr, Rightarrow] & P_4 \\ P_1 \arrow[r, Rightarrow] & P_2 \arrow[ur, Rightarrow] \arrow[dr, Rightarrow] & & \\ & & P_5 \arrow[r, Rightarrow] & P_6 \end{tikzcd}$$

No further implication is possible. The same is true if we replace the P's by the Q's.

<u>Proposition 3.8</u>. If $X_1,\ldots,X_n$ are i.i.d., $Y_1,\ldots,Y_n$ are i.i.d., and $X_1+\ldots+X_n \ \mathrm{st}\leq Y_1+\ldots+Y_n$, then $X_1 \ \mathrm{st}\leq Y_1$. In particular, if $X_1,\ldots,X_n$ are i.i.d., $Y_1,\ldots,Y_n$ are i.i.d., then P_1,P_2 and P_5 (Q_1,Q_2 and Q_5) are equivalent.

Because version 1 is too strong to be satisfied in most situations of interest, we shall use version 2 as our definition for stochastic majorization and weak majorizations (as in Nevius, Proschan and Sethuraman 1977). From now on, version 2 will be assumed if the version is not explicitly specified.

Version 3 has the following illustration:

<u>Proposition 3.9</u>. If $X_1,\ldots,X_n$ are exchangeable random variables and $\underline{u} \prec \underline{v}$, then $(u_1X_1,\ldots,u_nX_n) \ \mathrm{st}\prec (v_1X_1,\ldots,v_nX_n)$ in the sense of version 3. The same is true if we replace $\prec$ by $\underset{w}{\prec}$ ($\overset{w}{\prec}$) and $\mathrm{st}\prec$ by $\mathrm{st}\underset{w}{\prec}$ ($\mathrm{st}\overset{w}{\prec}$).

<u>Proof</u> : It suffices to prove the case n=2 because of Theorem 1.8. In this case, $(u_1,u_2) \prec (v_1,v_2) \Longrightarrow (u_1,u_2) = \alpha(v_1,v_2)+\bar{\alpha}(v_2,v_1)$

for some $\alpha,\bar{\alpha}>0$, $\alpha+\bar{\alpha}=1$. For any symmetric convex function F,

$$\begin{aligned} EF(u_1X_1,u_2X_2) &= EF[(\alpha v_1+\bar{\alpha}v_2)X_1,(\bar{\alpha}v_1+\alpha v_2)X_2] \\ &= EF[\alpha(v_1X_1,v_2X_2)+\bar{\alpha}(v_2X_1,v_1X_2)] \\ &\leq \alpha EF(v_1X_1,v_2X_2)+\bar{\alpha}EF(v_2X_1,v_1X_2) = EF(v_1X_1,v_2X_2). \end{aligned}$$

The proof for the P,Q versions are similar, except with the addition of monotonicity.

The M_3 version of the above proposition is proved by Marshall and Proschan (1965), who also illustrate by an example that further extension to M_2 is not possible. The P_3 version is proved by Chong (1976).

Intuitively, $\underline{X}$ st$\leq$ $\underline{Y}$ means that $\underline{Y}$ tends to majorize $\underline{X}$. As an illustration, consider a binomial random variable X with parameters N, p where $p \geq 1/2$, then X tends to take the integral values in [N/2,N] rather than in [0,N/2]; or equivalently, X tends to be greater than N-X. Consequently, (X+c,N-X) tends to majorize (X,N-X+c) for any $c\geq 0$. In fact, for any Schur convex function F,

$$EF(X+c,N-X) = \sum_{x=0}^{N} F(x+c,N-x)Pr\{X=x\} = \sum_{x=0}^{N} F(N-x+c,x)Pr\{X=N-x\}$$

so $EF(X+c,N-X)$

$$= \frac{1}{2}\sum_{x=0}^{N}[F(x+c,N-x)Pr\{X=x\}+F(N-x+c,x)Pr\{X=N-x\}] \quad . \qquad (1)$$

Similarly, $EF(X,N-X+c)$

$$= \frac{1}{2}\sum_{x=0}^{N}[F(x,N-x+c)Pr\{X=x\}+F(N-x,x+c)Pr\{X=N-x\}] \quad . \qquad (2)$$

But $F(x+c,N-x) = F(N-x,x+c) \overset{<}{\underset{>}{=}} F(N-x+c,x) = F(x,N-x+c)$

$$\Longleftrightarrow x \overset{<}{\underset{>}{=}} N/2 \Longleftrightarrow \Pr\{X=x\} \overset{<}{\underset{>}{=}} \Pr\{X=N-x\} ,$$

so that each summand in (1) is greater than the corresponding one in (2). Hence $(X,N-X-c)\,st\prec (X+c,N-X)$, which also means that $F(X,N-X+c)\ st\leq F(X+c,N-x)$ for any Schur convex function F. In particular, if f is convex, then $F(x_1,x_2)=f(x_1)+f(x_2)$ is Schur convex, and

$$f(X)+f(N-X+c)\ st\leq\ f(X+c)+f(N-X) ,$$

which is a lemma proved by Cohen and Sackrowitz (1975).

