MULTIVARIATE MAJORIZATION AND REARRANGEMENT INEQUALITIES WITH SOME APPLICATIONS TO PROBABILITY AND STATISTICS

BY

YOSSEF RINOTT

ABSTRACT

Multivariate generalizations of the concept of a Schur convex-function are defined and characterized. These characterizations are shown to be useful in obtaining majorization and rearrangement inequalities. We give simple derivations of known results as well as new ones with applications in probability and statistics.

1. Introduction

Rearrangement and majorization inequalities have been the subject of many recent papers, e.g. [4], [10], [12]. Our main stimulation derives from the works of Lorentz [9] and Fan and Lorentz [3] (see Section 2). The theory bears applications to the study of statistical comparison of distributions, reliability theory, etc.

Let $\boldsymbol{a} = (a_1, \cdots, a_n)$ be an n-tuple of real numbers. We denote by $\boldsymbol{a}^*$ the decreasing rearrangement of $\boldsymbol{a}$, i.e. $\boldsymbol{a}^*$ is composed of the components of $\boldsymbol{a}$, set in decreasing order.

Let $\mathscr{C}$ be a prescribed collection of real valued functions. A class of partial orderings of R^n can be defined as follows: for $\boldsymbol{a}, \prec \in R^n$, we write $\boldsymbol{a} \prec \boldsymbol{\alpha}$ if

$$\sum_{i=1}^{n} f(a_i) \leqq \sum_{i=1}^{n} f(\alpha_i) \quad \text{for all } f \text{ in } \mathscr{C}.$$

It is established in Hardy et al. [5] that when $\mathscr{C}$ is the class of all convex functions, the ordering $\boldsymbol{a} \prec \boldsymbol{\alpha}$ ia equivalent to the relations

$$\text{(1.2)} \qquad \sum_{i=1}^{k} a_i^* \leq \sum_{i=1}^{k} \alpha_i^*, \qquad k = 1, \cdots, n$$

Received December 5, 1972

with equality for $k = n$. In this case $\boldsymbol{a} \prec \boldsymbol{\alpha}$ can also be expressed by the relation $\boldsymbol{a} = \boldsymbol{\alpha} T$ valid for some doubly stochastic matrix T.

A wide class of ordering relations is generated by specifying $\mathscr{C}$ to be a cone of generalized convex functions (for this concept see Karlin and Studden [7, Ch. XI]). In this context the ordering $\boldsymbol{a} \prec \boldsymbol{\alpha}$ reduces to the expression that $\mu_\alpha - \mu_a$ belongs to the dual cone of $\mathscr{C}$ where μ_a and μ_α represent the measures concentrating unit masses at the points of $\{a_i\}$ and $\{\alpha_i)$ respectively.

Henceforth we concentrate on the case where $\mathscr{C}$ comprises the collection of all convex functions and the relation $\prec$ is that of "majorization" [5, p. 45] which is equivalent to (1.2).

In this paper we characterize the functions monotone with respect to the ordering relation of majorization. That is, we delimit the functions ψ which satisfy

$$\psi(\boldsymbol{a}^1, \cdots, \boldsymbol{a}^m) \leqq \psi(\boldsymbol{\alpha}^1, \cdots, \boldsymbol{\alpha}^m)$$

for every set of n_i-tuples $\boldsymbol{a}^i = (a_1^i, \cdots, a_{n_i}^i)$ and $\boldsymbol{\alpha}^i = (\alpha_1^i, \cdots, \alpha_{n_i}^i)$ satisfying $\boldsymbol{a}^i \prec \boldsymbol{\alpha}^i$ $i = 1, \cdots, m$.

This monotonicity property, and another one (both in Section 3) can be construed as multivariate analogues of the concept of Schur functions (Ostrowski [13]).

We show (in Section 4) that the theorem of Fan and Lorentz is a special case ot our characterization. Section 4 offers further examples of probability inequalities, with emphasis on the multinomial distribution. We finish by presenting some consequences emanating from the inequality of Lorentz and show that a number of recent publications are subsumed by [9].

2. Remarks on results of Fan and Lorentz

In this section we state the results of Lorentz [9] and Fan and Lorentz [3] and point out how the latter result can be obtained in a simple way from the former. We add a few remarks which will be useful later, in the applications.

A function of two variables $f(x_1, x_2)$ is said to be a positive set function if

$$f(x_1 + h, x_2 + k) - f(x_1 + h, x_2) - f(x_1, x_2 + k) + f(x_1, x_2) \geqq 0 \tag{2.1}$$

for h, $k \geqq 0$, and arbitrary choices of x_1, x_2. Note that when f has continuous second partial derivatives, condition (2.1) is equivalent to

$$\frac{\partial^2 f}{\partial x_1 \partial x_2} \geqq 0. \tag{2.2}$$

Let $a^i = (a_1^i, \cdots, a_n^i)$, $i = 1, \cdots, m$ be n-tuples of real numbers. Let α^i be the decreasing rearrangement of a^i so that $\alpha^i = a^{i*}$ in the notation of Section 1. Let $\phi(x_1, \cdots, x_m)$ be a positive set function in each pair of variables (the other $m-2$ kept fixed). Lorentz [9] established the following inequality

$$\sum_{j=1}^{n} \phi(a_j^1, \cdots, a_j^m) \leqq \sum_{j=1}^{n} \phi(\alpha_j^1, \cdots, \alpha_j^m). \tag{2.3}$$

The proof consists of the observation that the rearrangement of the vectors a can be done in steps, interchanging two terms of a vector at a time. This reduces the problem to the simple case where we have to show that if

$$a = (a_1, a_2), \qquad b = (b_1, b_2), \qquad a^* = \alpha, \qquad b^* = \beta$$

then

$$\phi(a_1, b_1) + \phi(a_2, b_2) \leqq \phi(\alpha_1, \beta_1) + \phi(\alpha_2, \beta_2)$$

which coincides with condition (2.1).

The result of Fan and Lorentz [3] is as follows: Let a^i be as above and let α^i be n-tuples of decreasing numbers satisfying now $a^i \prec \alpha^i$ (i.e. (1.2) holds) for $i = 1, \cdots, m$. Let ϕ be a positive set function in each pair of variables, which satisfies the additional condition that ϕ is convex in each variable; then (2.3) is again valid.

REMARKS.

