CHAPTER - II

MAJORIZATION AND RELATED TOPICS

Here we quote certain results which would form the foundation for the study of majorization. These include

A. Basic Notations

B. Weak Majorization

C. Doubly stochastic matrices and permutation matrices

D. Characterization of majorization using douby stochastic matrices

E. Schur-concavity and Schur convexity

F. Operations preserving majorization

2.A NOTATIONS

$R = (-\infty, \infty)$

$R_+ = [0, \infty]$

$R^n = \{(x_1, \ldots, x_n) \;;\; x_i \in R \text{ for all } i\}$

$R^n_+ = \{(x_1, \ldots, x_n) \;;\; x_i \geq 0 \text{ for all } i\}$

$D = \{(x_1, \ldots, x_n) \;;\; x_1 \geq \ldots \geq x_n\}$

$D_+ = \{(x_1, \ldots, x_n) \;;\; x_1 \geq \ldots \geq x_n \geq 0\}$

$x_{[i]}$ = i-th component of vector $x \in R^n$ when arranged in decreasing order.

$x_{(i)}$ = i-th component of vector $x \in R^n$ when arranged in increasing order

2.B WEAK MAJORIZATION

Let x and y be two vectors from R^n. x is said to be weakly submajorized by y (denoted as $x <_w y$) if

$$\sum_{i=1}^{k} x_{[i]} \leq \sum_{i=1}^{k} y_{[i]}; \qquad k = 1, \ldots, n.$$

Note that in this definition the second condition of majorization is replaced by $\leq$ constraint.

x is said to be weakly supermajorized by y (denoted as $x <^w y$) if

$$\sum_{i=1}^{k} x_{(i)} \geq \sum_{i=1}^{k} y_{(i)}; \qquad k = 1, \ldots, n.$$

In either case x is said to be weakly majorized by y.

2.B.1 Examples :

Let $x = (5, 3, 1)$ $\quad y = (7, 6, 1)$

$$x_{[1]} = 8 < y_{[1]} = 7$$

$$x_{[1]} + x_{[2]} = 8 < y_{[1]} + y_{[2]} = 13$$

$$x_{[1]} + x_{[2]} + x_{[3]} = 9 < y_{[1]} + y_{[2]} + y_{[3]} = 14$$

Hence $x <_w y$.

Let $x = (1, 3, 5)$ $\quad y = (0, 3, 4)$

$$x_{(1)} = 1 > y_{(1)} = 0$$

$$x_{(1)} + x_{(2)} = 4 > y_{(1)} + y_{(2)} = 3$$

$$x_{(1)} + x_{(2)} + x_{(3)} = 9 > y_{(1)} + y_{(2)} + y_{(3)} = 7$$

Hence $x \prec^{w} y$.

Weakly majorized from below and weakly majorized from above are two alternative terms for weakly submajorized and weakly super majorized respectively.

2.C DOUBLY STOCHASTIC MATRICES AND PERMUTATION MATRICES

An important result in the study of majorization is a theorem due to Hardy, Littlewood and Polya (1929) [10] which says that for $x, y \in R^n$;

$x \prec y$ if and only if $x = y P$ where P is a doubly stochastic matrix. Hence a brief account on doubly stochastic matrix.

2.C.1 Definition : An $n \times n$ matrix

$P = (P_{ij})$ is doubly stochastic if

$$P_{ij} \geq 0 \qquad \text{for all } i, j = 1, \ldots, n.$$

and

$$\sum_i P_{ij} = 1 ; \qquad \text{for all } j = 1, \ldots, n.$$

$$\sum_j P_{ij} = 1 ; \qquad \text{for all } i = 1, \ldots, n.$$

2.C.2 Examples :

(i) $$P = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1/3 & 2/3 \\ 0 & 2/3 & 1/3 \end{bmatrix}$$

(ii) $$P = \begin{bmatrix} 1/3 & 1/3 & 1/3 \\ 1/3 & 1/3 & 1/3 \\ 1/3 & 1/3 & 1/3 \end{bmatrix}$$

2.C.3 Definition : A square matrix Π is said to be permutation matrix if each row and each coloumn has a single unit and all other entries are zero.

2.C.4 Example :

$$\Pi = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 0 & 1 \\ 0 & 1 & 0 \end{bmatrix}$$ is a permutation matrix

2.C.5 Remarks : There are $n!$ such matrices of size n each of which is obtained by interchanging rows or coloumns of identity matrix.