C. <u>Preservation of Schur convexity</u>.

Now we come to the preservation theorem of Schur functions. First we introduce two definitions.

<u>Definition 3.10</u>. For $\Lambda\subset R$, $\psi(\lambda,x):\Lambda\times R\longrightarrow R$ is said to be totally positive of order 2 (TP_2) if

(1) $\psi(\lambda,x) \geq 0$,

(2) for $\lambda_1,\lambda_2\varepsilon\Lambda,\lambda_1\leq\lambda_2$ and $0\leq x_1\leq x_2$,

$$\psi(\lambda_1,x_1)\psi(\lambda_2,x_2) \geq \psi(\lambda_1,x_2)\psi(\lambda_2,x_1) .$$

<u>Definition 3.11</u>. If $\Lambda\subset R$ is such that $\lambda_1,\lambda_2\varepsilon\Lambda,\lambda_1>\lambda_2 \Longrightarrow \lambda_1\pm\lambda_2\varepsilon\Lambda$, then $\psi(\lambda,x):\Lambda\times R\longrightarrow R$ is said to satisfy the semi-group property ($\psi\psi$ is SGP) if

(1) $\psi(\lambda,x)=0$ for all $x\leq 0$,

(2) for any $\lambda_1,\lambda_2\varepsilon\Lambda,\psi(\lambda_1+\lambda_2,x) = \int\psi(\lambda_1,x-y)\psi(\lambda_2,y)d\mu(y)$,

where μ is Lebesque measure or counting measure.

From now on, we shall omit the dummy index in the summation $\sum$ and the product $\prod$ whenever it is clear from the context. The following is the preservation theorem of Proschan and Sethuraman (1977).

<u>Theorem 3.12</u>. If $\psi(\lambda,x):\Lambda\times R \rightarrow R$ is TP_2 and SGP, B is a non-negative function of two variables, then for any Schur convex function F , $H(\underline{\lambda})$ defined by

$$H(\underline{\lambda}) = \int\cdots\int F(\underline{x})B(\sum x_i,\sum\lambda_i)\prod\psi(\lambda_i,x_i)d\mu(x_1)\ldots d\mu(x_n)$$

is Schur convex in $\underline{\lambda}$, where the integral is assumed to exist.

<u>Proof</u> : It suffices to prove the case n=2 because of Theorem 1.8.
If $(\lambda_1,\lambda_2)\prec(\lambda_1', \lambda_2')$, then $\lambda_1+\lambda_2=\lambda_1'+\lambda_2'$.
Hence $G(x_1,x_2) \equiv F(x_1,x_2)B(x_1+x_2,\lambda_1+\lambda_2) = F(x_1,x_2)B(x_1+x_2,\lambda_1'+\lambda_2')$ is Schur convex in (x_1,x_2)
Because G is symmetric, it is easy to verify that H is also symmetric. Therefore we may assume that $\lambda_1'>\lambda_1\geq\lambda_2>\lambda_2'$.
Now $H(\lambda_1', \lambda_2')-H(\lambda_1,\lambda_2)$

$$= \iint G(x_1,x_2)[\psi(\lambda_1',x_1)\psi(\lambda_2',x_2)-\psi(\lambda_1,x_1)\psi(\lambda_2,x_2)]d\mu(x_1)d\mu(x_2)$$

$$= \iiint G(x_1,x_2)[\psi(\lambda_1,x_1-y)\psi(\lambda_2',x_2)-\psi(\lambda_1,x_1)\psi(\lambda_2,x_2-y)]$$
$$\times\psi(\lambda_1'-\lambda_1,y)d\mu(x_1)d\mu(x_2)d\mu(y)$$

$$= \int\psi(\lambda_1'-\lambda_1,y)\iint_{x_1>x_2}[\psi(\lambda_1,x_1)\psi(\lambda_2',x_2)-\psi(\lambda_1,x_2)\psi(\lambda_2',x_1)]$$
$$\times[G(x_1+y,x_2)-G(x_1,x_2+y)]d\mu(x_1)d\mu(x_2)$$

≥ 0 ,

where the second equality follows from being SGP and $\lambda_1'-\lambda_1=\lambda_2-\lambda_2'$; the third equality follows from cancelling the part $x_1=x_2$, changing variables and G being symmetric; the last inequality follows from being TP_2, G Schur convex and $x_1>x_2$, resulting in the integrand being non-negative.