1. It follows easily that if $\phi(x_1, \cdots, x_k, \cdots, x_m)$ determines a positive set function for every pair of variables x_i, x_j when $i, j > k$, or $i, j \leqq k$, and a negative set function provided $i \leqq k$ and $j > k$ then (2.3) takes the form

$$\sum_{j=1}^{n} \phi(a_j^1, \cdots, a_j^k, a_j^{k+1}, \cdots, a_j^m) \leqq \sum_{j=1}^{n} \phi(\tilde{a}_j^1, \cdots, \tilde{a}_j^k, a_j^{k+1*}, \cdots, a_j^{m*})$$

where $\tilde{a}^i = (\tilde{a}_1^i, \cdots, \tilde{a}_n^i)$ is the increasing rearrangement of a^i, $i = 1, \cdots, k$.

2. Let $\psi(a, b)$ be a function of two variables and introduce the matrix

$$\psi(a, b) = \begin{bmatrix} \psi(a_1, b_1) & \psi(a_1, b_2) & \cdots & \psi(a_1, b_n) \\ \cdot & & & \\ \cdot & & & \\ \psi(a_n, b_1) & & \cdots & \psi(a_n, b_n) \end{bmatrix}$$

$$a = (a_1, \cdots, a_n), \qquad b = (b_1, \cdots, b_n).$$

A positive function ψ is called totally positive of order $2\,(TP_2$; see Karlin [6]) if every second order subdeterminant of $\psi(\boldsymbol{a},\boldsymbol{b})$ is non-negative. It is readily checked that a positive ψ is TP_2 iff $\phi(a,b) = \log\psi(a,b)$ is a positive set function. Thus a version of (2.3) can be cast in the form

$$\prod_{j=1}^{n} \psi(a_j^1,\cdots,a_j^m) \leqq \prod_{j=1}^{n} \psi(a_j^{1*},\cdots,a_j^{m*}) \tag{2.4}$$

for any positive function ψ which is TP_2 as a function of any two of its variables.

3. Let W be a permutation matrix. Observe that (2.3) can be expressed for the case $m = 2$ in

$$\operatorname{tr} W\phi(\boldsymbol{\alpha},\boldsymbol{\beta}) \leqq \operatorname{tr}\phi(\boldsymbol{\alpha},\boldsymbol{\beta}) \qquad (\operatorname{tr} A = \text{trace of } A) \tag{2.5}$$

where $\phi(a,\beta)$ is a positive set function, $\boldsymbol{\alpha},\boldsymbol{\beta}$ are n-tuples of decreasing numbers. Since any doubly stochastic matrix can be written as a convex combination of permutation matrices, it follows that (2.5) applies for any doubly stochastic W.

4. Proof of the inequality of Fan and Lorentz: For ease of notation, we restrict ourselves to the case $m = 2$ (see also Section 4). Let $\boldsymbol{a},\boldsymbol{b}$ be n-tuples of real numbers and $\boldsymbol{\alpha},\boldsymbol{\beta}$ n-tuples of decreasing numbers satisfying $\boldsymbol{a} \prec \boldsymbol{\alpha}$, $\boldsymbol{b} \prec \boldsymbol{\beta}$. By a theorem of Hardy, Littlewood and Pólya [5] there exist doubly stochastic matrices S and T such that $\boldsymbol{a} = \boldsymbol{\alpha}S$, $\boldsymbol{b} = \boldsymbol{\beta}T$. Writing out the indicated expressions and invoking the convexity of ϕ in each variable we obtain

$$\operatorname{tr}\phi(\boldsymbol{a},\boldsymbol{b}) = \operatorname{tr}\phi(\boldsymbol{\alpha}S,\boldsymbol{\beta}T) \leqq \operatorname{tr} ST'\phi(\boldsymbol{\alpha},\boldsymbol{\beta}) \leqq \operatorname{tr}\phi(\boldsymbol{a},\boldsymbol{\beta})$$

where the last inequality results from Remark 3 above. But the inequality $\operatorname{tr}\phi(\boldsymbol{a},\boldsymbol{b}) \leqq \operatorname{tr}\phi(\boldsymbol{\alpha},\boldsymbol{\beta})$ is exactly (2.3) in the case $m = 2$.

In Section 4 we shall indicate that this result of Fan and Lorentz can be derived as a special case of our treatment of multivariate Schur convex functions.

3. The main theorems

DEFINITION 1. Let $\boldsymbol{a}^i = (a_1^i,\cdots,a_{n_i}^i)$, $i = 1,\cdots,m$ denote n_i-tuples of decreasing numbers. A function $\psi(\boldsymbol{a}^1,\cdots,\boldsymbol{a}^m)$ is said to be a multivariate Schur function if $\psi(\boldsymbol{a}^1,\cdots,\boldsymbol{a}^m) \leqq \psi(\boldsymbol{\alpha}^1,\cdots,\boldsymbol{\alpha}^m)$ for every set of n_i-tuples $\boldsymbol{\alpha}^i = (\alpha_1^i,\cdots,\alpha_{n_i}^i)$, $i = 1,\cdots m$, of decreasing numbers such that $\boldsymbol{a}^i \prec \boldsymbol{\alpha}^i$ (i.e. relations (1.2) hold). For $m = 1$ this reduces to the old concept of a Schur function (Ostrowski [13]). A function of n variables is said to be a Schur function if $\psi(\boldsymbol{a}) \leqq \psi(\boldsymbol{\alpha})$ for any n-tuples of decreasing numbers $\boldsymbol{a} = (a_1,\cdots,a_n)$, $\boldsymbol{\alpha} = (\alpha_1,\cdots,\alpha_n)$ which satisfy $\boldsymbol{a} \prec \boldsymbol{\alpha}$.

It is well known that a symmetric convex function is a Schur function. It can be easily shown that a symmetric function ϕ which is convex in each one of its variables and which satisfies

$$\phi(a_1, a_1, a_3, \cdots, a_n) + \phi(a_2, a_2, a_3, \cdots, a_n) \leqq 2\phi(a_1, a_2, a_3, \cdots, a_n)$$

is also a Schur function. Note that if ϕ is a negative set function (i.e. (2.1) holds with the inequality reversed) then the above inequality clearly holds.

We first examine the case $m = 1$ and stipulate that $\psi(a_1, \cdots, a_m)$ is differentiable. (Henceforth we shall use differentiability as needed without stating it as one of the conditions).

