Sankhyā : The Indian Journal of Statistics
1988, Volume 50, Series A, Pt. 2, pp. 199-204

MULTIVARIATE MAJORIZATION AND DIRECTIONAL MAJORIZATION ; POSITIVE RESULTS

By SUBIR KUMAR BHANDARI
Indian Statistical Institute

SUMMARY. General relations between directional majorization and multivariate majorization are studied, and sufficient conditions for directional majorization to imply multivariate majorization are obtained.

1. Introduction

For two matrices $X^{(m\times n)}$ and $Y^{(m\times n)}$, Marshall and Olkin (1979) defined X to be majorized by Y, written as $X \prec Y$, if $X = YP$, for some $n\times n$ doubly stochastic matrix P. Following Marshall and Olkin (1979, p. 433), we define X to be directionally majorized by Y, written as $X \prec_d Y$, if $aX \prec aY$ for all $a \in R^m$. Marshall and Olkin (p. 433) posed the open question whether $X \prec_d Y$ implies $X \prec Y$. A more general problem stated in Marshall and Olkin is whether $AX \prec AY$ for all $A : k\times m$ (for fixed k) implies $X \prec Y$. In this paper, sufficient conditions are given under which directional majorization implies multivariate majorization.

It will be reported in a subsequent communication that the above implication is not true under some specified conditions.

2. Main results

Theorem 2.1 : *For a fixed Y, $X^{(2\times n)} \prec_d Y^{(2\times n)}$ implies $X \prec Y$ for all $X^{(2\times n)}$, if all the column vectors of Y (in R^2) are boundary points in the convex hull of the column vectors of Y, and this convex null has 2-dimensional positive volume.*

Theorem 2.2 : *Suppose every column vector of $Y : m\times n$ is an extreme point in the convex hull generated by the columns of Y, which has r-dimensional positive volume, and at least $(n-r+2)$ of these column vectors are co-planner. Then $X \prec_d Y$ implies $X \prec Y$ for all X. Moreover, $AX \prec AY$ for all A : $k\times m$ implies $X \prec Y$.*

AMS (1980) *subject classification :* 15A39, 15A99.

Key words and phrases : Majorization, directional majorization, multivariate majorization, convex hull, hyperplane.

3. Proof of the results

Definition 3.1 : A function $f: R^m \to R$ is said to be *directional convex function*, if it is of the form $f(x) = g(\alpha . x)$, for fixed $\alpha \epsilon R^m$ and $g: R \to R$ convex.

Note that directional convex functions are convex functions.

Lemma 3.1 : *For* $X(m \times n) = (x_1^c, \ldots, x_n^c)$, $Y(m \times n) = (y_1^c, \ldots, y_n^c)$, $X \prec_d Y$ *if and only if,*

$$\sum_{i=1}^{n} F(x_i^c) \leqslant \sum_{i=1}^{n} F(y_i^c),$$

for all functions F which are sums of finitely many directional convex functions.

Proof : First note that for $x = (x_1, \ldots, x_n)$, $y = (y_1, \ldots, y_n)$, $x \prec y$ iff

$$\sum_{i=1}^{n} g(x_i) \leqslant \sum_{i=1}^{n} g(y_i)$$

for all convex functions $g: R \to R$ (see Marshall and Olkin, 1979, p. 108 or Hardy, Littlewood and Polya, 1934).

Now for $X: m \times n$, $Y: m \times n$,

$X \prec_d Y$

$\Leftrightarrow \alpha X \prec \alpha Y$, for all $\alpha \epsilon R^m$

$\Leftrightarrow (\alpha x_1^c, \ldots, \alpha x_n^c) \prec (\alpha y_1^c, \ldots, \alpha y_n^c)$, for all $\alpha \epsilon R^m$

$\Leftrightarrow \sum_{i=1}^{n} g(\alpha x_i^c) \leqslant \sum_{i=1}^{n} g(\alpha y_i^c)$, for all $\alpha \epsilon R^m$ and all convex functions $g: R \to R$.

$\Leftrightarrow \sum_{i=1}^{n} F(x_i^c) \leqslant \sum_{i=1}^{n} F(y_i^c)$, for all directional convex functions F.

$\Leftrightarrow \sum_{i=1}^{n} F(x_i^c) \leqslant \sum_{i=1}^{n} F(y_i^c)$, for all F which are sums of finitely many directional convex functions.

