CHAPTER - I

INTRODUCTION

Majorization is concerned with the comparison of degrees of diversity between two vectors. In order to understand the conceptual developments in this topic, it is felt that one ought to know the formal definition of Majorization. Hence we start this chapter with the definition of majorization. The chapter contains the following topics.

A. Definition of Majorization.

B. Examples.

C. Conceptual background and historical developments.

D. Geometrical aspects of majorization.

1.A DEFINITION :

For $x, y \in R^n$; x is said to be majorized by y (denoted as $x < y$) if

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

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

$x_{[i]}$'s represents the decreasing numerical order of x_i's.

1.B EXAMPLES :

Here we present some general examples on Majorization.

1.B.1 Let $a = (1, 3, 8)$

$b = (3, 4, 5)$

On re-arranging in descending order we get

$a_{[1]} = 8 \qquad a_{[2]} = 3 \qquad a_{[3]} = 1$

$b_{[1]} = 5 \qquad b_{[2]} = 4 \qquad b_{[3]} = 3$

$a_{[1]} = 8 > b_{[1]} = 5$

$a_{[1]} + a_{[2]} = 11 > b_{[1]} + b_{[2]} = 9$

$a_{[1]} + a_{[2]}\ a_{[3]} = 12 = b_{[1]} + b_{[2]} + b_{[3]} = 12$

Hence $a > b$.

1.B.2 Let $x = (8, 5, 2) \qquad y = (7, 7, 1)$

be two vectors which are already in descending numerical order.

$x_{[1]} = 8 > y_{[2]} = 7$

$x_{[1]} + x_{[2]} = 13 < y_{[1]} + y_{[2]} = 14$

$x_{[1]} + x_{[2]}\ x_{[3]} = 15 = y_{[1]} + y_{[2]} + y_{[3]} = 15$

Hence $a \not> b$ or $b \not> a$

This example illustrates that even if the sum of the components of two vectors are equal it is not essential that one would majorize the other.

1.B.3 $(\frac{1}{n}, \ldots\ldots, \frac{1}{n}) < (\frac{1}{n-1}, \ldots\ldots, \frac{1}{n-1}, 0) < \ldots\ldots$

$\ldots < (\frac{1}{2}, \frac{1}{2}, 0, \ldots 0) < (1, 0, \ldots, 0)$

I.B.4 In general

$$(\underbrace{\alpha c, \ldots, \alpha c}_{m}, 0, \ldots, 0) < (\underbrace{c, \ldots, c}_{\ell}, 0, \ldots, 0)$$

where $m \geq \ell$; $\ell c = m \alpha c$; $\alpha = \frac{\ell}{m} \leq 1$.

1.B.5 $$\left(\frac{1}{n}, \ldots, \frac{1}{n}\right) < (a_1, \ldots, a_n) < (1, 0, \ldots, 0)$$

where $a_i \geq 0$; $\Sigma a_i = 1$.

1.B.6 $$(x_1 + c, \ldots, x_n + c) / (2x_i + nc)$$

$$< (x_1, \ldots, x_n) / (\Sigma x_i)$$

1.C CONCEPTUAL BACKGROUND AND HISTORICAL DEVELOPMENTS

According to Marshall and Olkin [1] the origin of the concept of majorization can be traced into

- Extension of Inequalities
- Mathematical Origins
- Economics

1.C.1 Extension of Inequalities

Consider the function

$$f(x_1, x_2) = x_1^2 + x_2^2$$

It is easy to varify that

$$f(\overline{x}, \overline{x}) \leq f(x_1, x_2)$$

where $$\overline{x} = \frac{x_1 + x_2}{2}$$

It is natural that one may aspire for more general comparisions like

$$\varphi(x_1,\ldots,x_n) \le \varphi(y_1,\ldots,y_n)$$

where $x_1,\ldots,x_n$ is less spread out than that of $y_1,\ldots,y_n$ and φ is a convex function

In 1929 Hardy, Littlewood and Polya [10] proved a similar result when they were searching for conditions on

$x_1,\ldots,x_n$ and $y_1,\ldots,y_n$ so that

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

for all convex function g.

They found that a necessary and sufficient condition for (1) to be true is that x should be majorized by y.

1.C.2 Mathematical Origin

In 1923 Schur [2] used the concept of majorization as a preliminary to proving Hadamard's determinant inequality. Schur showed that the diagonal elements $a_1,\ldots,a_n$ of a positive semi definite Hermitian Matrix is majorized by their characteristic roots $\lambda_1,\ldots,\lambda_n$.

i.e., $$(a_1,\ldots,a_n) \prec (\lambda_1,\ldots,\lambda_n) \qquad (2)$$

This is illustrated in an example given below

Let $$A = \begin{bmatrix} 5 & 2-3i \\ 2+3i & 3 \end{bmatrix}$$

It is obvious that A is a positive definite Hermitiam matrix.