Because F is Schur convex iff $-F$ is Schur concave, Theorem 3.12 is also true if we replace Schur convexity by Schur concavity. Furthermore, from Mudholkar's result (1966), $F(x_1,x_2)=f(x_1)f(x_2)$ is Schur convex (concave) iff f is log-convex (log-concave). So we have the following corollary:

<u>Corollary 3.13</u>. (Proschan and Sethuraman 1974). If $\psi(\lambda,x)$ is TP_2 and SGP and f is log-convex (log-concave), then $h(\lambda)$ defined by $h(\lambda) = \int\psi(\lambda,x)g(x)d\mu(x)$ is log-convex(log-concave).

We give below some commonly encountered functions that are well known to be TP_2 and SGP. Verification will be omitted.

<u>Proposition 3.14</u>. The following $\psi(\lambda,x)$ are TP_2 and SGP:

(1) $\psi(\lambda,x) = \frac{\lambda^x}{x!}$ for $x=0,1,2,\ldots,$ $\lambda>0,$

(2) $\psi(\lambda,x) = \left\{{\lambda \atop x}\right\}$ for $x=0,1,2,\ldots,$ $\lambda=1,2,\ldots,$

here we adopt the convention that $\binom{\lambda}{x}=0$ if $x>\lambda$,

(3) $\psi(\lambda,x) = \frac{x^{\lambda-1}}{\Gamma(\lambda)}$ for $x>0,$ $\lambda>0,$

(4) $\psi(\lambda,x) = \frac{\Gamma(\lambda+x)}{x!\Gamma(\lambda)}$ for $x=0,1,2,\ldots,\quad \lambda>0,$

with "$\psi(\lambda,x)=0$ otherwise" being assumed.

In case that $\phi(\underline{x};\underline{\lambda})=cB(\sum x_i,\sum\lambda_i)\prod\psi(\lambda_i,x_i)$ is the density of a random vector $\underline{X}_{\underline{\lambda}}$, $H(\underline{\lambda})$ is the expectation $E_{\underline{\lambda}}(F(\underline{X}))$. Theorem 3.12 can be written in the form $\underline{\lambda} \prec \underline{\lambda}' \Longrightarrow \underline{X}_{\underline{\lambda}}$ st$\prec$ $\underline{X}_{\underline{\lambda}'}$. For convenience, we introduce the following:

<u>Definition 3.15</u>. A family of n-dimensional distribution functions $\{\Phi_{\underline{\lambda}};\underline{\lambda}\varepsilon\Lambda^n\}$ (or random vectors $\{\underline{X}_{\underline{\lambda}}:\underline{\lambda}\varepsilon\Lambda^n\}$) is said to be a Schur convex family (in $\underline{\lambda}$) if $H(\underline{\lambda})\equiv E_{\underline{\lambda}}(F(\underline{X}))$ is Schur convex in $\underline{\lambda}$ whenever F is Schur convex.

<u>Proposition 3.16</u>. $\{\Phi_{\underline{\lambda}}\}$ is a Schur convex family if $\underline{X}_{\underline{\lambda}}=(X_{\lambda_1},\ldots,X_{\lambda_n})$ is composed of independent components of the same family, and X_{λ_i} $i=1,\ldots,n$ has density from either of the following families:

(a) Poisson (Rinott 1973):

$$\Phi(x;\lambda) = \frac{\lambda^x}{x!}e^{-\lambda} \text{ for } x=0,1,2,\ldots,\quad \lambda>0 .$$

(b) Binomial (Nevius, Proschan and Sethuraman 1977):

$$\Phi(x;\lambda) = \binom{\lambda}{x}\theta^x(1-\theta)^{\lambda-x} \text{ for } x=0,1,2,\ldots,\ \lambda=1,2,\ldots,$$

and fixed $\theta>0$.

(c) Gamma (Nevius, Proschan and Sethuraman 1977):

$$\phi(x;\lambda) = \frac{\theta^{\lambda} x^{\lambda-1}}{\Gamma(\lambda)} e^{-\theta x} \quad \text{for } x>0,\ \lambda>0 \text{ and fixed } \theta>0.$$

<u>Proof</u> : This is because the joint density $\prod\phi(x_i,\lambda_i)$ is $cB(\sum x_i,\sum\lambda_i)\prod\psi(\lambda_i,x_i)$ with ψ in the form shown in Proposition 3.14.