The following lemma is established in [13]. Our simple proof may be of some interest.

LEMMA 1. *ψ is a Schur function, i.e. ψ satisfies*

$$\psi(a_1, \cdots, a_n) \leqq \psi(\alpha_1, \cdots, \alpha_n)$$

for any n-tuples of decreasing numbers $\boldsymbol{a} = (a_1, \cdots, a_n)$, $\boldsymbol{\alpha} = (\alpha_1, \cdots, \alpha_n)$ satisfying $\boldsymbol{a} \prec \boldsymbol{\alpha}$, if and only if

(3.1) $\quad \dfrac{\partial}{\partial a_k}\psi(\boldsymbol{a}) - \dfrac{\partial}{\partial a_j}\psi(\boldsymbol{a}) \geqq 0$ *for all $k < j$ and all decreasing n-tuples $\boldsymbol{a}$.*

PROOF. Define ϕ by $\phi(z_1, \cdots, z_n) = \psi(z_1, z_2 - z_1, \cdots, z_n - z_{n-1})$. We have $\psi(a_1, \cdots, a_n) = \phi(a_1, a_1 + a_2, \cdots, a_1 + a_2 + \cdots + a_n)$.
Recall that $\boldsymbol{a} \prec \boldsymbol{\alpha}$ iff $a_1 + a_2 + \cdots + a_k \leqq \alpha_1 + \alpha_2 + \cdots + \alpha_k$ for $k \leqq n - 1$ and thus ψ is a Schur function if and only if ϕ is monotone in its first $n - 1$ variables, on the domain where $z_n = a_1 + \cdots + a_n$ is constant. Differentiating for $1 \leqq k < n$ gives

$$0 \leqq \frac{\partial}{\partial z_k}\phi(z_1, \cdots, z_n) = \frac{\partial}{\partial z_k}\psi(z_1, z_2 - z_1, \cdots, z_n - z_{n-1}) = \frac{\partial \psi}{\partial a_k}(\boldsymbol{a}) - \frac{\partial \psi}{\partial a_{k+1}}(\boldsymbol{a}).$$

This implies (3.1) and the proof is complete.

Since $\psi(\boldsymbol{a}^1, \cdots, \boldsymbol{a}^m)$ is a multivariate Schur function if and only if it is a Schur function as a function of any one of its vector variables, the others kept constant, we deduce

THEOREM 1. *$\psi(\boldsymbol{a}^1, \cdots, \boldsymbol{a}^m)$ is a multivariate Schur function if and only if*

$$\frac{\partial\psi}{\partial a_k^i}(a^1,\cdots,a^m) - \frac{\partial\psi}{\partial a_j^i}(a^1,\cdots,a^m) \geqq 0 \tag{3.2}$$

for all $1 \leqq k < j \leqq n_i$, $i = 1,\cdots,m$ *where* a^i *are* n_i*-tuples of decreasing numbers.*

DEFINITION 2. Let $a^1,\cdots,a^m$ be m n-tuples of real numbers. A function $\psi(\alpha^1,\cdots,\alpha^m)$ is said to be a multivariate *symmetric* Schur function if

$$\psi(\alpha^1 T,\cdots,\alpha^m T) \leqq \psi(\alpha^1,\cdots,\alpha^m)$$

for any $n \times n$ doubly stochastic matrix T, and any n-tuples $\alpha^1,\cdots,\alpha^m$.

The definition tacitly presupposes certain symmetry properties to the extent that the value of the function is unchanged when the same permutation is applied to all vectors. This fact follows since the inverse of a permutation matrix is also a permutation matrix and both are doubly stochastic.

For $m = 1$ our definition coincides with that of a symmetric Schur function, i.e. a symmetric function satisfying $\psi(\alpha) \leqq \psi(\alpha)$ whenever $a \prec \alpha$ holds. (Here a and α are not necessarily in decreasing order.)

THEOREM 2. *ψ is a symmetric multivariate Schur function if and only if*

$$\sum_{i=1}^{m} (\alpha_k^i - \alpha_j^i)\left[\frac{\partial\psi}{\partial\alpha_k^i}(\alpha^1,\cdots,\alpha^m) - \frac{\partial\psi}{\partial\alpha_j^i}(\alpha^1,\cdots,\alpha^m)\right] \geqq 0$$

for all $k, j \leqq n.$

Manifestly the above inequality is a weaker condition than (3.2).

PROOF. Assume ψ is Schur, and take

$$T = \begin{pmatrix} t & 1-t & 0 & \cdot & \cdot & \cdot & 0 \\ 1-t & t & 0 & \cdot & \cdot & \cdot & 0 \\ 0 & 0 & 1 & 0 & \cdot & \cdot & 0 \\ \cdot & \cdot & & 1 & & & \cdot \\ \cdot & \cdot & & & \cdot & & \cdot \\ \cdot & \cdot & & & & \cdot & \cdot \\ & & & & & \cdot & 0 \\ 0 & 0 & \cdot & \cdot & \cdot & 0 & 1 \end{pmatrix}$$

where $0 < t < 1$.

By adding and subtracting terms in an obvious way, we get, invoking the mean value theorem

$$0 \leqq \frac{\psi(\alpha^1,\cdots,\alpha^m) - \psi(\alpha^1 T,\cdots,\alpha^m T)}{1-t} \underset{t\to 1}{\to} \sum_{i=1}^{m} (\alpha_1^i - \alpha_2^i)\left(\frac{\partial\psi}{\partial x_1^i} - \frac{\partial\psi}{\partial x_2^i}\right).$$

Conversely, define

$$\phi(t) = \psi(\alpha^1 T,\cdots,\alpha^m T)$$

for T explicitly as above. Then

$$\phi'(t) = \sum_{i=1}^{m} (\alpha_1^i - \alpha_2^i)\left[\frac{\partial\psi}{\partial \alpha_1^i}(a^1,\cdots,a^n) - \frac{\partial\psi}{\partial \alpha_2^i}(a^1,\cdots,a^n)\right]$$

where $a^i = \alpha^i T$.

