To Prof. Jayanta K. Ghosh
with best regards,
Subir
19. 8. 87

SOME ASPECTS OF MAJORIZATION
AND
THEIR APPLICATIONS IN STATISTICS

By

SUBIR KUMAR BHANDARI

INDIAN STATISTICAL INSTITUTE
CALCUTTA
1987

SOME ASPECTS OF MAJORIZATION
AND
THEIR APPLICATIONS IN STATISTICS

By

SUBIR KUMAR BHANDARI

Stat-Math. Division
INDIAN STATISTICAL INSTITUTE
CALCUTTA

1987

[Thesis submitted to the Indian Statistical Institute in partial fulfilment of the requirements for the award of the degree of Doctor of Philosophy.]

Dedicated to

TUTUL
and
ANI

ACKNOWLEDGEMENT

I gratefully acknowledge my indebtedness to Professor Somesh DasGupta under whose supervision and constant encouragement this thesis was written. The results in this thesis were developed in a very hard time of my life. Without the sincere care and suggestions of Professor DasGupta, my works would have been a failure. Mere words can not express my sense of gratitude to him.

I gratefully remember the invaluable suggestions and help of Professor J.K. Ghosh. He suggested this interesting area of research and introduced me to my supervisor.

I can never forget the help of Dr. Rahul Mukerjee, both as a teacher and as a friend.

This thesis has been written on the basis of ten of my papers, two of which have been written jointly with Professor DasGupta, one with Dr. R. Mukerjee, and two with Dr. A. Bose. I wish to express my deep gratitude to Professor DasGupta, Dr. Mukerjee and Dr. Arup Bose for giving me permission to present these joint works in this dissertation. These form the materials of Chapter 1 and Sections 2B, 4B, 4C and 5C.

The acknowledgement will remain incomplete unless I mention the sincere and hearty help of my sisters in all possible ways. They not only took care of my physical and mental health, but also assisted me in copying the manuscripts of my papers.

Last but not the least, thanks are due to Atasi Basu (Ray Chaudhuri) for her encouragement, suggestions and help. Also I take this opportunity to thank Joydeep Bhanja, Sumitra Purkayastha, Goutam Mukherjee, Dr. Arup Bose and others in the department for encouragement and discussions.

I thank Shri Bidhan Chandra Chatterjee for his excellent typing work.

S. K. Bhandari
January, 1987

C O N T E N T S

SUMMARY OF THE THESIS

SUMMARY OF THE THESIS

The works in this dissertation are primarily based on different concepts of majorization and the results thereof. The first part of this dissertation is a study on different concepts of univariate and multivariate majorization. The latter part of this dissertation includes studies on some problems in Sample Survey, and problems relating to ranking and selection with the use of some results, old and new, in majorization.

The concept of univariate majorization has been considered by economists in relation to Lorenz curve, as well as by mathematicians and statisticians, especially in the field of reliability. It appears that the results relating the different concepts that are available in the literature are not widely known; as a matter of fact, it appears often from some papers in economics that the respective authors are not familiar with some of the relevant results published earlier in journals of mathematics or statistics. In the first chapter we have a brief review of the results in univariate majorization and brought out a unified relationship among different concepts of majorization available in the literature. The extension of these concepts to the multivariate case is then studied. Certain concepts on multivariate majorization have been presented along with some

new results. These results can be related to problems in economics; with that in view some sufficient conditions for concave utility function have been presented.

An important tool in the theory of majorization is a theorem due to Hardy, Littlewood and Polya (1929), which says that for $P : n \times n$, yP in majorized by y for all $y \in R^n$ if and only if P is a doubly stochastic matrix. But similar results on weak supermajorization was an open question [Marshall and Olkin (1979)]. Such a result has been developed in the first part of Chapter 2. In particular, it has been proved that a non-negative matrix $P : n \times n$ is doubly superstochastic if and only if yP is weakly supermajorized by y, for all y with all components positive. This result is based on the following fact that a non-negative matrix $P : n \times n$ is doubly superstochastic if and only if it satisfies the following condition.

<u>Condition</u>: For $1 \leq k, \ell \leq n$, and any $k \times \ell$ submatrix B of P, total sum of the entries in B is greater than or equal to $(k + \ell - n)$.

The latter part of Chapter 2 is devoted to some mathematical problems in multivariate majorization. For two matrices $X : m \times n$ and $Y : m \times n$, Marshall and Olkin (1979) defined X to be majorized by Y, if $X = YP$, for some $n \times n$ doubly stochastic matrix P. Following Marshall and Olkin (p.433) we define X to be directionally majorized by Y, if

aX is majorized aY for all $a \in R^m$. They have posed the open question whether these two types of matrix majorizations are equivalent. Here we give some sufficient conditions under which directional majorization implies multivariate majorization. In particular, for $m = 2$, it has been proved that if all the column vectors of $Y : 2 \times n$ are boundary points in the convex hull of the column vectors of Y and this convex hull has two dimensional positive volume, then directional majorization implies multivariate majorization.