Definition 3.2 : For $a, b \epsilon R$, a line L in R^2 (having equation $l(x) = 0$ for $x \epsilon R^2$) for a point $Z \epsilon R^2$ with $Z \notin L$, define $C_{L,a,b,Z}: R^2 \to R$ by

$$C_{L,a,b,Z}(x) = a.d(L, x), \text{ if } l(x).\, l(Z) \geqslant 0$$

$$= b.d(L, x), \text{ if } l(x).\, l(Z) < 0,$$

where for $A \subseteq R^2$, $p \epsilon R^2$

$$d(A, p) = \inf \{\|q - p\| : q \epsilon A.\}$$

Clearly $C_{L,a,b,Z}$ is a directional convex function for $a \geqslant 0$, $b \geqslant 0$.

Lemma 3.2 : *For* $m = 2$, $X \prec_d Y$ *implies that the columnvectors of* X *are in the convex hull of the column vectors of* Y.

Proof: Let C denote the convex hull of the column vectors of Y. Suppose that for some i, the ith column vector x_i^c of X is not in C. As C is closed, there exists a line L which separates x_i^c from C and does not contain x_i^c. Now consider the directional convex function $\varphi = C_{L,1,0,x_i^c}$. Note that

$$\sum_{j=1}^{n} \varphi(y_j^c) = 0 < \sum_{j=1}^{n} \varphi(x_j^c),$$

since $\varphi(x_i^c) > 0$. This contradicts Lemma 3.1. Hence x_i^c is in C.

Proof of Theorem 2.1 : First note that $X : (m \times n) \prec Y : (m \times n)$ iff

$$\sum_{i=1}^{n} \varphi(x_i^c) \leqslant \sum_{i=1}^{n} \varphi(y_i^c), \qquad \dots \quad (3.1)$$

for all convex functions $\varphi : R^m \to R$ (See Fischer and Holbrook, 1977, p. 564 or Blackwell, 1953). Hence it is sufficient to show the above inequality for our case $m = 2$.

Let the polygon $C \subseteq R^2$ denote the convex hull of the column vectors of Y which are assumed to be distinct. Suppose y_i^c's are the n vertices of C, i.e. y_i^c's are the extreme points. We name these vertices by $A_1, A_2, \dots, A_n$ in consecutive order. By Lemma 3.2, all x_i^c's are in C. Consider a convex function φ on C. Define $\alpha_i = \varphi(A_i)$, $\alpha = (\alpha_1, \dots, \alpha_n)$, and

$$F_\alpha = \sup \{f : f \text{ convex on } C, f(A_i) = \alpha_i \text{ for all } i\}. \qquad \dots \quad (3.2)$$

In view of Lemmas 3.1 and 3.2 it is sufficient to show that F is the sum of finitely many directional convex functions, since

$$\varphi(x_i^c) \leqslant F_\alpha(x_i^c), \; \varphi(y_i^c) = F_\alpha(y_i^c) \text{ for all } i. \qquad \dots \quad (3.3)$$

We can assume $\alpha_1 = \alpha_2 = 0$, since otherwise we can make $\alpha_1 = \alpha_2 = 0$ by adding a suitable affine function to φ. We can furthermore assume that $\alpha_i \geqslant 0$ for all $i > 2$, since this can be achieved by adding the affine function $C_{L,s,-s,A_2}$ for suitable large $s > 0$, where L is the line joining the distinct points A_1 and A_2 and $i > 2$. Note that $C_{L,a,b,z}$ is affine if $a = -b$.

Consider $C_{L,t,0,A_2}$ for $t \geqslant 0$ and note that this function is affine on C. For $t = 0$

$$C_{L,t,0,A_2}(A_i) \leqslant \alpha_i, \text{ for all } i > 2. \qquad \dots \quad (3.4)$$

Now as we increase t, at some point (say at $t = t_0$) at least one equality in (3.4) will be attained preserving the other inequalities. Let $A_{i_1}, A_{i_2}, \dots, A_{i_r}$

$(r \leqslant n-2)$ be the vertices at which the equality in (3.4) is attained. If $r = n-2$, define $F_\alpha = C_{L, t_0, 0, A_3}$ and we are done.

If $r < n-2$, consider the following possible configuration :

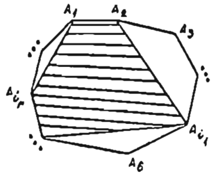

Fig. 1

Let $\beta_i = \alpha_i - C_{L, t_0, 0, A_3}(A_i)$. Then

$$\beta_i \geqslant 0, \; \beta_1 = \beta_2 = \beta_{i_1} = \dots = \beta_{i_r} = 0. \qquad \dots (3.5)$$

Consider now the polygon $A_1 A_2 A_{i_1} A_{i_2} \dots A_{i_r}$. Note that $C - A_1 A_2 A_{i_1} \dots A_{i_r}$ is the union of disjoint polygons. Because of (3.5) we can apply the above operation on each of those polygons, taking the two initial vertices to that β_i's are zero on them.