Now $a_1^i - a_2^i = (2t-1)(\alpha_1^i - \alpha_2^i)$, and so by our assumption $\phi'(t) \geqq 0$ for $t > \frac{1}{2}$. But $\phi(t)$ is symmetric about $t = \frac{1}{2}$ which implies

$$\phi(1) \geqq \phi(t) \text{ for } 0 \leqq t \leqq 1,$$

i.e. $\psi(\alpha^1,\cdots,\alpha^m) \geqq \psi(\alpha^1 T,\cdots,\alpha^m T)$.

For a general doubly stochastic matrix the result is obtained by interchanging the first and k-th rows and the first and j-th columns of T, for $k, j \leqq n$ and repeating the above argument, and then applying the fact that every doubly stochastic matrix is a product of a finite number of matrices of the above form [5, p. 47].

4. Applications

EXAMPLE 1. From our general setting we indicate how to extract the inequality of Fan and Lorentz which we now restate in a slightly more general form than in Section 2. Let a^i, $i = 1,\cdots,m$, be n-tuples of decreasing numbers, and let a^0 be an n-tuple of increasing numbers which will be kept fixed. Let $\phi(x_0, x_1,\cdots,x_m)$ be a function of $m+1$ real variables and define

$$\psi(a^1,\cdots,a^m) = \sum_{j=1}^{n} \phi(a_j^0, a_j^1,\cdots,a_j^m).$$

We prove that ψ is a multivariate Schur function as a function of $a^1,\cdots,a^m$, with fixed (but arbitrary) a^0, if and only if

$$\frac{\partial^2}{\partial x_i \partial x_k}\phi(y_0, y_1,\cdots,y_m) \geqq 0, \qquad 1 \leqq i,k \leqq m$$

(4.1)

$$\frac{\partial^2}{\partial x_i \partial x_0}\phi(y_0, y_1,\cdots,y_m) \leqq 0, \qquad 1 \leqq i \leqq m$$

for all real $y_0,\cdots,y_m$, in the domain of ϕ.

PROOF. Since a^i are vectors of decreasing numbers we consider for $1 \leqq i \leqq m$, $1 \leqq k < n$, the expression

$$\frac{\partial \psi}{\partial a_k^i}(a^1, \cdots, a^m) - \frac{\partial \psi}{\partial a_{k+1}^i}(a^1, \cdots, a^m).$$

This equals

$$\frac{\partial \phi}{\partial x_i}(a_k^0, a_k^1, \cdots, a_k^m) - \frac{\partial \phi}{\partial x_i}(a_{k+1}^0, a_{k+1}^1, \cdots, a_{k+1}^m) = \sum_{j=0}^{m} (a_k^j - a_{k+1}^j) \frac{\partial^2}{\partial x_i \partial x_j} \phi(y)$$

where $y = (y_0, y_1, \cdots, y_m)$ is a suitable point arising by application of the mean value theorem. It follows that

$$\frac{\partial \psi}{\partial a_k^i}(a^1, \cdots, a^m) - \frac{\partial \psi}{\partial a_{k+1}^i}(a^1, \cdots, a^m) \geqq 0$$

if the conditions of (4.1) are satisfied. By making a special choice of the terms of a^i it is easy to see that conditions (4.1) are also necessary. Appealing to Theorem 1 the proof is complete.

By a standard approximation procedure we can deduce that (4.1) is necessary and sufficient for ϕ to satisfy

$$\int_0^1 \phi(t, f_1(t), \cdots, f_m(t))dt \leqq \int_0^1 \phi(t, g_1(t), \cdots, g_m(t))dt$$

for every system of decreasing bounded functions f_i, g_i, $i = 1, \cdots, m$ such that $f_i \prec g_i$, i.e.

$$\int_0^x f_i(t)dt \leqq \int_0^x g_i(t)dt, \quad 0 \leqq x \leqq 1 \text{ with equality for } x = 1 .$$

EXAMPLE 2. Let $X = (X_1, \cdots, X_k)$ have the multinomial distribution

$$P(X = x) = \binom{N}{x_1, \cdots, x_k} \prod_{i=1}^{k} \theta_i^{x_i}$$

where $x = (x_1, \cdots, x_k)$, $\sum_{i=1}^k x_i = N$, and $\sum_{i=1}^k \theta_i = 1$.
Let $\phi(x_1, \cdots, x_k)$ be a symmetric Schur function (i.e. ϕ is symmetric and satisfies $\phi(x) \leqq \phi(y)$ whenever $x \prec y$). Define the function $\psi(\theta)$ as the expectation

$$\psi(\theta) = E_\theta \phi(X) = \sum \phi(x) \binom{N}{x_1, \cdots, x_k} \prod_{i=1}^{k} \theta_i^{x_i}$$

where the sum extends over all k-tuples of non-negative integers $\mathbf{x} = (x_1, \cdots, x_k)$ for which $\sum_{i=1}^{k} x_i = N$ holds.

PROPOSITION. *$\psi(\boldsymbol{\theta})$ is a Schur function.*

PROOF. We show that condition (3.1) of Lemma 1 (Section 3) is satisfied.

$$\frac{\partial\psi(\boldsymbol{\theta})}{\partial\theta_1} - \frac{\partial\psi(\boldsymbol{\theta})}{\partial\theta_2} = N! \ \Sigma \ \phi(\mathbf{x}) \prod_{i=1}^{k} \left(\frac{\theta_i^{x_i}}{x_i!}\right) \left(\frac{x_1}{\theta_1} - \frac{x_2}{\theta_2}\right).$$

Rearranging the order of summation the sum becomes

$$N! \ \Sigma \prod_{i=1}^{k} \frac{\theta_i^{y_i}}{y_i!} \left(\phi(y_1 + 1, y_2, \cdots, y_k) - \phi(y_1, y_2 + 1, y_3, \cdots, y_k)\right)$$

where the sum now extends over all k-tuples $\mathbf{y} = (y_1, \cdots, y_k)$ for which $\sum_{i=1}^{k-1} y_i = N - 1.$