Chapter 3 is devoted to some inequalities relating to random replacement schemes introduced by Karlin (1974). In particular, a conjecture of Karlin (1974) has been studied in this context. The theory of majorization plays a central role in those problems. Neither part of Karlin's conjecture holds to be true, as has been observed by different authors [Krafft and Schaefer (1984), Schaefer (1987)]. In the first part we give short and elegant proofs of some of the existing results [Krafft and Schaefer (1984)]. In the latter, we analyse the problem from a different view point and give a large class of Schur-concave (convex) functions for which the conjecture holds. Some other related inequalities have also been derived.

The problem of selecting the most (or, least) likely event in multinomial population (using indifference zone approach) has drawn the attention of many researchers in recent years [Chen and Hwang (1984), Chen (1986), Bhandari and Bose (1987) etc.]. In those problems the technique of deriving the least favourable configuration (L.F.C.) is usually hard and cumbersome. Marshall and Olkin (1979) have introduced some applications of majorization to tackle such problems. We have applied the majorization concept in these problems following the works of Marshall and Olkin. The first part of Chapter 4 deals with the problem of selecting the cell associated with the largest probability. We have assumed the following constraint:

$$\theta_{(k)} \geq a \ \theta_{(k-1)} + b,$$

and studied the problem of deriving the L.F.C. for different possible values of a and b, where

$\theta_{(k)} \geq \theta_{(k-1)} \geq \cdots \cdots \geq \theta_{(1)}$ are the ordered values of the cell probabilities. In particular, we have disproved a conjecture of Marshall and Olkin (1979) on the form of L.F.C. for $a = 1$. Moreover, our results provide partial answers to all the four conjectures of Chen and Hwang (1984) on the form of L.F.C. The latter part of Chapter 4 deals with the problem of selecting the cell associated with the smallest probability. In that context we have assumed the following constraint:

$$\theta_{(1)} \leq a \ \theta_{(2)} - c,$$

and studied the problem of deriving the L.F.C. for different possible values of a and c. In particular, we have derived certain known results through simpler and more elegant proofs.

For the non-stochastic set-up the concept of Lorenz - dominance coincides with that of majorization between two vectors with positive components and the same total sum of the components. A desirable property for the measures of income inequality is to maintain this order of Lorenz - dominance. The first part of Chapter 5 generalises some already existing scattered results [of Bhattacharya (1963) and Ord et al (1983) etc.] on characterisation of the parent distribution by inequality measures on its truncations. In particular, it has been shown that if for some measures of inequality (in income), the upper α-truncated distributions corresponding to two income distributions F and G have the same inequality measure for every α in $(0, 1)$, then F and G are equal except for possible change in scale. Some results on Lorenz - dominance has been used to prove this. The latter part of Chapter 5 develops some inequalities among different Schur - convex functions which are mostly used as quantitative measures of income inequality.

REFERENCES

[1] Bhandari, S.K. (1986) : Characterisation of the parent distribution by inequality measures on its truncations. Sankhya, Ser. B, Vol.48, Pt.3, 297 - 300.

[2] Bhandari S.K. (1987) : On a conjecture of Karlin in sampling theory. Sankhya, Ser. A, Vol.49, Pt.1, 58 - 62.

[3] Bhandari, S.K. (1987) : On some families of Schur-convex functions those satisfy Karlin's conjecture in random replacement schemes. Proceedings of the International Conference on Multivariate Analysis, Springer - Verlag (to appear).

[4] Bhandari, S.K. (1988) : Multivariate majorization and directional majorization: Positive results. Sankhya, Ser. A, Vol. 50, Pt. 1 (to appear).

[5] Bhandari, S.K. and DasGupta, S. (1985) : Two characterizations of doubly super-stochastic matrices. Sankhya, Ser. A, Vol. 47, Pt. 3, 357 - 365.

[6] Bhandari, S.K. and Mukerjee, R. (1986) : Some relations among inequality measures. Sankhya, Ser. B, Vol. 48, Pt. 2, 258 - 261.

[7] Bhandari, S.K. and DasGupta, S. (1987) : Multivariate Majorization. Tech. Rep. No.12/87, Stat-Math. Division, Indian Statistical Institute, Calcutta.

[8] Bhandari, S.K. and Bose, A. (1987) : On selecting the most likely event. Journal of Statistical Planning and Inference. (to appear.).

[9] Bhattacharya, N. (1963) : A property of pareto distribution. Sankhya, Ser. B, Vol.25, 195-196.

[10] Chen, P. (1986) : On the least favourable configuration in multinomial selection problems. Commun. Statist. Theor. Meth. 15 (2), 367-385.

[11] Chen, R.W. and Hwang, F.K. (1984) : Some theorems, counter examples, and conjectures in multinomial selection theory.. Commun. Statist. Theor. Meth. 13 (10), 1289 - 1298.

[12] Hardy, G.H., Littlewood, J.E. and Polya, G. (1929) : Some simple inequalities satisfied by convex functions, Messenger Math. 58, 145-152.