<u>Proposition 3.17</u>. $\{\Phi_{\underline{\lambda}}\}$ is a Schur convex family if $\underline{X}_{\underline{\lambda}}$ has density in the following form:

(1) Multinomial (Rinott 1973):

$$\phi(\underline{x};\underline{\lambda}) = N!\prod\frac{\lambda_i^{x_i}}{x_i!}$$

for $x_i=0,1,\ldots,N$, $\sum x_i=N$, $\lambda_i>0$, $\sum\lambda_i=1$.

(2) Dirichlet (Nevius, Proschan and Sethuraman 1977):

$$\phi(\underline{x};\underline{\lambda}) = \frac{\Gamma(\beta+\sum\lambda_i)}{\Gamma(\beta)}(1-\sum x_i)^{\beta-1}\prod\frac{x_i^{\lambda_i-1}}{\Gamma(\lambda_i)}$$

for $x_i\geq 0$, $\sum x_i\leq 1$, $\lambda_i>0$, and fixed $\beta>0$.

(3) Inverted Dirichlet (Hollander, Proschan and Sethuraman, to appear):

$$\phi(\underline{x};\underline{\lambda}) = \frac{\Gamma(\theta+\sum\lambda_i)}{\Gamma(\theta)(1+\sum x_i)^{\theta+\sum\lambda_i}}\prod\frac{x_i^{\lambda_i-1}}{\Gamma(\lambda_i)}$$

for $x_i>0$, $\lambda_i>0$, and fixed $\theta>0$.

(4) Negative multinomial (Marshall and Olkin, to appear);

$$\phi(\underline{x};\underline{\lambda}) = \frac{\Gamma(k+\sum x_i)}{\Gamma(k)}(1-\sum\lambda_i)^{k}\prod\frac{\lambda_i^{x_i}}{x_i!}$$

for $x_i=0,1,2,\ldots$, $\lambda_i>0$, $\sum\lambda_i<1$ and fixed $k>0$.

(5) Multivariate negative binomial (Nevius, Proschan and Sethuraman 1977):

$$\phi(\underline{x};\underline{\lambda}) = \frac{\Gamma(N+\sum x_i)}{\Gamma(N)}(1+\sum\lambda_i)^{-N-\sum x_i}\prod\frac{\lambda_i^{x_i}}{x_i!}$$

for $x_i=0,1,2,\ldots$, $\lambda_i>0$; and fixed $N>0$.

(6) Dirichlet compound negative multinomial (Hollander, Proschan and Sethuraman, to appear):

$$\phi(\underline{x};\underline{\lambda}) = \frac{\Gamma(N+\sum x_i)\Gamma(\theta+\sum\lambda_i)\Gamma(N+\theta)}{\Gamma(N)\Gamma(\theta)\Gamma(N+\theta+\sum\lambda_i+\sum x_i)}\prod\frac{\Gamma(x_i+\lambda_i)}{x_i!\Gamma(\lambda_i)}$$

for $x_i=0,1,2,\ldots$, $\lambda_i>0$; and fixed $N>0$, $\theta>0$.

(7) Multivariate hypergeometric (Nevius, Proschan and Sethuraman 1977):

$$\phi(\underline{x};\underline{\lambda}) = \binom{M}{N}^{-1}\prod\binom{\lambda_i}{x_i}$$

for $x_i=0,1,\ldots$, $\sum x_i=N$, $\lambda_i=1,2,\ldots$, $\sum\lambda_i=M$; and fixed positive integers M, N with $M\geq N$.

(8) Multivariate inverse hypergeometric:

$$\phi(\underline{x};\underline{\lambda}) = \binom{M}{k+\sum x_i-1}^{-1}\binom{M-\sum\lambda_i}{k-1}\frac{M-\sum\lambda_i-k+1}{M-\sum x_i-k+1}\prod\binom{\lambda_i}{x_i}$$

for $x_i=0,1,2,\ldots,\lambda_i=1,2,\ldots,\sum\lambda_i\leq M-k$; and fixed positive integers M,k with $k<M$.

(9) Negative multivariate hypergeometric (Nevius, Proschan and Sethuraman 1977):

$$\phi(\underline{x},\underline{\lambda}) = \frac{N!\Gamma(M)}{\Gamma(N+M)}\prod\frac{\Gamma(x_i+\lambda_i)}{x_i!\Gamma(\lambda_i)}$$

for $x_i=0,1,2,\ldots$, $\sum x_i=N$, $\lambda_i>0$, $\sum\lambda_i=M$; and fixed $M>0$, and positive integer N.