A permutation matrix is a stochastic matrix as each of the row or coloumn sums are equal to one.

It is straight farward to varify that the set of $n \times n$ doubly stochastic matrices is convex and the permutation matrices are the extreme points of this set. Convex hull of permutation matrices coinsides with the set of doubly stochastic matrices.

2.C.7 Theorem (Birkhoff 1946) :

The permutation matrices constitute the extreme points of the set of doubly stochastic matrices. Moreover the set of doubly stochastic matrices is the convex hull of permutation matrices.

We omit the proof of this theom as it is beyond the scope of this study. However it could be found in [1].

<u>2.C.8 Theorem</u> :

An $n \times n$ matrix $P = (P_{ij})$ is doubly stochastic if and only if $yP \prec y$ for all $y \in R^n$.

<u>Proof</u> :

Assume that $yP \prec y$ for all $y \in R^n$.

Hence $eP \prec e$; where $e = (1,\ldots,1)$. But for some vector z, $z \prec e$ would mean $z = e$. (This is because all components of e are equal and there is no vector which has got components less scattered than itself.)

Hence $eP = e$.

i.e., $\sum_i P_{ij} = 1.$ for all j's (1)

Next take $y = e_i$ (i.e., $y_i = 1$, $y_j = 0$ if $j \neq i$)

We get $e_i P \prec e_i$

i.e., $(P_{i1}, P_{i2}, \ldots, P_{in}) \prec e_i$.

From the definition of majorization

we get $\sum_j P_{ij} = 1.$ for all i's (2)

i.e., $P e' = e'$

This also means that $P_{ij} \geq 0$ (3)

Since $a \prec b$ implies $\min_i a_i \geq \min_i b_i$.

From (1), (2) and (3) it follows that P is doubly stochastic

Suppose P is doubly stochastic, let $x = y P$, also suppose that $x_1 \geq \ldots, \geq x_n$; $y_1 \geq \ldots\ldots, \geq y_n$.

(Otherwise rewrite $x = y P$ as

$$x R = y Q Q^{-1} P R$$

where Q and R are permutation matrices chosen such that $y Q$ and $x R$ have decreasing components).

Then

$$\sum_{j=1}^{k} x_j = \sum_{j=1}^{k} \sum_{i=1}^{n} y_i P_{ij} = \sum_{i=1}^{n} y_i t_i,$$

where

$$0 \leq t_i = \sum_{j=1}^{k} P_{ij} \leq 1 \quad \text{and} \quad \sum_{i=1}^{n} t_i = k.$$

Thus

$$\sum_{j=1}^{k} x_j - \sum_{j=1}^{k} y_i = \sum_{i=1}^{n} y_i t_i - \sum_{i=1}^{k} y_i$$

$$= \sum_{i=1}^{n} y_i t_i - \sum_{i=1}^{k} y_i + y_k (k - \sum_{i=1}^{n} t_i)$$

$$= \sum_{i=1}^{k} (y_i - y_k)(t_i - 1) + \sum_{i=k+1}^{n} t_i (y_i - y_k)$$

$$\leq 0 \qquad (4)$$

Also $\sum_{i=1}^{n} x_i = y P e' = y e' = \sum_{i=1}^{n} y_i$ (5)

From (4) and (5) it follows that

$$y P \prec y$$

Hence the proof.

2.D CHARACTERIZATION OF MAJORIZATION USING DOUBLY STOCHASTIC MATRICCES

For the purpose of proving the theorem due to Hardy, Littlewood and Polya (1929) which states that $x \prec y$ if and only if $x = y P$ for some doubly stochastic matrix P, a preliminary Lemma is proved which is perhaps of greater importance.

2.D.1 T-transform : It is a special kind of linear transformation. The matrix of T-transform has the form

$$T = \lambda I + (1 - \lambda) Q$$

where $0 \leq \lambda \leq 1$ and Q is a permutation matrix that just interchanges two coordinates. Thus $x T$ has the form

$$x T = (x_1, \ldots, x_{j-1}, \lambda x_j + (1 - \lambda) x_k, x_{j+1}, \ldots$$
$$\ldots x_{k-1}, \lambda x_k + (1 - \lambda) x_j, x_{k+1}, \ldots, x_n).$$

2.D.2 Lemma (Muirhead, Hardy Littlewood and Polya)

If $x \prec y$ then x can be derived from y by successive applications of a finite number of T-transforms.