Invoking the symmetry of ϕ we finally obtain

$$\frac{\partial\psi(\boldsymbol{\theta})}{\partial\theta_1} - \frac{\partial\psi(\boldsymbol{\theta})}{\partial\theta_2}$$

$$= N! \ \Sigma \left(\frac{\theta_1^{y_1}\theta_2^{y_2}\cdots\theta_k^{y_k}}{y_1!y_2!\cdots y_k!} - \frac{\theta_1^{y_2}\theta_2^{y_1}\theta_3^{y_3}\cdots\theta_k^{y_k}}{y_1!y_2!\cdots y_k!}\right) \Big(\phi(y_1 + 1, y_2, \cdots, y_k) - \phi(y_1, y_2 + 1, \cdots, y_k)\Big)$$

where the sum extends over $(y_1, \cdots, y_k)$ as above satisfying the additional condition $y_1 > y_2$. Since ϕ is a Schur function

$$\phi(y_1 + 1, y_2, \cdots, y_k) - \phi(y_1, y_2 + 1, y_3, \cdots, y_k) \geqq 0$$

for $y_1 > y_2$. Since

$$\theta_1^{y_1}\theta_2^{y_2} - \theta_1^{y_2}\theta_2^{y_1} \geqq 0 \quad \text{for} \quad y_1 > y_2 \quad \text{and} \quad \theta_1 > \theta_2$$

we have

$$\frac{\partial\psi(\boldsymbol{\theta})}{\partial\theta_1} - \frac{\partial\psi(\boldsymbol{\theta})}{\partial\theta_2} \geqq 0$$

for $\theta_1 > \theta_2$ and the proof is complete.

As an example consider the expectations

$$\psi_1(\boldsymbol{\theta}) = E_{\boldsymbol{\theta}}(\max_{1 \leqq i \leqq k} X_i)^\alpha$$

$$\psi_2(\boldsymbol{\theta}) = E_{\boldsymbol{\theta}}(\min_{1 \leqq i \leqq k} X_i)^\alpha$$

$$\alpha > 0$$

where X_i are distributed as above. Since $(\max_{1 \leqq i \leqq k} x_i)^\alpha$ and $-(\min_{1 \leqq i \leqq k} x_i)^\alpha$ are both symmetric Schur functions, the proposition implies that $\psi_1(\boldsymbol{\theta})$ and $-\psi_2(\boldsymbol{\theta})$ are Schur functions.

Thus we obtain for example that $E_{\boldsymbol{\theta}}(\max_{1 \leqq i \leqq k} X_i)$ is minimized for $\boldsymbol{\theta} = (1/k, \cdots, 1/k)$ since for any $\boldsymbol{\theta} = (\theta_1, \cdots, \theta_k)$ where $\sum_{i=1}^k \theta_i = 1$ we have $(1/k, \cdots, 1/k) \prec (\theta_1, \cdots, \theta_k)$.

Results of this kind are applicable in determining least favorable distributions for testing multinomial hypotheses. (See, for example, Gupta and Nagel [4], where some tables of $\psi_1(\boldsymbol{\theta})$ and $\psi_2(\boldsymbol{\theta})$ are given.)

In the same way we obtain that the function $\psi_1(\boldsymbol{\theta}) = -P_{\boldsymbol{\theta}}(X_1 \geqq r, \cdots, X_n \geqq r)$ and $\psi_2(\boldsymbol{\theta}) = -P_{\boldsymbol{\theta}}(X_1 < r, \cdots, X_k < r)$ (r fixed) are Schur functions. This is so because

$$P_{\boldsymbol{\theta}}(X_1 \geqq r, \cdots, X_k \geqq r) = E_{\boldsymbol{\theta}} \phi_r(X)$$

$$P_{\boldsymbol{\theta}}(X_1 < r, \cdots, X_k < r) = E_{\boldsymbol{\theta}} \chi_r(X)$$

where

$$\phi_r(x) = \begin{cases} 1 & \min_{1 \leqq i \leqq k} x_i \geqq r \\ 0 & \min_{1 \leqq i \leqq k} x_i < r \end{cases}$$

$$\chi_r(x) = \begin{cases} 1 & \max_{1 \leqq i \leqq k} x_i < r \\ 0 & \max_{1 \leqq i \leqq k} x_i \geqq r \end{cases}$$

ans both $-\phi_r(x)$ and $-\chi_r(x)$, being monotone functions of $\min x_i$ and $\max x_i$ respectively, are symmetric Schur functions. For a different proof of this result see Olkin [12].

COROLLARY. *Let X_i, $i = 1, \cdots, k$ be independent random variables with the Poisson distribution, i.e.*

$$P(X_i = x) = e^{-\lambda_i} \frac{\lambda_i^x}{x!}, \qquad x = 0, 1, \cdots$$

If $\phi(x)$ is a symmetric Schur function then $E_\lambda \phi(X_1, \cdots, X_k)$ is a Schur function.

The proof follows from the fact that the distribution of X given $\Sigma_{i=1}^k X_i$ is the multinomial distribution, and from the fact that the collection of Schur functions is a convex cone.

EXAMPLE 3. Let $X_1, \cdots, X_n$ be independent binomial random variables with $P(X_i = 1) = p_i$. Denote $S = \Sigma_{i=1}^n X_i$. We show

$$\psi(p_1, \cdots, p_n) = E_{p_1, \cdots, p_n} \phi(S)$$

is a Schur function if ϕ is concave. For this we use Lemma 1,

$$\psi(p_1, \cdots, p_n) = \Sigma \; p_1^{x_1}(1 - p_1)^{1-x_1} \cdots p_n^{x_n}(1 - p_n)^{1-x_n} \phi\left(\sum_{i=1}^n x_i\right)$$

where the sum extends over all vectors $(x_1, \cdots, x_n)$ of zeros and ones.

$$\frac{\partial \psi}{\partial p_1} - \frac{\partial \psi}{\partial p_2} = -(p_1 - p_2) \; \Sigma \; p_3^{x_3} \cdots p_n^{x_n} \left[\phi\left(2 + \sum_{i=3}^n x_i\right) - 2\phi\left(1 + \sum_{i=3}^n x_i\right) + \phi\left(\sum_{i=3}^n x_i\right) \right]$$

which is non-negative by the concavity of ϕ. For a different proof see [7].