The proof follows directly from Theorem 3.12 and Proposition 3.14.

The distributions (1), (7) and (9) are (n-1)-dimensional because $\underline{x}$ is confined on the hyperplane $x_1+\ldots+x_n=N$.

<u>Proposition 3.18</u>. (1) If $\{\Phi_{\underline{\lambda}}\}$ is a Schur convex family, then $\Pr\{r_1<(\text{or}\leq)X_i<(\text{or}\leq)r_2 \text{ for all } i|\underline{\lambda}\}$ is a Schur concave function of $\underline{\lambda}$, where $r_1\geq-\infty$, $r_2\leq\infty$.

(2) In addition, if $\underline{X}_{\underline{\lambda}}$ is non-negative and $Z_{\underline{\lambda}}$ is the number of zero components of $\underline{X}_{\underline{\lambda}}$, then $\underline{\lambda}\prec\underline{\lambda}' \Rightarrow Z_{\underline{\lambda}} \text{ st} \leq Z_{\underline{\lambda}'}$.

<u>Proof</u> : (1) It is easy to verify that $F(\underline{x})$ defined by

$$F(\underline{x}) = \begin{cases} 1 & \text{if } \min_i\{x_i\}>(\text{or}\geq)r_1 \text{ and } \max\{x_i\}<(\text{or}\leq)r_2 \\ 0 & \text{otherwise} \end{cases}$$

is Schur concave. Therefore

$$E_{\underline{\lambda}}F(\underline{X}) = \Pr\{r_1<(\text{or}\leq)X_i<(\text{or}\leq)r_2 \text{for all } i|\underline{\lambda}\}$$

is Schur concave in $\underline{\lambda}$.

(2) This follows from considering the indicator functions of

$$\{\underline{x}\geq 0: \sum \mathbb{1}_{\{x_i=0\}}\leq k\} \quad k>0,$$

which are Schur convex on R_+^n.

The multinomial case of (1) with $r_2=\infty$ corresponds to the result of Olkin (1972) and with $r_1=-\infty$ corresponds to that of Rinott (1973). The multinomial case of (2) corresponds to the result of Wong and Yue (1973).

D. Preservation of monotonicity.

For the extension to stochastic weak majorization, we want

$$\underline{\lambda} \underset{w}{\prec} (\overset{w}{\prec}) \underline{\lambda}' \Longrightarrow \underline{X}_{\underline{\lambda}} \text{ st}\underset{w}{\prec} (\text{st}\overset{w}{\prec}) \underline{X}_{\underline{\lambda}'} \Longleftrightarrow E_{\underline{\lambda}}F(\underline{X}) \leq E_{\underline{\lambda}'}F(\underline{X})$$

for every Schur convex increasing (decreasing) F. Thus, in addition to the preservation of Schur convexity, we need a preservation of monotonicity as well.

Theorem 3.19. Let $\phi(\underline{x};\underline{\lambda})$ be a density of the form

$$\phi(\underline{x};\underline{\lambda}) = cB(\sum x_i, \sum \lambda_i)\prod\psi(\lambda_i, x_i) ,$$

where ψ is SGP and B satisfies:

$$B(x,\lambda) = \int B(x+y,\lambda+\alpha)\psi(\alpha,y)d\mu(y) ,$$

whenever $\alpha,\lambda\varepsilon\Lambda$, $\alpha>0$.

Then the transformation $F \longrightarrow H$ defined by

$$H(\underline{\lambda}) = \int\cdots\int F(\underline{x})\phi(\underline{x};\underline{\lambda})d\mu(x_1)\ldots d\mu(x_n)$$

preserves monotonicity, i.e. if F is increasing or decreasing, then so is H .

Proof : Suppose F is increasing. We only need to show that H is increasing in its kth component for $k=1,\ldots,n$.