[13] Karlin, S. (1974) : Inequalities for symmetric sampling plans I. Ann. Statist. 2, 1065 - 1094.

[14] Krafft, O. and Schäefer, M. (1984) : On Karlin's conjecture for random replacement sampling plans. Ann. Statist. 12, 1528 - 1535.

[15] Marshall, A.W. and Olkin, I. (1979) : Inequalities : Theory of Majorization and Its Applications. Academic Press, New York.

[16] Ord, J.K., Patil, G.P. and Taillie, C. (1983) : Truncated distributions and measures of income inequality. Sankhya, Ser. B, Vol.45, 413 - 430.

[17] Schaefer, M. (1987) : A counterexample to Karlin's conjecture for random replacement sampling plans. Sankhya, Vol.49, Series A, Pt.1.

CHAPTER 1

UNIVARIATE AND MULTIVARIATE MAJORIZATION

CHAPTER 1 UNIVARIATE AND MULTIVARIATE MAJORIZATION

SECTION 1A: INTRODUCTION

The concept of univariate majorization has been considered by economists in relation to Lorenz curve, as well as by mathematicians and statisticians, especially in the field of reliability. It appears that the relationship among the different scattered results that are available in the literature is not widely known; as a matter of fact, it often appears from some papers in economics that the respective authors are not familiar with some of the relevant results published earlier. Here we first give a brief review of univariate majorization and bring out the relationship among different conditions.

The problem is then posed whether these conditions could be extended to the multivariate case. Certain concepts on multivariate majorization have been presented along with some new results. The question of imposing a concave utility function for multiple commodities has been examined in terms of some conceptually understandable and realistic axioms.

SECTION 1B: UNIVARIATE MAJORIZATION

1. Lorenz Curve and Univariate Majorization

In order to measure inequality of incomes or of wealth in a given population of individuals Lorenz (1905) introduced a curve, known as the Lorenz curve, which he described as follows: "Plot along one axis cumulated percents of the population from poorest to richest, and along the other the percent of the total wealth held by these percents of the population." Let $x_1, \ldots, x_n$ denote the incomes (or wealth) of n individuals in a given population, and let $x_{(1)} \leq x_{(2)} \leq \cdots \leq x_{(n)}$ denote the ordered values of the x_i's. Then the Lorenz curve for this income distribution is the polygon graph obtained by joining the consecutive points of the sequence $(0, 0)$, $(k/n, \sum_{i=1}^{k} x_{(i)}/T)$, $k = 1, \ldots, n$, by straight lines, where $T = \sum_1^n x_i$. Following Lorenz, the distribution of incomes corresponding to a population π_1 is said to be "more even" than the income distribution of another population π_2, if the Lorenz curve for π_1 lies above the Lorenz curve for π_2; such a relation is often stated by saying that the income distribution of π_1 is Lorenz-dominated by that of π_2.

The above concept is closely related to the concept of majorization introduced by Schur (1923). Given two vectors $x = (x_1, \ldots, x_n)$ and $y = (y_1, \ldots, y_n)$ in R^n, x is said to be Schur-majorized by y, written as $x \prec y$, if

$$\sum_1^k x_{(i)} \geq \sum_1^k y_{(i)}, \quad k = 1, \ldots, n, \quad \sum_1^n x_i = \sum_1^n y_i .$$

The above relation may also be expressed equivalently as

$$\sum_1^k x_{[i]} \leq \sum_1^k y_{[i]}, \quad k = 1, \ldots, n, \quad \sum_1^n x_i = \sum_1^n y_i ,$$

where $x_{[n]} \leq \dots \leq x_{[1]}$ are the ordered values of x_i's.

The following basic result on Schur-majorization is due to Hardy, Littlewood and Polya (1929, 1934, 1952) :

<u>Theorem 1.1</u>: The following conditions are equivalent:

(a) $x \prec y$

(b) $x = yP$ for some doubly stochastic matrix P

(c) $\sum_1^n \phi(x_i) \leq \sum_1^n \phi(y_i)$ for all continuous convex functions ϕ

(d) $\sum_1^n x_i = \sum_1^n y_i$, and $\sum_1^n (x_i - a)^+ \leq \sum_1^n (y_i - a)^+$, for all real a, where $(x)^+ = \max(x, 0)$.

We shall name the above four relations (a), (b), (c) and (d) as 'the rearrangement condition', 'the structure condition', 'the convexity condition', and 'the residual condition', respectively.

It may be noted that the income distribution $x = (x_1, \dots, x_n)$ is Lorenz-dominated by another income distribution $y = (y_1, \dots, y_n)$ with $\sum_1^n y_i = \sum_1^n x_i$ if, and only if, $x \prec y$; however, for Schur-majorization the components of the vectors x and y need not be non-negative.