Let $|A - x I| = 0$ be the characteristic equation

i.e., $$\begin{bmatrix} 5-x & 2-3i \\ 2+3i & 3-x \end{bmatrix} = 0$$

i.e., $$(5-x)(3-x) - (2-3i)(2+3i) = 0$$

i.e., $$x^2 - 8x + 2 = 0$$

On solving this equation we would get

$$\lambda_1 = 7.74 \qquad \lambda_2 = .26.$$

The diagonal elements of A are $(5,3)$ call them (a_1, a_2)

$$a_{[1]} = 5 \quad < \quad \lambda_{[1]} = 7.74$$

$$a_{[1]} + a_{[2]} = 8 \quad = \quad \lambda_{[1]} + \lambda_{[2]} = 8$$

Hence $(a_1, a_2) < (\lambda_1, \lambda_2)$.

In 1954 Horn [3] gave a new interpretation to Schur's result. By identifying all functions φ which satisfy the relation $x < y$ implies $\varphi(x) \leq \varphi(y)$ (whenever $x, y \in R^n_+$) Schur identified all possible inequalities for a positive semidefinite Hermitian matrix. The comparison is between the functional values of the diagonal elements with the same functional values of the characteristic roots. There are other inequalities in mathematics which could be characterised through majorization.

1.C.3 Studies in Economics

Economists were interested in finding a measure to characterize the inequalities in distribution of wealth or income. In 1905 Lorenz [4] introduced what is known as Lorenz curve.

Lorenz curve of distribution of income (wealth) is the graph of the fraction of total income possessed by the lowest p-th fraction of the population as a function of P. $(0 \leq P \leq 1)$ (fig.1.)

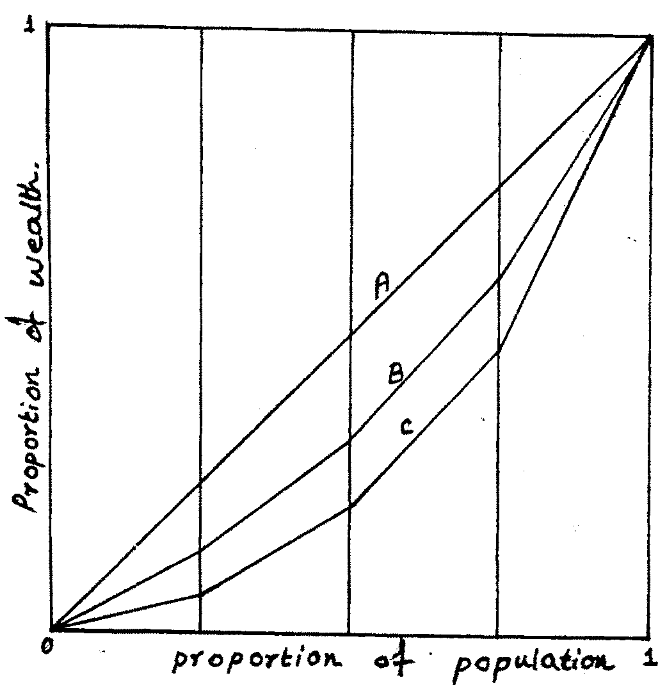


fig.1.

Consider the wealth of n individuals x_i; $i = 1,\ldots,n$. Plot the points $(k/n, s_k/s_n)$, $k = 0,\ldots,n$ where $s_0 = 0$

and $s_k = \sum_{i=1}^{k} x_{(i)}$ is the wealth of the poorest k individuals in the population. Join these points by line segments to obtain a curve connecting the origin with $(1,1)$. If the total wealth is uniformly distributed we are bound to get a straight line. If not it would be a convex curve. This is illustrated in fig.1. A represents a uniform distribution but B is more bent in the middle. Which shows an uneven distribution; whereas C is further bent in the middle and is most uneven among the three distributions. Let x_i; $i = 1,\ldots,n$ be the distribution of total wealth T according to curve A. Let y_i; $i = 1,\ldots,n$ be the distribution of total wealth T according to curve B. From the graph it is evident that

$$\sum_{i=1}^{t} x_{(i)} \geq \sum_{i=1}^{t} y_{(i)}; \qquad t = 1,\cdots,n-1$$

$$\sum_{i=1}^{n} x_i = \sum_{i=1}^{n} y_i = T$$

This implies that x is majorized by y. (Of course the arrangement is in increasing order eventhough the definition of majorization demands an arrangement in decreasing order).