EXAMPLE 4. We now generalize the preceding example to the case of multinomial variables. Consider the $(m + 1)$-dimensional random variables

$$(X_k^0, X_k^1, \cdots, X_k^m) = \begin{cases} (1, 0, 0, \cdots, 0) & \text{with probability } p_k^0 \\ (0, 1, 0, \cdots, 0) & \text{with probability } p_k^1 \\ \cdot & \\ \cdot & \\ \cdot & \\ (0, \cdots, 0, 1) & \text{with probability } p_k^m \end{cases}$$

$$\sum_{i=0}^m p_k^i = 1, \qquad k = 1, \cdots, n.$$

Set $\boldsymbol{p}^i = (p_1^i, \cdots, p_n^i)$, $i = 1, \cdots, m$ and define

$$\Psi(\boldsymbol{p}^1, \cdots, \boldsymbol{p}^m) = E_{\boldsymbol{p}^1 \cdots \boldsymbol{p}^m} \Phi \left(\sum_{k=1}^n X_k^1, \sum_{k=1}^n X_k^2, \cdots, \sum_{k=1}^n X_k^m \right).$$

We shall derive conditions on Φ under which Ψ is a (symmetric) multivariate Schur function. Note that

$$\Psi(p^1,\cdots,p^m) = \Sigma \prod_{k=1}^{n} (p_k^0)^{x_k^0}(p_k^1)^{x_k^1}\cdots(P_k^m)^{x_k^m}\,\Phi\left(\sum_{k=1}^{n} x_k^1,\cdots,\sum_{k=1}^{n} x_k^m\right)$$

where the sum extends over all $(m+1)$-tuples of zeros and one $x_k^0,\cdots,x_k^m$ such that

$$\sum_{i=0}^{m} x_k^i = 1 \quad , \qquad k = 1,\cdots,n.$$

Differentiation produces

$$\frac{\partial \Psi}{\partial p_1^1} = \Sigma \prod_{k=2}^{n} (p_k^0)^{x_k^0}\cdots(p_k^m)^{x_k^m}\left[\Phi\left(1+\sum_{k=2}^{n} x_k^1,\ \sum_{k=2}^{n} x_k^2,\cdots,\sum_{k=2}^{n} x_k^m\right)\right.$$
$$\left. -\Phi\left(\sum_{k=2}^{n} x_k^1,\cdots,\sum_{k=2}^{n} x_k^m\right)\right].$$

For simplicity of notation we replace in the sequel each sum of the form $\sum_{k=3}^{n} x_k^i$ by γ^i; thus for example

$$\Phi(2+\gamma^1,\gamma^2,\cdots,\gamma^m) = \Phi\,(2+\sum_{k=3}^{n} x_k^1,\ \sum_{k=3}^{n} x_k^2,\cdots,\sum_{k=3}^{n} x_k^m) \text{ etc.}$$

We get

$$\frac{\partial \Psi}{\partial p_1^1} - \frac{\partial \Psi}{\partial p_2^1} = \Sigma \prod_{k=3}^{n} (p_k^0)^{x_k^0}\cdots(p_k^m)^{x_k^m} A$$

where the sum is as above only for $k = 3,\cdots,m$, and

$$A = (p_2^1 - p_1^1)\,[\Phi(2+\gamma^1,\gamma^2,\cdots,\gamma^m) - 2\Phi(1+\gamma^1,\gamma^2,\cdots,\gamma^m) + \Phi(\gamma^1,\cdots,\gamma^m)]$$
$$+ \sum_{i=2}^{m} (p_2^i - p_1^i)\,[\Phi(1+\gamma^1,\gamma^2,\cdots,\gamma^{i-1},1+\gamma^i,\gamma^{i+1},\cdots,\gamma^m)$$
$$- \Phi(1+\gamma,\gamma^2,\cdots,\gamma^m) - \Phi(\gamma^1,\cdots,\gamma^{i-1},1+\gamma^i,\gamma^{i+1},\cdots,\gamma^m) + \Phi(\gamma^1,\cdots,\gamma^m)].$$

By Theorem 1 we infer that if Φ is concave and a negative set function in each two variables then Ψ is a multivariate Schur function (consult Definition 1).

We now check the condition of Theorem 2. Using the above computation one gets

$$\sum_{i=1}^{m} (p_1^i - p_2^i)\left(\frac{\partial \Psi}{\partial p_1^i} - \frac{\partial \Psi}{\partial p_2^i}\right) = -\Sigma \prod_{k=3}^{n} (p_k^0)^{x_k^0} \cdots (p_k^m)^{x_k^m} \cdot Q'(\Delta_{ij}\Phi)\, Q$$

where $Q' = (p_1^1 - p_2^1, p_1^2 - p_2^2, \cdots, p_1^m - p_2^m)$

and $(\Delta_{ij}\Phi)$ denotes the $m \times m$ symmetric matrix with entries

$$\Delta_{ij}\Phi = \Phi(\gamma^1, \cdots, \gamma^{i-1}, 1+\gamma^i, \cdots, \gamma^{j-1}, 1+\gamma^j, \gamma^{j+1}, \cdots, \gamma^m)$$
$$- \Phi(\gamma^1, \cdots, \gamma^{i-1}, 1+\gamma^i, \gamma^{i+1}, \cdots, \gamma^m)$$
$$- \Phi(\gamma^1, \cdots, \gamma, \cdots, \gamma^{j-1}, 1+\gamma^j, \gamma^{j+1}, \cdots, \gamma^m) + \Phi(\gamma^1, \cdots, \gamma^m)$$

for $1 \leqq i < j \leqq m$, and for $i = 1, \cdots, m$

$$\Delta_{ii}\Phi = \Phi(\gamma^1, \cdots, \gamma^{i-1}, 2+\gamma^i, \gamma^{i+1}, \cdots, \gamma^m)$$
$$- 2\Phi(\gamma^1, \cdots, \gamma^{i-1}, 1+\gamma^i, \gamma^{i+1}, \cdots, \gamma^m) + \Phi(\gamma^1, \cdots, \gamma^m).$$

Thus Ψ is a symmetric multivariate Schur function (Definition 2) provided $(\Delta_{ij}\,\Phi)_{i,j=1}^m$ is non-positive definite. Observe that $(\Delta_{ij}\Phi)_{i,j=1}^m$ is the discrete analogue of the second differential, so that the condition becomes that Φ is a discrete-concave function.

EXAMPLE 5. Let X_i, $i = 1, \cdots, n$ be independent random variables with distribution given by $p(X_i = k) = (1/s_i)(1 - 1/s_i)^k$, $k = 0, 1, 2, \cdots$ where $s_i > 1$, $i = 1, \cdots, n$, i.e. X_i has a geometric distribution with

$$(2) \qquad p_i = 1 - \frac{1}{s_i}, \qquad i = 1, \cdots, n.$$

We set

$$\Psi(s_1, \cdots, s_n) = E_{s_1, \ldots, s_n}\Phi(X_1, \cdots, X_n)$$

where Φ is a symmetric function of n variables and we determine conditions under which $\Psi(s_1,\cdots,s_n)$ is a Schur function. The case where Φ is a function of the sum $\sum_{i=1}^{n} x_i$ was treated in [7].