The area between the Lorenz curve of a given income distribution and the egalitarian line (i.e., the line joining (0 , 0) and (1 , 1) is called the Lorenz area or the area of concentration, and is used as a 'measure of inequality' (see Nygard and Sandstrom(1981)). The condition (c) can be interpreted from the viewpoint of economics by considering the total utility for incomes $(x_1, \ldots, x_n)$ as $\sum_1^n u(x_i)$, where u is a concave function.

2. Stochastic Majorization

The above four conditions (a) - (d) have been extended to a stochastic set-up and the result analogous to Theorem 1.1 has also been obtained in the literature. We shall briefly review the results.

Consider two random variables X and Y with distributions F and G, respectively, and with first moment distributions F_1 and G_1 respectively. When X and Y are non-negative with finite and non-zero means, the concept of Lorenz-domination can be extended as follows. The Lorenz curve corresponding to F is defined by $(p, F_1^{-1}(p)/E_F X)$, $0 \le p \le 1$; the Lorenz-curve corresponding to Y is similarly defined. We say that X is Lorenz-dominated by Y, written as $X \prec_L Y$, if, and only if

$$F_1^{-1}(p)/E_F X \geq G_1^{-1}(p)/E_G Y,$$

for all p in [0, 1].

Following Hardy, Littlewood and Polya (1929), the stochastic version of ordered partial sum $\sum_1^k x_{(i)}$ in case of a random variable X with distribution F would be

$$\int_{-\infty}^{F^{-1}(p)} x dF(x).$$

So the rearrangement condition ($\bar{a}$) can be expressed as follows:

$$(A) \qquad \int_{-\infty}^{F^{-1}(p)} x dF(x) \geq \int_{-\infty}^{G^{-1}(p)} y dG(y), \; 0 \leq p \leq 1 \; ; E_F X = E_G Y .$$

The above relation of course assumes the finiteness of the means of X and Y. We shall write $X \prec Y$ for the relation (A).

The convexity condition (C) has the following straightforward extension:

$$(C) \qquad E_F \phi(X) \leq E_G \phi(Y)$$

for all continuous convex function ϕ for which the above expectations exist and are finite.

Similarly the stochastic version of the residual condition (d) is the following:

$$(D) \qquad E_F (X-a)^+ \leq E_G (Y-a)^+$$

for all real a, and $E_F X = E_G Y$.

The structure condition (b) does not have a straightforward stochastic analogue. Ryff (1965) has introduced doubly stochastic operator to get the stochastic extension of the structure condition. A different development for the structure condition is given in the paper by Rothschild and Stiglitz (1970). The most satisfactory analogue of the structure condition may be formulated as follows using the concept of 'dilatation'; this is due to Strassen (1965).

Condition (B): There exists a probability space and random variables U and V associated with it such that the distribution of U is the same as that of X, the distribution of V is the same as that of Y, and $E(V \mid U = u) = u$ almost sure.

Next we consider the question whether a stochastic analogue of Theorem 1.1 could be obtained with the conditions (a) - (d) replaced by (A) - (D). The equivalence between the conditions (A) and (D) was proved by Atkinson (1970) for non-negative random variables with finite means; his proof can be easily extended to the general case for random variables with finite means. The equivalence between the conditions (B) and (C) was obtained by Strassen (1965) following the general results on dilatation theory. The equivalence between the conditions (C) and (D) essentially follows from Karamata's theorem (1932), although similar result was also obtained by Levin and Steckin (1948), Brunk (1956) and Ross (1983). In this development it has been tacitly assumed that X and Y have finite expectations.

3. Weak Majorizations

The concept of weak (Schur-) majorization between two n-vectors $x = (x_1, \ldots, x_n)$ and $y = (y_1, \ldots, y_n)$ is introduced in the literature to deal with the case when $\sum_1^n x_i \neq \sum_1^n y_i$. The vector x is said to be weakly submajorized by y, written as $x \prec_w y$, if

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

The vector x is said to be weakly supermajorized by y , written as $x \prec^{w} y$, if

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

The following basic results on weak majorization are quoted from Marshall and Olkin (1979).

__Theorem 3.1__: The following relations are equivalent:

(e) $x \prec_{w} y$

(f) $x = y P$ for some doubly substochastic matrix P

(g) $\sum_{1}^{n} \phi(x_i) \leq \sum_{1}^{k} \phi(y_i)$, for all continuous nondecreasing convex functions ϕ

(h) $\sum_{1}^{n} (x_i - a)^{+} \leq \sum_{1}^{n} (y_i - a)^{+}$, for all real a.

__Theorem 3.2__: The following relations are equivalent:

(i) $x \prec^{w} y$

(j) $x = y P$ for some doubly superstochastic matrix P.

(k) $\sum_{1}^{n} \phi(x_i) \leq \sum_{1}^{n} \phi(y_i)$ for all continuous nonincreasing convex functions ϕ.

(l) $\sum_{1}^{n} (a - x_i)^{+} \leq \sum_{1}^{n} (a - y_i)^{+}$ for all real a .

The stochastic versions of the conditions (h) and (l) are given respectively as follows:

(H) $E(X - a)^+ \leq E(Y - a)^+$, for all real a.