In 1912 Pigou [5] introduced the concept of principle of transfers. This was illustrated by Dalton [6] in 1920 through income distribution. If there are two income-receivers and a transfer takes place from the richer to the poorer; the inequality is diminished. This may continue till both of them receives the average income; which makes the inequality vanish. He could further observe that if y_k is the income of the individual k; $k = 1,\dots,n$ and if an amount of income Δ is to be transferred from individual j to i then the inequality is diminished provided $\Delta \le y_j - y_i$; $y_j \ge y_i$.

In 1903 Muirhead [7] discussed this concept of transfer in his paper generalizing the arithmetic-geometric mean inequality. He proves that if the components of two vectors x and y are non-negative integers then the following conditions are equivalent.

(i) x can be derived from y by a finite number of transfers (each satisfying Dalton's ristriction).

(ii) The sum of k largest components of x is less than or equal to the sum of k largest components of y; $k = 1,2,\dots,n$ with equality when $k = n$.

The second condition is as good as that of the formal definition of majorization quoted in 1.A.

1.D GEOMETRICAL ASPECTS OF MAJORIZATION

Let $(y_1,\dots,y_n)$ be the income of n individuals. According to Hardy, Littlewood and Polya [8] repeated averages of two incomes at a time can produce the same result as the replacement of y_i by an arbitary average of the form

$$x_i = y_1 p_{1j} + \dots\dots + y_n p_{nj}; \quad j = 1,\dots,n$$

where $p_{ij} \geq 0$ for all i and j.

$$\sum_{i=1}^{n} p_{ij} = 1 \quad \text{for all } j$$

and

$$\sum_{j=1}^{n} p_{ij} = 1 \quad \text{for all } i$$

This could be written as

$$x = y\,p.$$

Where p is doubly stochastic.

This could be better illustrated through an example.

Let $y = (10,\ 5,\ 3)$ be the income of 3 individuals.

By taking repeated averages two times we would get a vector (7, 7, 4). Call it x. According to Hardy, Littlewood and Polya we should be able to find a dcubly stochastic matrix P such that $x = y\,P$.

i.e.,

$$(7,7,4) = (10,5,3)\begin{bmatrix} P_{11} & P_{12} & P_{13} \\ P_{21} & P_{22} & P_{23} \\ P_{31} & P_{32} & P_{33} \end{bmatrix}$$

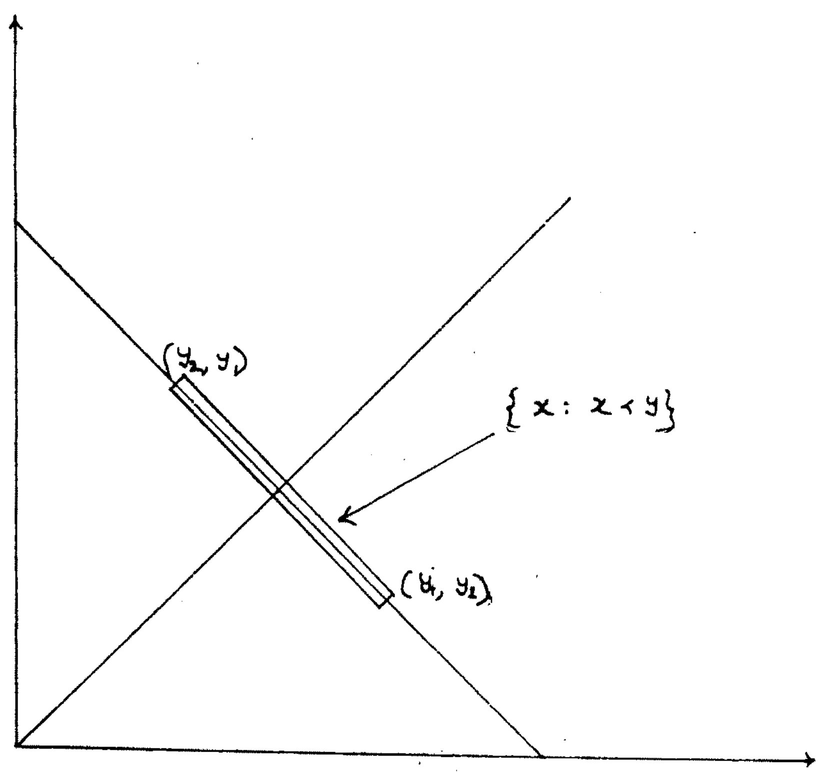
Now using the fact that each row-sums as well as column-sums should be equal to one; we would reduce our task to finding four unknowns rather than nine. This would result in solving for four unknowns from three equation. Hence the doubly stochastic matrix P need not be unique.

$$P = \begin{bmatrix} 1/2 & 3/7 & 1/14 \\ 1/4 & 1/2 & 1/4 \\ 1/4 & 1/4 & 19/28 \end{bmatrix}$$ is one such solution.