For $\alpha,\lambda_i\varepsilon\Lambda$ and $\alpha>0$,

$$H(\lambda_1,\ldots,\lambda_k+\alpha,\ldots,\lambda_n)$$

$$=c\int\cdots\int F(\underline{x})B(\sum x_i,\sum\lambda_i+\alpha)\prod_{i\neq k}\psi(\lambda_i,x_i)\psi(\lambda_k+\alpha,x_k)\prod d\mu(x_i)$$

$$=c\int\cdots\iint F(\underline{x})B(\sum x_i,\sum\lambda_i+\alpha)\prod_{i\neq k}\psi(\lambda_i,x_i)\psi(\lambda_k,x_k-y)\psi(\alpha,y)\prod d\mu(x_i)d\mu(y)$$

$$=c\int\cdots\iint F(x_1,\ldots,x_k+y,\ldots,x_n)B(\sum x_i+y,\sum\lambda_i+\alpha)\prod\psi(\lambda,x_i)\psi(\alpha,y)$$

$$\times\prod d\mu(x_i)d\mu(y)$$

$$\geq c\int\cdots\iint F(\underline{x})B(\sum x_i+y,\sum\lambda_i+\alpha)\prod\psi(\lambda_i,x_i)\psi(\alpha,y)\prod d\mu(x_i)d\mu(y)$$

$$=c\int\cdots\int F(\underline{x})B(\sum x_i,\sum\lambda_i)\prod\psi(\lambda_i,x_i)\prod d\mu(x_i) = H(\underline{\lambda}) ,$$

where the second equality follows from the SGP; the third equality follows from a change of variable; the next inequality follows from F being increasing and $y\geq 0$ (integrand is 0 for $y<0$); and the next equality follows from the hypothesis on B .

The condition to be satisfied by B is simply that the marginal density is in a similar family, only with a lower dimension;

i.e. if $\quad \phi_{X_1\ldots X_n}(x_1,\ldots,x_n) = cB(\sum_{i=1}^{n} x_i,\sum_{i=1}^{n}\lambda_i)\prod_{i=1}^{n}\psi(\lambda_i,x_i)$,

then $\quad \phi_{X_1\ldots X_{n-1}}(x_1,\ldots,x_{n-1}) = cB(\sum_{i=1}^{n-1} x_i,\sum_{i=1}^{n-1}\lambda_i)\prod_{i=1}^{n-1}\psi(\lambda_i,x_i)$.

In particular, if the multivariate density is a product of univariate densities, each being SGP, then the multivariate density is of the form $\prod\psi(\lambda_i,x_i)$, so it preserves monotonicity. Furthermore, if $\phi(\underline{x};\underline{\lambda})$ satisfies the hypothesis of the above

theorem, so will its marginal densities.

Proposition 3.20. The Schur convex families of independent Poisson, Binomial and Gamma of Proposition 3.16 preserve monotonicity.

The proof follows from the following obvious lemma:

Lemma 3.21. If $\psi(\lambda,x)$ is TP_2(or SGP), then so are $a^x\psi(\lambda,x)$ and $b^\lambda\psi(\lambda,x)$ where $a,b>0$.

Now we give some examples of multivariate distributions that satisfy the hypothesis of Theorem 3.19. Some appear slightly different from their previous forms shown in Proposition 3.17 because we require the distributions to be n-dimensional.

(1) Multinomial with fixed N:

$$\phi(\underline{x};\underline{\lambda}) = N!\frac{(1-\sum\lambda_i)^{N-\sum x_i}}{(N-\sum x_i)!}\prod\frac{\lambda_i^{x_i}}{x_i!}$$

for $x_i=0,1,2,\ldots,\ \sum x_i\le N,\ \lambda_i>0,\sum\lambda_i<1.$

Here $\quad B(x,\lambda) = \dfrac{(1-\lambda)^{N-x}}{(N-x)!}\ ,\quad \psi(\lambda,x) = \dfrac{\lambda^x}{x!}\ .$

(2) Multivariate hypergeometric with fixed M,N:

$$\phi(\underline{x};\underline{\lambda}) = \binom{M}{N}^{-1}\binom{M-\sum\lambda_i}{N-\sum x_i}\prod\binom{\lambda_i}{x_i}$$

for $x_i=0,1,2,\ldots,\ \sum x_i\le N;\ \lambda_i=1,2,\ldots,\ \sum\lambda_i<M$.

Here $B(x,\lambda) = \binom{M-\lambda}{N-x}$, $\psi(\lambda,x) = \binom{\lambda}{x}$.

(3) Negative multivariate hypergeometric with fixed M,N:

$$\phi(\underline{x};\underline{\lambda}) = \frac{N!\Gamma(M)}{\Gamma(N+M)} \frac{\Gamma(M+N-\sum x_i-\sum\lambda_i)}{(N-\sum x_i)!\Gamma(M-\lambda_i)} \prod \frac{\Gamma(x_i+\lambda_i)}{x_i!\Gamma(\lambda_i)}$$

for $x_i=0,1,2,\ldots,\ \sum x_i \leq N,\ \lambda_i>0,\ \sum\lambda_i<M.$

Here $B(x,\lambda) = \dfrac{\Gamma(M+N-x-\lambda)}{(N-x)!\Gamma(M-\lambda)}$, $\psi(\lambda,x) = \dfrac{\Gamma(x+\lambda)}{x!\Gamma(\lambda)}$.