(L) $E(a - X)^+ \leq E(a - Y)^+$, for all real a.

The relation (H) and the relation (L) are used in the literature to define convex ordering, written as $X \leq_c Y$, and concave ordering, written as $Y \leq_{cv} X$, respectively; see Karlin and Novikoff (1963), Bessler and Veinott (1966), Marshall and Proschan (1970), and Stoyan (1983).

It follows from Karamata's theorem (1932) that the condition H (resp. Condition L) is equivalent to the following:

Condition G (resp. Condition K):

$$E_F\, \emptyset(X) \leq E_G\, \emptyset(Y)$$

for all continuous nondecreasing (nonincreasing) convex functions $\emptyset$ for which the above expectations exist and are finite.

Moreover, in relation to the structure conditions, it follows from the results of Strassen (1965) that the condition G (Condition K) is equivalent to the following:

Condition F (Condition J):

There exists a probability space and random variables U and V associated with it such that the distribution of U is the same as that of X, the distribution of V is the same as that of Y, and $E(V/U = u) \geq (\leq) u$, almost sure.

Following the proof given in Atkinson's paper (1970) it can be shown that the condition H and the condition L are respectively equivalent to the following conditions E and I:

<u>Condition E</u>: $\int_{F^{-1}(p)}^{\infty} x\,dF(x) \leq \int_{G^{-1}(p)}^{\infty} y\,dG(y), \quad 0 \leq p \leq 1$

<u>Condition I</u>: $\int_{-\infty}^{F^{-1}(p)} x\,dF(x) \geq \int_{-\infty}^{G^{-1}(p)} y\,dG(y), \quad 0 \leq p \leq 1.$

4. <u>Convex Ordering</u>

Certain results relating the convex ordering seem to be important and useful. Karlin and Novikoff (1963) have introduced the <u>cut criterion</u> which may be described as follows: Suppose X and Y have finite first moments with $EX \leq EY$. Then there exists a finite point ξ such that $F(x) \leq G(x)$ for $x < \xi$ and $F(x) > G(x)$ for $x > \xi$. It turns out that the cut criterion is a sufficient condition for convex ordering; see Stoyan (1983).

In this context Marshall and Proschan (1970) have introduced the concept of <u>star-shaped function</u>; a function η is said to be star-shaped if $\eta(cX) < c \cdot \eta(X)$, $0 < c < 1$, $X \geq 0$. A sufficient condition for two distribution functions F and G to cut one another at most once is that the function $\eta(X) = G^{-1}F(X)$ is star-shaped; such a relation is written as $F <_{*} G$. It is easy to see that $EX \leq EY$ and $F <_{*} G$ together imply

$X \leq_c Y$; see Stoyan (1983).

Stochastic majorization and especially convex ordering have been used extensively in the theory of statistical reliability. In that context, a nonnegative random variable X (or its distribution F) is said to be NBUE (new better than used in expectation) if $EX_\tau \leq EX$ for all $\tau > 0$, where X_τ is defined by

$$P[X_\tau > x] = P[X > x + \tau \mid X > \tau].$$

It has been proved that if F is NBUE with mean m, then $F \leq_c \mathrm{Exp}(m^{-1})$, where $\mathrm{Exp}(m^{-1})$ stands for the exponential distribution with mean m; see Stoyan (1983). A weaker result is given in Chandra and Singpurwalla (1981). The above result shows that the Lorenz curves of all NBUE distributions with mean m are enclosed within the Lorenz curve of $\mathrm{Exp}(m^{-1})$.

Suppose, in particular, that X_1 and X_2 are distributed as lognormal distribution with parameters (μ_1, σ_1) and (μ_2, σ_2), respectively. Then it can be shown easily that $\mu_1 \leq \mu_2$ and $\sigma_1 \leq \sigma_2$ together imply $X_1 \leq_c X_2$; see Stoyan (1983).

SECTION 1C : MULTIVARIATE MAJORIZATION

1. Introduction

The ordering of univariate populations does not have a straightforward extension to the case when the ordering is based on observations on multiple characteristics in the (experimental) units. The primary difficulty in extending the Lorenz curve, in particular, is due to the fact that there is no unique or natural way for defining an ordinal scale (poor to rich) to describe the units to start with, although attempts for such an extension has been made by Taguchi (1968). It seems that any such concept of ordering should depend on the objectives and possible uses of such a study; besides, the physical nature of the problem as manifested in concrete situations may call for some specific types of ordering on the basis of relevant auxiliary information. Any abstract formulation of the concept of ordering would be primarily a mathematical exercise, although such a formulation often may give insight into various relationships underlying it.