We prove:

PROPOSITION. *If Φ is convex in each variable and if the function*

$$g(x_1,\cdots,x_{i-1},u,x_{i+1},\cdots,x_{j-1},v,x_{j+1},\cdots,x_n)$$
$$=\Phi(x_1,\cdots,x_{i-1},\frac{u+v}{2},x_{i+1},\cdots,x_{j-1},\frac{u-v}{2},x_{j+1},\cdots,x_n)$$

is nondecreasing in v and a positive set function in u and v for all $i\neq j$, $i,j=1,\ldots,n$ then $\Psi(s_1,\cdots,s_n)$ is a Schur function.

It is easily checked that $\Phi(x_1,\cdots,x_n)=\sum_{i=1}^{n}\gamma(x_i)$ satisfies the above conditions if both $\gamma(x)$ and its first derivative are convex and we conclude that $\psi(s_1,\cdots,s_n)=E_{s_1,\ldots,s_n}\sum_{i=1}^{n}\gamma(X_i)$ is a Schur function. In particular $E_{s_1,\ldots,s_n}\sum_{i=1}^{n}X_i^{\gamma}$, $\gamma\geqq 2$ is a Schur function.

PROOF OF PROPOSITION.

$$\Psi(s_1,\cdots,s_n)=\prod_{i=1}^{n}(1-p_i)\sum_{k_1,\ldots,k_n=0}^{\infty}p_1^{k_1}\cdots p_n^{k_n}\Phi(k_1,\cdots,k_n).$$

Differentiating with respect to s_1 and s_2 we get

$$\frac{\partial}{\partial s_1}\Psi-\frac{\partial}{\partial s_2}\Psi=\prod_{i=1}^{n}(1-p_i)\cdot\{(p_1-p_2)\sum_{k_1,\ldots,k_n=0}^{\infty}p_1^{k_1}\cdots p_n^{k_n}\Phi(k_1,\cdots,k_n)$$
$$+(1-p_1)^2\sum_{k_1,\ldots,k_n=0}^{\infty}k_1p_1^{k_1-1}p_2^{k_2}\cdots p_n^{k_n}\Phi(k_1,\cdots,k_n)$$
$$-(1-p_2)^2\sum_{k_1,\ldots,k_n=0}^{\infty}k_2p_1^{k_1}p_2^{k_2-1}p_3^{k_3}\cdots p_n^{k_n}\Phi(k_1,\cdots,k_n)\}.$$

We now regard the latter expression as a polynomial in $p_1,\cdots,p_n$ and rearrange the order of summation by collecting the coefficients of terms of equal degree (say $m+1$) in $p_1,\cdots,p_n$. We thus obtain

$$\frac{\partial}{\partial s_1}\Psi-\frac{\partial}{\partial s_2}\Psi=\prod_{i=1}^{n}(1-p_i)(p_1-p_2)\sum_{k_3,\ldots,k_n=0}^{\infty}A(k_3,\cdots,k_n)$$

where

$$A(k_3,\ldots,k_n) = \sum_{m=0}^{\infty} \sum_{k=0}^{[m/2]} B_{m-2k}p_1^k p_2^k \cdot \{(m-2k+1)\,[\Phi(m-k,k,k_3,\cdots,k_n)$$
$$-2\Phi(m-k+1,k,k_3,\cdots,k_n) + \Phi(m-k+2,k,k_3,\cdots,k_n)]$$
$$+[(k+1)\Phi(m-k+2,k,k_3,\cdots,k_n)$$
$$-(k+1)\Phi(m-k+1,k+1,k_3,\cdots,k_n)$$
$$-k\Phi(m-k+1,k-1,k_3,\cdots,k_n) + k\Phi(m-k,k,k_3,\cdots,k_n)]\}$$

and $B_n = \sum_{i=0}^{n} p_1^i p_2^{n-i}$.

By Lemma 1 we deduce that in order for Ψ to be a Schur function it is sufficient that Φ is convex in each variable and satisfies for every pair of variables the inequality

$$(k+1)\Phi(m-k+2,k) - (k+1)\Phi(m-k+1,k+1) - k\Phi(m-k+1,k-1) + k\Phi(m-k,k) \geqq 0 \tag{3}$$

where we have suppressed the variables which are unchanged. To simplify (3) define

$$g(u,v) = \Phi\left(\frac{u+v}{2}, \frac{u-v}{2}\right);$$

then $\Phi(x,y) = g(x+y, x-y)$ and (3) becomes

$$\{g(m+2, m-2k+2) - g(m+2, m-2k)\} + k\{g(m+2, m-2k+2) - g(m+2,m-2k) - g(m,m-2k+2) + g(m,m-2k)\} \geqq 0. \tag{4}$$

(4) will be satisfied if g is monotone in the second variable and a positive set function.

In the case that Φ is a function of $\sum_{i=1}^{n} x_i$ the expression in (3) vanishes and thus Ψ will be a Schur function if Φ is convex.

5. Some applications of the theorem of Lorentz

In this section we indicate how several recent results can be derived directly from the inequality of Lorentz [9]. Some of these observations were also made by Day [2].

1. Ruderman [14].

The function $\phi(x_1,\cdots,x_m) = \prod_{i=1}^{m} x_i$ is a positive set function in each pair of variables, for $x^i \geqq 0$, $i = 1,\cdots,m$. Thus for nonnegative n-tuples $\boldsymbol{a}^i$ $i = 1,\cdots,m$

$$\sum_{j=1}^{n} \prod_{i=1}^{m} a_j^i \leqq \sum_{j=1}^{n} \prod_{i=1}^{m} a_j^{i*} \quad \text{by (2.3).}$$

Note that the function $\Phi(x_1, \cdots, x_m) = \log(\sum_{i=1}^{m} x_i)$ is a negative set function in each pair of variables for $x_i > 0$, $i = 1, \cdots, m$. Remark 2 of Section 2 implies

$$\prod_{j=1}^{n} \sum_{i=1}^{m} a_j^i \geqq \prod_{j=1}^{n} \sum_{i=1}^{m} a_j^{i*} .$$

2. Minc [10].

The function $\Phi(x_1, \cdots, x_m) = \min_{1 \leqq i \leqq m} x_i$, $x_i > 0$ is both TP_2 and a positive set function in each pair of variables. Remark 2 and the inequality of Lorentz (2.3) imply, for $\boldsymbol{a}^i$, $i = 1, \cdots, m$ n-tuples of positive numbers

$$\prod_{j=1}^{n} \min_{1 \leqq i \leqq m} a_j^i \leqq \prod_{j=1}^{n} \min_{1 \leqq i \leqq m} a_j^{i*}$$

and

$$\sum_{j=1}^{n} \min_{1 \leqq i \leqq m} a_j^i \leqq \sum_{j=1}^{n} \min_{1 \leqq i \leqq m} a_j^{i*}$$

which are Theorems 3 and 5 in [10].