(4) Multivariate negative binomial with fixed N :

$$\phi(\underline{x};\underline{\lambda}) = \frac{\Gamma(N+\sum x_i)}{\Gamma(N)}(1+\sum\lambda_i)^{-N-\sum x_i} \prod \frac{\lambda_i^{x_i}}{x_i!}$$

for $x_i=1,2,\ldots,\ \lambda_i>0$.

Here $B(x,\lambda) = \Gamma(N+x)(1+\lambda)^{-N-x}$, $\psi(\lambda,x) = \dfrac{\lambda^x}{x!}$.

(5) Inverted Dirichlet with fixed θ :

$$\phi(\underline{x};\underline{\lambda}) = \frac{\Gamma(\theta+\sum\lambda_i)}{\Gamma(\theta)(1+\sum x_i)^{\theta+\sum\lambda_i}} \prod \frac{x_i^{\lambda_{i-1}}}{\Gamma(\lambda_i)}$$

for $x_i>0,\ \lambda_i>0$.

Here $B(x,\lambda) = \Gamma(\theta+\lambda)(1+x)^{-\theta-\lambda}$, $\psi(\lambda,x) = \dfrac{x^{\lambda-1}}{\Gamma(\lambda)}$.

(6) Dirichlet compound negative multinomial with fixed N,θ:

$$\phi(\underline{x};\underline{\lambda}) = \frac{\Gamma(N+\theta)}{\Gamma(N)\Gamma(\theta)} \frac{\Gamma(\theta+\sum\lambda_i)\Gamma(N+\sum x_i)}{\Gamma(N+\theta+\sum\lambda_i+\sum x_i)} \prod \frac{\Gamma(x_i+\lambda_i)}{x_i!\Gamma(\lambda_i)}$$

for $x_i=0,1,2,\ldots,\ \lambda_i>0$.

Here $B(x,\lambda) = \dfrac{\Gamma(N+x)\Gamma(\theta+\lambda)}{\Gamma(N+\theta+\lambda+x)}$, $\psi(\lambda,x) = \dfrac{\Gamma(x+\lambda)}{x!\Gamma(\lambda)}$.

Proposition 3.22. If $\underline{X}_{\underline{\lambda}}$ has its distribution in the form of (1) to (6) discussed above, then for $m \leq n$:

(1) $(\lambda_1,\ldots,\lambda_m) \underset{w}{\prec} (\overset{w}{\prec})(\lambda_1',\ldots,\lambda_m') \Longrightarrow$

$(X_1,\ldots,X_m|\lambda_1,\ldots,\lambda_m)\ \mathrm{st}\underset{w}{\prec}\ (\mathrm{st}\overset{w}{\prec})(X_1,\ldots,X_m|\lambda_1',\ldots,\lambda_m')$

(2) $\Pr\{r_1<(\text{or}\leq)X_i<(\text{or}\leq)r_2$ for $i=1,\ldots,m|\lambda_1,\ldots,\lambda_m\}$ is a decreasing Schur concave function of $(\lambda_1,\ldots,\lambda_m)$.

(3) $\Pr\{$number of zero components of $(X_1,\ldots,X_m)>k|\lambda_1,\ldots,\lambda_m\}$ is a decreasing Schur convex function of $(\lambda_1,\ldots,\lambda_m)$.

Proof : (1) follows from Theorem 3.13, 3.16 and the remark concerning the function B . The proofs for (2) and (3) follow the same line as that of Proposition 3.18, together with the application of (1).

E. Application to hypothesis testing.

Consider a Schur convex family, parametrized by $\underline{\lambda}$, $\lambda_i>0$. Given $\lambda_1+\ldots+\lambda_n=nk$ where k is either a positive integer or just a positive number, depending on the form of Λ, we want to test the null hypothesis H_0: $\lambda_1=\ldots=\lambda_n=k$ versus the alternative H_1: not all the λ_i equal k .