One such exercise is presented in the paper by Sen (1976) in relation to the problem of ordering communities of individuals on the basis of consumption data on multiple commodities. One of Sen's objectives was to compare the communities with respect to total social welfare as well as welfare standard. To briefly describe Sen's work let us consider a matrix $X = (x_{ij}) : n \times m$, where x_{ij} denotes the amount of commodity j going to person i. On the basis of certain axioms, Sen has suggested a criterion of the form $R X Q'$ for comparing social standards, where Q is the price vector of the commodities and R is the vector of ranks corresponding to the money incomes of the individuals. Although some of his axioms are far from natural, it turns out that the above criterion is approximately proportional to $e(1-G)$ for large n, where G is the Gini-index and e stands for the average money income.

To illustrate his axiomatic development, Sen has considered the problem of comparing the welfare standards of the fifteen states in India based on rural consumption data (1961-62 data). The values of $e(1-G)$ are computed for each state based on the price vectors of all the fifteen states. For any given state, the set of all states, which have lower standard than the given state in respect to their respective price vectors, are obtained. This leads to a tree-diagram or hierarchy of states with respect to the welfare standard. Although this method gives a nice comparative picture of the different states, the basis of this method is questionable. It may be noted that the criterion of welfare standard, as given by $e(1-G)$, is strongly related to inflationary prices; moreover, the pattern of consumption generally depends on prices of different commodities. The method suggested by Sen may even lead to inconsistent description.

It is not clear whether a single-variable measure of social welfare is realistic. It seems that more empirical studies in a broader perspective are needed to make a realistic formulation of the social welfare function in a given situation. It may be plausible to assume that a measure of welfare standard is invariant under identical repetitions of the set of individuals (along with their social behaviour) in a community; however, popular consumption of a certain commodity may affect the relative weight given to that commodity in determining the welfare standard.

One of the approaches for introducing partial ordering among populations based on multiple characteristics is to consider a real-valued function of these characteristics and then invoke partial ordering for univariate measurements. Such a function may be a utility function or total income based on consumption of different commodities and the associated price vector.

Partial ordering may be introduced with respect to each of the characteristics separately. Consider two $n \times m$ matrices X and Y. Following Marshall and Olkin (1979), X is said to be column-majorized by Y, written as $X \prec^{col} Y$, when

$$X_i^c \prec Y_i^c, \quad i = 1, \ldots, m$$

where X_i^c and Y_i^c stand for the ith column of X and Y, respectively. The above relation is equivalent to the following: There exist doubly stochastic matrices $D_i (i = 1, \ldots, m)$ such that $X_i^c = D_i Y_i^c$, $i = 1, \ldots, m$. It may be noted that such partial ordering is a weak relation, and ignores the association among the columns of the matrices that may play a significant role in concrete situations.

Marshall and Olkin (1979) have considered the following extension of majorization to the multivariate case. A matrix

$X: n \times m$ is said to be majorized by another matrix $Y: n \times m$, written as $X \prec Y$, if there exists a doubly stochastic matrix $D: n \times n$ such that $X = DY$. It can be seen that this partial ordering invokes a strong relation, through one such matrix D.

It can be seen that the convexity condition (c) and the structure condition (b) can easily be extended to the multivariate case if we treat each x_i in (b) and (c) as an n-vector. However, the rearrangement condition (a) and the residual condition (d) do not have any unique or natural extension to the multivariate case.

2. <u>Column-majorization</u>

We now define a convexity condition for column-majorization and show that it is equivalent to the structure condition. For a row vector $Z^R = (Z_1, \ldots, Z_m)$ define

$$g(Z^R) = g_1(Z_1) + \ldots + g_m(Z_m),$$

where g_i's are convex functions. For a matrix $X: n \times m$ with rows given by $X_1^R, \ldots, X_n^R$, define

$$\phi(X) = \sum_{i=1}^{n} g(X_i^R).$$

Let Φ be the set of all such functions ϕ.

<u>Theorem 2.1</u>: For any two $n \times m$ matrices X and Y, $X \prec^{col} Y$ if, and only if, $\phi(X) \leq \phi(Y)$ for all ϕ in Φ.

Proof: Suppose $X \prec^{col} Y$. Then there exist doubly stochastic matrices D_i such that $X_i^c = D_i Y_i^c$, $i = 1, \ldots, m$. With this structure $\phi(X) \leq \phi(Y)$ holds trivially.

Suppose now $\phi(X) \leq \phi(Y)$ for all ϕ in Φ. Taking g_j's as constant functions $(j \neq k)$, $\phi(X) \leq \phi(Y)$ implies

$$\sum_{i=1}^{n} g_k(X_{ik}) \leq \sum_{i=1}^{n} g_k(Y_{ik})$$

for all convex functions g_k. This, in turn, implies $X_k^c \prec Y_k^c$. The above result can be interpreted from economic viewpoint. To make the result correspond to the usual notions in economics we proceed as follows. Consider a matrix $X = (X_{ij})$: $n \times m$, where x_{ij} denote the amount of the jth commodity going to the individual i. Let the welfare function be defined by

$$W(X) = \sum_{i=1}^{n} U(X_i^R),$$

where X_i^R, the ith row vector of X, corresponds to the ith individual. The function U is defined as follows:

$$U(X_i^R) = \sum_{j=1}^{m} U_j(X_{ij}),$$

where U_j's are increasing and concave functions. Let $\mathcal{W}$ be the class of all such functions W. Then we get the following theorem, the standard proof of which is omitted.