Proof : Since permutation matrices Q are T-transforms in case $\lambda = 0$ and since any permutation matrix is the product of such simple permutation matrices we assume that x is not obtainable from y by permuting arguments.

Also we assume without loss of generality that

$$x_1 \geq \ldots\ldots \geq x_n \; ; \quad y_1 \geq \ldots\ldots y_n$$

Let j be the largest index such that $x_j < y_j$ and let k be the smallest index greater than j such that $x_k > y_k$. Such a pair j, k must exist, since the largest index i for which $x_i \neq y_i$ must satisfy $x_i > y_i$, by choice of j and k

$$y_j > x_j \geq x_k > y_k \qquad (1)$$

Let $\quad d = \min(y_j - x_j,\ x_k - y_k)$

$$1 - \lambda = d \,/\, (y_j - y_k) \text{ and let}$$

$$y^* = (y_1, \ldots, y_{j-1},\ \ y_j - d,\ y_{j+1}, \ldots, y_{k-1},\ y_k + d,\ y_{k+1}, \ldots, y_n)$$

It follows from (1) that $0 < \lambda < 1$ and it is easy to varify that

$$y^* = \lambda\, y + (1 - \lambda)\, (y_1, \ldots, y_{j-1},\ y_k,\ \ y_{j+1}, \ldots, y_{k-1},\ y_j,\ y_{k+1}, \ldots, y_n)$$

Thus $y^* = y\,T$ for $T = \lambda\, I + (1 - \lambda)\, Q$ where Q interchanges the j-th and k-th coordinates.

Consequently $y^* < y$.

Also $x < y^*$ since

$$\sum_1^v y_i^* = \sum_1^v y_1 \geq \sum_1^v x_i \qquad v = 1, \ldots, j-1$$

$$y_j^* \geq x_j \,, \qquad y_i^* = y_i \qquad i = j+1, \ldots, k$$

$$\sum_1^v y_i^* = \sum_1^v y_i \geq \sum_1^v x_i \qquad v = k+1, \ldots, n.$$

$$\sum_1^n y_i^* = \sum_1^n y_i = \sum_1^n x_i$$

Hence $x < y^*$

For any two vectors u, v let $b(u, v)$ be the number of non-zero differences $u_i - v_i$.

Since $y_j^* = x_j$ if $d = y_j - x_j$ and $y_k^* = x_k$ if $d = x_k - y_k$, it follows that

$$b(x, y^*) \leq b(x, y) - 1$$

Hence y can be derived from x by a finite number of T-transformations.

<u>2.D.3 Remark</u> : It can be observed from the above proof that if $x < y$ then x can be derived from y by successive applications of at most $(n - 1)$ T-transforms. This is because $b(u, v) \leq n$ and $b(u, v) \neq 1$ (otherwise $\Sigma u_i \neq \Sigma v_i$)

<u>2.D.4 Theorem</u> : (<u>Hardy, Littlewood and Polya (1929</u>))

A necessary and sufficient condition that $x < y$ is that there exist, a doubly stochastic matrix P such that $x = y P$.

<u>Proof</u> :

First assume that there exists a doubly stochastic matrix

P such that

$$x = y\,P$$

Then by 2.C.8 $\quad x \prec y$

Now assume that $\quad x \prec y$.

Since T-transforms are doubly stochastic, the product of T-transforms is doubly stochastic. Thus there exists a doubly stochastic matrix such that $x = y\,P$.

2.D.5 Example :

Let $\quad x = (3.5,\ 3,\ 3.5) \qquad y = (6,\ 3,\ 1)$

Obviously $\quad x \prec y$

Hence $\quad x = y\,P$ for some doubly stochastic matrix P.

On solving we get

$$P = \begin{bmatrix} .5 & 0 & .5 \\ 0 & 1 & 0 \\ .5 & 0 & .5 \end{bmatrix}$$

2.E SCHUR CONCAVE AND SCHUR CONVEX FUNCTIONS

For a given ordering on a set $\mathcal{X}$, the real valued function of which satisfy $f(x) \leq f(y)$ is referred to as order-preserving function. In 1923 I.Schur [2] studied the ordering on majorization, which made them known as Schur-concave or Schur-convex functions. Many of the inequalities that arise from majorization can be obtained by identifying an appropriate order-preserving function. Hence the importance of Schur-concave and Schur-convex functions.

2.E.1 Definition : A real valued function f defined on a set $A \subset R^n$ is said to be Schur-convex on A if $x \prec y$ on A implies $f(x) \le f(y)$.