Intuitively, symmetry of the distribution function implies that if the λ_i are equal, then it is more probable that the components X_i do not differ greatly, which also means that

$X_i/\sum X_j$ do not differ greatly. Formulated in terms of Schur functions, this means that when F is Schur convex (Schur concave), $F(X_1/\sum X_j,\ldots,X_n/\sum X_j)$ tends to be smallest (largest) under H_0. So we can take as a test function

$$T(\underline{x}) = \begin{cases} 1 & \\ \gamma & \text{if } F(x_1/\sum x_j,\ldots,x_n/\sum x_j) \quad >(<) \\ & \quad = c \\ 0 & \quad <(>) \end{cases}$$

where F is some Schur convex (Schur concave) function, $0\leq\gamma\leq 1$ and c are chosen so as to achieve the desired size. It is easy to verify that T is a Schur convex function in both cases, so that the power function $\beta_T(\underline{\lambda}) \equiv E_{\underline{\lambda}}T(\underline{X})$ is Schur convex in $\underline{\lambda}$. Therefore, β_T attains its minimum at the null point $(k,\ldots,k)$. Consequently, the test is unbiased.

The case of the multinomial family, with $F(x_1/\sum x_j,\ldots,x_n/\sum x_j) \equiv F(x_1/N,\ldots,x_n/N) = \sum h(x_i)$, h convex, is first found by Cohen and Sackrowitz (1975). A generalization of $\sum h(x_i)$ to Schur convex functions $G(\underline{x})$ is found by Perlman and Rinott (to appear). Cohen and Sackrowitz (1975) also point out that the class $\sum h(x_i)$ includes the likelihood ratio test with $h(x)=x\log x$ and the goodness of fit test with $h(x)=x^2$. We can verify that the goodness of fit test with $h(x)=x^2$ also applies to the hypergeometric distribution.

Bibliography

[1] Chong, K.M. (1976). An induction theorem for rearrangements. Canadian Journal of Mathematics 28, 154-160.

[2] Cohen, A. and Sackrowitz, H.B. (1975). Unbiasedness of the Chi-square, likelihood ratio, and other goodness of fit tests for the equal cell case. Annals of Statistics 3, 959-964.

[3] Fuchs, L. (1947). A new proof of an inequality of Hardy-Littlewood-Polya. Matematisk Tidsskrift B, 53-54.

[4] Hardy, G.H., Littlewood, J.E., and Pólya, G. (1929). Some simple inequalities satisfied by convex functions. Messenger of Mathematics 58, 145-152.

[5] Hardy, G.H., Littlewood, J.E. and Pólya, G. (1952). Inequalities (second edition). Cambridge University Press.

[6] Hollander, M., Proschan, F. and Sethuraman, J. (to appear). Functions decreasing in transposition and their applications in ranking problems.

[7] Marshall, A.W. and Olkin, I. (to appear). Inequalities via majorization.

[8] Marshall, A.W. and Proschan, F. (1965). An inequality for convex functions involving majorization. Journal of Mathematical Analysis and Applications 12, 87-90.

[9] Mirsky, L. (1959). Inequalities for certain class of convex functions. Proceedings of Edinburgh Mathematical Society 11, 231-235.

[10] Mitronović, D.S. (1970). Analytic inequalities. Springer-Verlag, New York.

[11] Mudholkar, G.S. (1966). The integral of an invariant unimodal function over an invariant set- an inequality and applications. Proceedings of American Mathematical Society 17, 1327-1333.

[12] Nevius, S.E., Proschan, F. and Sethuraman, J. (1977) Schur functions in statistics II: Stochastic majorization. Annals of Statistics 5, 263-273. Also, Florida State University Technical Report M306, (1974).

[13] Nevius, S.E., Proschan, F. and Sethuraman, J. (1974) Schur functions in Statistics III: A stochastic version of weak majorization, with applications. Florida State University Technical Report M307.

[14] Olkin, I. (1974). Monotonicity properties of Dirichlet integrals with applications to the multinomial distribution and the analysis of variance. Biometrika 59, 303-307.

[15] Ostrowski, A. (1952). Sur quelques applications des functions convexes et concaves au sens de I. Schur. Journal de Mathématiques Pure et appliquées 31, 253-292.

[16] Perlman, M.D. and Rinott, Y. (to appear) On the unbiasedness of goodness-of-fit tests.

[17] Proschan, F. and Sethuraman, J. (1977). Schur functions in statistics I: The preservation theorem. Florida State University Technical Report M254-RR.

[18] Rinott, Y. (1973). Multivariate majorization and rearrangement inequalities with some applications to probability and statistics. Israel Journal of Mathematics 15, 60-77.

[19] Tomić, M. (1949). Gauss' theorem on the centroid and its application. Bulletin de la Société des Mathématiciens et Physiciens de la R.P. de Serbie 1, 31-40.

[20] Wong, C.K. and Yue, P.C. (1973). A majorization theorem for the number of distinct outcomes in N independent trials. Discrete Mathematics 6, 391-398.