<u>Theorem 2.2</u>: For two $n \times m$ matrices X and Y, $W(X) \geq W(Y)$ for all functions W in $\mathcal{W}$ if, and only if, there exist doubly stochastic matrices D_i such that $X_i^c \geq D_i Y_i^c$ $(i = 1, \ldots, m)$,

where the inequality is defined componentwise.

In relation to the partial ordering defined by majorization with respect to each of the commodities separately, we may define appropriate inequality measures for X. Let $I(X) = h(\phi_1(X_1^c), \ldots, \phi_m(X_m^c))$, where h is increasing and ϕ_i's are Schur-concave. Then it can be seen easily that $X \prec^{col} Y$ implies $I(X) \geq I(Y)$. In particular, one such inequality measure for $X \geq 0$ can be given as follows:

$$I_0(X) = \frac{1}{m} \Sigma \left[\frac{1}{2(n-1)} \frac{\sum\limits_{j \neq j'} \sum |X_{ij} - X_{ij'}|}{\sum\limits_j X_{ij}} \right]$$

In the above general definition of I(X) one may take ϕ_i's as univariate inequality measures which are Schur-concave.

3. Matrix majorization

Some important results relating matrix majorization and discussion on certain assumptions will be given in this subsection. First we give a useful necessary condition for matrix majorization.

Theorem 3.1: For two $n \times m$ matrices X and Y, $X \prec Y$ implies $Y'Y - X'X$ is positive semidefinite.

Proof: The result follows from the fact that for any doubly stochastic matrix $D: n \times n$ the matrix $I_n - D'D$ is positive semidefinite.

Example : Consider two 4×2 matrices X and Y given as follows :

$$X' = \begin{pmatrix} 2.5 & 2.5 & 2.5 & 2.5 \\ 4 & 4 & 1 & 1 \end{pmatrix}, \quad Y' = \begin{pmatrix} 4 & 4 & 1 & 1 \\ 4 & 4 & 1 & 1 \end{pmatrix}$$

Note that $X \prec^{col} Y$, but $X \not\prec Y$, which can be seen by invoking Theorem 3.1.

The equivalence between the structure condition and the convexity condition for matrix majorization has been proved by Karlin and Rinott (1983) following the general result on dilations (see Meyer (1966)); the result is stated below.

Theorem 3.2: For any two $n \times m$ matrices X and Y the following conditions are equivalent.

(i) $X = DY$ for some doubly stochastic matrix D.

(ii) $\sum_{i=1}^{n} f(X_i^R) \leq \sum_{i=1}^{n} f(Y_i^R)$, for every convex and continuous function f on R^m.

Note : Suppose X and Y denote two consumption matrices of n individuals on m commodities. Suppose W is a welfare function such that $X \prec^{col} Y$ implies $W(X) \geq W(Y)$. This means that the total welfare would be increased if the amount of each commodity for any individual is replaced by a weighted average of the quantities of that commodity as distributed over n individuals. However, it is not clear averaging separately for each commodity should lead to more welfare. For example, if the amount of farm equipment is redistributed after averaging without any change in the farm land there would be inappropriate use of the equipment which may not result into more welfare.

On the otherhand, the matrix majorization requires the <u>same</u> type of averaging for every commodity; this requirement is not only restrictive, but presupposes a definite relationship among the commodities.

Neither the column majorization nor the matrix majorisation seems to be satisfactory. However, the following aspect of matrix majorization appears to be convincing.

<u>Definition</u>: A matrix $X: n \times m$ is said to be weakly supermajorized by a matrix $Y: n \times m$, written as $X \prec^{w} Y$, if $X \geq DY$ for some doubly stochastic matrix D.

<u>Theorem 3.3</u>: For two $n \times m$ matrices X and Y, $X \prec^{w} Y$ if, and only if

$$(U(X_1^R), \ldots, U(X_n^R)) \prec^{w} (U(Y_1^R), \ldots, U(Y_n^R))$$

for all increasing concave functions U on R^m.

The proof of the above theorem depends on the following result which can be easily obtained from the development in Karlin and Rinott (1983).

<u>Theorem 3.4</u>: For two $n \times m$ matrices X and Y, the following are equivalent:

(i) $X \prec^{w} Y$

(ii) $\sum_{1}^{n} U(X_i^R) \geq \sum_{1}^{n} U(Y_i^R)$

for all increasing concave functions U on R^m.

Proof of Theorem 3.3: The condition in Theorem 3.3 implies that $\sum_{1}^{n} U(X_i^R) \geq \sum_{1}^{n} U(Y_i^R)$. Hence, by virtue of Theorem 3.4, $X \prec^{w} Y$.