2.E.2 Example :

Let $f(x) = \sqrt{\frac{1}{n} \Sigma (x_i - \bar{x})^2}$

Let $x = (1, 2, 3)$ $y = (0, 2, 4)$

Obviously $x \prec y$

$\bar{x} = 2$ $\bar{y} = 2$

$$f(x) = \sqrt{\frac{(1-2)^2 + 0 + (2-3)^2}{3}}$$

$$= \sqrt{\frac{2}{3}} = \sqrt{.66}$$

$$f(y) = \sqrt{\frac{(0-2)^2 + 0 + (4-2)^2}{3}}$$

$$= \sqrt{\frac{8}{3}} = \sqrt{2.66}$$

Hence $f(x) \le f(y)$

2.E.3 Remarks :

i) The example discussed above is the S.D. of a set of numbers, which is a measure of diversity. As majorization also characterizes diversity Schur-convexity reflects this property of the function.

ii) Marshall and Olkin [1] has suggested that Schur-increasing as an appropriate title for Schur-convex functions. But Schur-convex is by now well entrenched in the literature.

iii) If $f(x) < f(y)$ whenever $x \prec y$ but x is not a permutation of y, then f is said to be strictly Schur-convex on A.

<u>2.E.4 Definition</u> : f is said to be Schur-concave on A if $x \prec y$ on A implies $f(x) \geq f(y)$.

<u>2.E.5 Example</u> :

Let $$f(x) = \frac{1}{1 + \Sigma(x_i - \bar{x})^2}$$

It is easy to varify that $f(x)$ is Schur concave as $\Sigma(x_i - \bar{x})^2$ is Schur-convex function.

<u>2.E.6 Remarks</u> :

i) f is said to be strictly Schur-concave on A if strict inequality $f(x) > f(y)$ holds when x is not a permutation of y.

ii) It is obvious that f is Schur-concave if and only if $-f$ is Schur-convex.

iii) Also because of the ordering on R^n has the property that $x \prec x\Pi \prec x$ for all permutation matrices Π, it follows that f is Schur-convex or Schur-concave on a symmetric set A (i.e., a set of A with the property that $x \in A$ implies $x\Pi \in A$ for all permutations Π), then f is symmetric on A. (i.e., $f(x) = f(x\Pi)$ for all permutations Π).

2.F OPERATIONS PRESERVING MAJORIZATION

From available literature one can list out a number of operations that can generate a pair of vectors with some sort of majorization from

i) another pair in which majorization is already present

ii) a pair of vectors not necessarily having majorization as a property between them.

Here we list out some such operations. These theorem
s are
quoted from Marshall and Olkin [1] without proofs.

2.F.1 Theorem : For all convex functions g

$$x \prec y \implies (g(x_1),\dots,g(x_n)) \prec_w (g(y_1),\dots,g(y_n))$$

for all concave functions g

$$x \prec y \implies (g(x_1),\dots,g(x_n)) \prec^w (g(y_1),\dots,g(y_n))$$

2.F.2 Example :

Since $g(x) = x^2$ is a convex function

$$x \prec y \implies (x_1^2,\dots,x_n^2) \prec_w (y_1^2,\dots,y_n^2)$$

also as $g(x) = \sqrt{x}$ is a concave function

$$x \prec y \implies (\sqrt{x_1},\dots,\sqrt{x_n}) \prec^w (\sqrt{y_1},\dots,\sqrt{y_n})$$

2.F.3 Theorem :

i) For all increasing convex functions g

$$x <_w y \Longrightarrow (g(x_1),\ldots,g(x_n) <_w (g(y_1),\ldots,g(y_n))$$

ii) For all increasing concave functions g

$$x <^w y \Longrightarrow (g(x_1),\ldots,g(x_n) < (g(y_1),\ldots,g(y_n))$$

iii) For all decreasing convex functions g

$$x <^w y \Longrightarrow (g(x_1),\ldots,g(x_n) <_w (g(y_1),\ldots,g(y_n))$$

iv) For all decreasing concave functions g

$$x <_w y \Longrightarrow (g(x_1),\ldots,g(x_n) <^w (g(y_1),\ldots,g(y_n))$$

2.F.4 Example :

Examples quoted in 2.F.2 are increasing functions. Hence these are good enough for (i) and (ii)

Let $f(x) = -x^{1/2}$ which is a decreasing convex function. Hence it holds for (iii).

Let $f(x) = -x^2$ which is a decreasing concave function. Hence it holds for (iv).