3. London [8].

For $a > 0$, $b > 0$ let $\Phi(a, b) = F(\log(1 + b/a))$ where $F(x)$ is convex and increasing for $x \geqq 0$.

Then $\partial^2 \Phi / \partial a \partial b \leqq 0$ which implies using Remark 1 and London's notation $F(x) = f(e^x)$

$$\sum_{i=1}^{n} f\left(1 + \frac{\tilde{b}_i}{\tilde{a}_i}\right) \leqq \sum_{i=1}^{n} f\left(1 + \frac{\tilde{b}_i}{a_i}\right) \leqq \sum_{i=1}^{n} f\left(1 + \frac{\tilde{b}_i}{a_i^*}\right)$$

for $\boldsymbol{a}, \boldsymbol{b}$ n-tuples of positive numbers. This is Theorem 1 in [8].

4. The next example is similar to London's Theorem 2 [8]. Both are direct consequences of the inequality of Lorentz. Consider $\Phi(a, b) = f(a - b)$ where f is convex. Φ is a negative set function for $-\infty < a, b < \infty$ which implies (see Remark 1)

$$\sum_{i=1}^{n} f(a_i - b_i) \leqq \sum_{i=1}^{n} f(a_i^* - \tilde{b}_i).$$

In particular

$$\sum_{i=1}^{n} |a_i - b_i|^p \leqq \sum_{i=1}^{n} |a_i^* - \tilde{b}_i|^p \text{ for } p \geqq 1.$$

5. The following example generalizes a result of Abramovich [1]. Let $y = (y_1, \cdots, y_{2n})$ and consider the function of $2n$ variables Ψ defined by

$$\Psi(y) = \sum_{i=1}^{n} \Phi(y_{2i-1}, y_{2i}).$$

An easy induction argument shows that if $\Phi(a,b)$ is symmetric then Ψ satisfies

$$\Psi(y) \leqq \Psi(y^*) \text{ for all } y \in R^{2n}$$

if and only if Φ is a positive set function. Setting $y_{2i-1} = a_i$, $y_{2i} = b_i$, $i = 1, \cdots, n$ we can rewrite the above inequality in the form

$$\sum_{i=1}^{n} \Phi(a_i, b_i) \leqq \sum_{i=1}^{n} \Phi(a_i^*, b_i^*) \leqq \sum_{i=1}^{n} \Phi(y_{2i-1}^*, y_{2i}^*)$$

where on the left we have applied Lorentz's inequality. On the right the vector y^* is a decreasing rearrangement of y whereas on the left a and b are rearranged separately. Invoking Remark 2 of Section 2, this implies that the function $\tilde{\psi}$ defined by

$$\tilde{\psi}(y) = \prod_{i=1}^{n} F(y_{2i-1}, y_{2i})$$

where F is positive and symmetric satisfies

$$\tilde{\psi}(y) \leqq \tilde{\psi}(y^*) \text{ for all } y \in R^{2n}$$

if and only if F is TP_2. In particular if we choose

$$F(x,y) = (xy)^m + a_1(xy)^{m-1} + \cdots + a_m$$

$x, y \geqq 0$, $a_i \geqq 0$, $i = 1, \cdots, m$, then $F(x,y)$ is TP_2 (see Karlin [6, p. 101]). We obtain that

$$\prod_{i=1}^{n} [(y_{2i-1}\, y_{2i})^m + a_1(y_{2i-1}\, y_{2i})^{m-1} +, \cdots, + a_m]$$

attains its maximum when y is arranged in decreasing order. This analyses Theorem 2 in Abramovich [1].

ACKNOWLEDGMENT

I wish to express my great indebtedness to Professor S. Karlin for his generous help and constant encouragement throughout this work.

REFERENCES

1. S. Abramovich, *The increase of sums and products dependent on* $(y_1,\ldots,y_n)$ *by rearrangement of this set*, Israel J. Math. **5** (1967), 177-181.

2. P. W. Day, *Rearrangement inequalities*, to appear.

3. K. Fan and G. G. Lorentz, *An integral inequality*, Amer. Math. Monthly **61** (1954), 626-631.

4. S. S. Gupta and U. Nagel, *On selection and ranking procedures and order statistics from the multinomial distribution*, Sankhya Ser. B **29** (1967), 1-34.

5. G. H. Hardy, J. E. Littlewood and G. Pólya, *Inequalities*, Cambridge University Press, 1959.

6. S. Karlin, *Total Positivity, Vol. I*, Stanford University Press, 1968.

7. S. Karlin and W. Studden, *Tchebycheff Systems with Application in Analysis and Statistics*, Interscience, New York, 1966.

8. D. London, *Rearrangement inequalities involving convex functions*, Pacific J. Math. **34** (1970), 749-753.

9. G. G. Lorentz, *An inequality for rearrangement*, Amer. Math. Monthly **60** (1953), 176-179.

10. H. Minc, *Rearrangements*, Trans. Amer. Math. Soc. **159** (1971), 69-72.

11. D. S. Mitrinovic, *Analytic Inequalities*, Springer-Verlag, Berlin-Heidelberg-New York, 1970.

12. I. Olkin, *Monotonicity properties of Dirichlet integrals with application to the multinominal distribution and the analysis of variance*, Biometrika **59** (1972), 303-307.

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

14. H. D. Ruderman, *Two new inequalities*, Amer. Math. Monthly **59** (1952), 29-32.

THE WEIZMANN INSTITUTE OF SCIENCE
REHOVOT, ISRAEL