Ultimately adding these $C_{L, a, b, z}$ functions obtained at each stage from each of those polygons we get a function which is F_α ; this follows from the fact that for each point in the polygon $A_1 A_2 \dots A_n$, there exists a sub-polygon with vertices in $\{A_1, A_2, \dots, A_n\}$ on which the derived function is affine.

This construction shows that F_α, as derived above, is the sum of finitely many directional convex functions. This proves Theorem 2.1 when all the column vectors of Y are extreme points.

Now suppose the vertices of C are $V_1, V_2, \dots, V_k$, arranged in consecutive order, and B is a column vector of Y which lies on the segment $V_1 V_2$ closest to V_1. Then we shall follow the above initial operation with $A_1 = V_1$ and $A_2 = B$. By making $\alpha_1 = \alpha_2 = 0$, we can ensure that α at all column vectors lying on $V_1 V_2$ is $\geqslant 0$. The above proof can now be followed stage by stage.

When the column vectors of Y are not distinct, the above operation is used only on distinct column vectors of Y ; the desired result then follows from (3.3).

Lemma 3.3 : *Let the convex hull of the column vectors of* $Y : m \times n$ *have* r*-dimensional positive volume,* $r < m$*. Then the problem of equivalence of* $X \prec_d Y$ *and* $X \prec Y$ *reduced to the corresponding problem in* r*-dimension.*

Proof : For some nonsingular $A : m \times m$ and suitable $b : m \times 1$ and $Y_1 : r \times n$, we have

$$A[Y+(b, \ldots, b)] = \begin{pmatrix} Y_1 \\ 0 \end{pmatrix}.$$

Following the line of proof of Lemma 3.2, we can show that $X \prec_d Y$ implies that every column vector of X is in the convex hull of the column vectors of Y. Thus

$$A[X+(b, \ldots, b)] = \begin{pmatrix} X_1 \\ 0 \end{pmatrix},$$

for some X_1. It can be shown now that $X \prec_d Y \Leftrightarrow X_1 \prec_d Y_1$ and $X \prec Y \Leftrightarrow X_1 \prec Y_1$.

Proof of Theorem 2.2 : In view of Lemma 3.3 we may assume, without any loss of generality, that $r = m$.

Hence our assumption entails that at least $(n-m+2)$ of the Y_i's are co-plannor, i.e. they belong to a 2-dimensional affine space of R^m. Let those vectors be represented by the points $A_1, A_2, \ldots, A_{n-m+2}$ and their convex hull be a polygon denoted by $A_1 A_2 \ldots A_{n-m+2}$, written in consecutive order.

The convex hull of A_1, A_2 and the $(m-2)$ column vectors of Y outside the above plane has $(m-1)-$dimensional positive volume; let this convex hull be contained in a hyperplane H.

Note that H does not contain the polygon $A_1 A_2 \ldots A_{n-m+2}$. Since A_1 and A_2 are in H, the other A_i's $(i = 3, 4, \ldots, n-m+2)$ are on one side of H.

Following Definition 3.2, define

$$C_{H,t,0,A_3}(x) = \begin{cases} t d(H, x), & \text{if } l(x)\, l(A_3) \geqslant 0 \\ 0 & \text{if } l(x)\, l(A_3) < 0 \end{cases}$$

where $t \geqslant 0$ and $l(x) = 0$ is the equation of H.

To complete the proof we follow the operations employed in the proof of Theorem 2.1 with a hyperplane H taking the role of the line defining the C-function. Note that initially we can make α_i to be zero at A_1, A_2 and $(m-2)$ points lying outside the plane by adding a suitable affine function to φ.

Acknowledgement. The author is grateful to Professor Somesh Das Gupta for his help.

REFERENCES

BHANDARI, S. K. (1984): Multivariate majorization and directional majorization. *Tech. Report* No. 20/84, Stat-Math. Division, Indian Statistical Institute.

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

FISCHER, P. and HOLBROOK, J. A. R. (1977): Matrices doubly stochastic by blocks. *Can. J. Math.* **29**, 559-577.

HARDY, G, H., LITTLEWOOD, J. E. and POLYA, G. (1934): *Inequalities*, Cambridge University Press.

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

ROCKAFELLAR, R. T. (1970): *Convex Analysis*, Princeton Univ. Press, Princeton.

SHERMAN, S. (1951): On a theorem of Hardy, Littlewood, Polya and Blackwell. *Proc. Nat. Acad. Sci.* **37**, 826-831.

Paper received : September, 1984.

Revised : December, 1986.