Birkhoff's theorem [9] states that doubly stochastic matrices constitute the convex hall of permutation matrices.

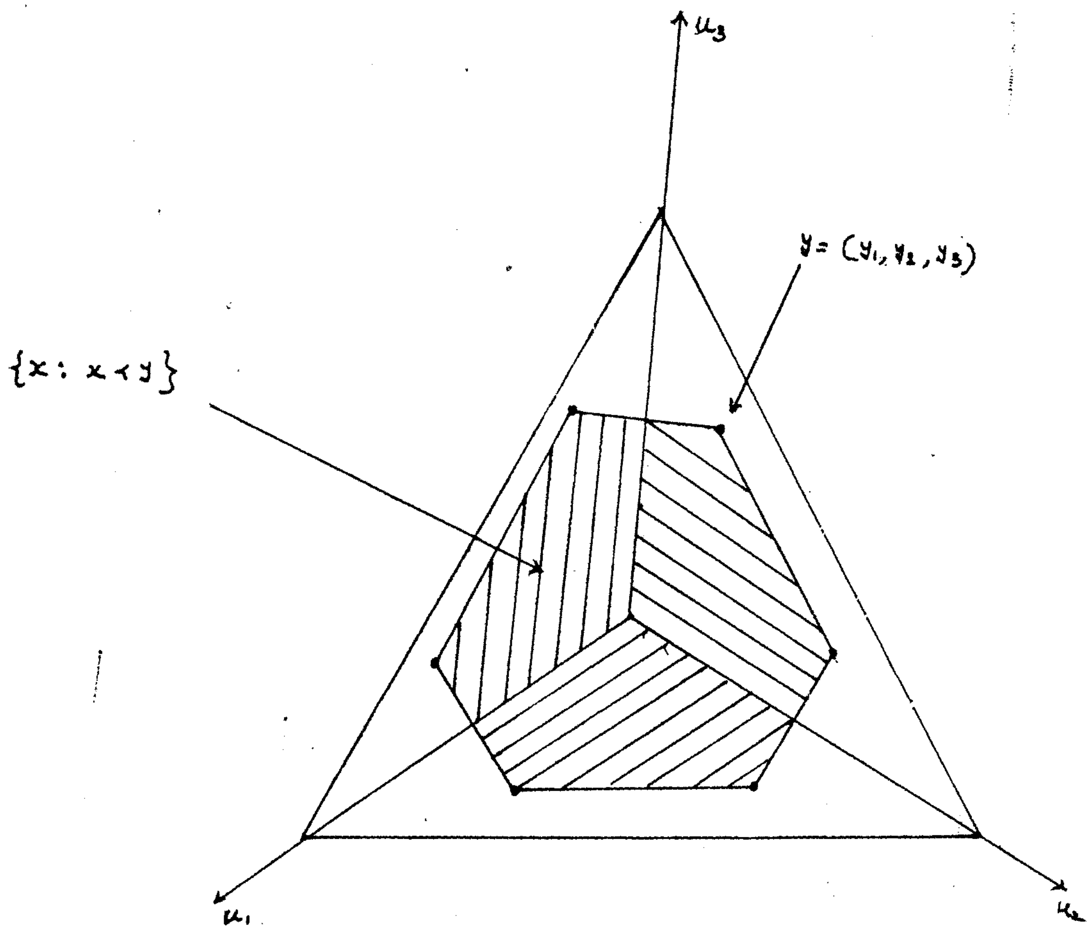
Thus if $x \prec y$ so that $x = y P$ for some doubly stochastic matrix P, then there exists constants $a_i \geq 0$; $\Sigma a_i = 1$ such that

$x = y (\Sigma a_i \Pi_i) = \Sigma a_i (y \Pi_i)$ where Π_i's are permutation matrices. This means that x is in the convex hall of the orbit of y under the group of permutation matrices. (As shown in fig.2a and 2b).



Orbit of y under permutations and the set $\{x : x \prec y\}$

fig 2. a



Orbit of y under permutations and the set $\{x : x \prec y\}$

fig 2.6