Now, suppose $X \prec^{w} Y$. Then $X \geq DY$ for some doubly stochastic matrix D. Now $U(X_i^R) \geq (\sum_{k=1}^{n} d_{ik} Y_k^R)$, where $D = (d_{ij})$

$$\geq \sum_{k=1}^{n} d_{ik} U(Y_k^R) .$$

Let

$$a = (U(X_1^R), \ldots, U(X_n^R))$$

$$b = (U(Y_1^R), \ldots, U(Y_n^R)) .$$

Since the above result shows that $a \geq bD$ for some doubly stochastic D, it follows from Marshall and Olkin (1979) that $a \prec^{w} b$.

Note: Theorem 3.3 seems to be useful from the viewpoint of economics. However, it is conceptually difficult to understand a concave utility function on R^n. We shall next pose some easily understandable axioms for U to satisfy and show that these axioms imply that U is concave.

4. Concavity of Utility Function

First we shall consider the case $m = 2$. We postulate the following axioms.

Axiom 1: U is strictly increasing.

Axiom 2: U is concave in the positive direction, i.e., for $x \geq y$ and $0 \leq \lambda \leq 1$

$$U(\lambda x + (1-\lambda)y) \geq \lambda U(x) + (1-\lambda) U(y).$$

Axiom 3: U is continuously twice differentiable.

Axiom 4: Given $x_1^* > x_1^{**}$, x_2, $\Delta x_2 > 0$, define Δx_1^* and Δx_1^{**} by

$$U(x_1^*, x_2) = U(x_1^* + \Delta x_1^*, x_2 - \Delta x_2)$$

$$U(x_1^{**}, x_2) = U(x_1^{**} + \Delta x_1^{**}, x_2 - \Delta x_2).$$

Then $\Delta x_1^* > \Delta x_1^{**}$

Theorem 4.1: Under the axioms 1-4, U is a concave function on R^2.

Proof: Let

$$(4.1) \qquad U_{ij} = \frac{\partial^2 U(x_1, x_2)}{\partial x_i \, \partial x_j}; \quad i, j = 1, 2.$$

t is sufficient to show that the matrix $[U_{ij}]$ is negative semidefinite for all x_1, x_2.

Axiom 2 implies that for fixed $x \geq y$,

$$(4.2) \qquad H(\lambda) \equiv U(\lambda x + (1-\lambda) y)$$

s a concave function of λ in $[0, 1]$. This, in turn, implies hat $H''(\lambda) \leq 0$. It can be easily seen that

$$(4.3) \quad H''(\lambda) = \sum_i \sum_j (x_i - y_i)(x_j - y_j) \frac{\partial^2 U(w)}{\partial w_i \; \partial w_j} \Bigg|_{w = \lambda x + (1-\lambda) y}$$

Given $w > 0$ and $a > 0$ there exist x, y with $x \geq y$ and $0 < \lambda < 1$ such that

$$(4.4) \qquad w = \lambda x + (1 - \lambda) y ,$$

and

$$(4.5) \qquad \sum_i \sum_j a_i a_j \frac{\partial^2 U(w)}{\partial w_i \; \partial w_j} \leq 0 .$$

To see this, note that there exists $\varepsilon > 0$ such that $x = w + \varepsilon a > 0$, $y = w - \varepsilon a > 0$, and use the fact that $H''(\lambda) \leq 0$. It follows from the above development that (4.5) also holds for $a < 0$.

We want to show that (4.5) holds for all a. Suppose $U_{12} > 0$ for $x = w$. Since U is concave in each argument (by axiom 2), $U_{11} < 0$, $U_{22} < 0$. Hence (4.5) holds when $a_1 a_2 < 0$.

Now suppose that $U_{12} < 0$ for $x = w$. It follows from Axiom 4 that

$$\frac{\partial}{\partial x_1} \left(- \frac{U_2}{U_1}\right) < 0$$

where

$$U_i = \frac{\partial U(x_1 , x_2)}{\partial x_i}$$

Thus

$$U_2 U_{11} - U_1 U_{12} < 0.$$

Reversing the role of x_1 and x_2 we get

$$U_1 U_{22} - U_2 U_{12} < 0.$$

From the above two relations we get

$$U_{11} U_{22} > U_{12}^2,$$

since $U_{12} < 0$, $U_{11} < 0$, $U_{22} < 0$, $U_1 > 0$, and $U_2 > 0$. Thus the proof of the theorem is complete.

Next we consider the case $m > 2$. We define a new characteristic (or a commodity) by a mixture of the m given characteristics (or commodities) in fixed proportions. We modify axioms 1-4 so that they hold for <u>any</u> such two new characteristics. Under these modified axioms the utility function U is concave on R^m.

To see the above claim, take any two fixed points x and y in R^m, and consider the plane $\underline{P}$ passing through x and y and the origin 0 It is now sufficient to prove that U is concave on the plane $\underline{P}$. Consider the convex cone which is the intersection of the plane $\underline{P}$ and the positive orthant, and let Q_1 and Q_2 be the unit vectors corresponding to the two extreme rays of this cone. All points on $\underline{P}$ can be considered as linear combinations of Q_1 and Q_2, i.e., for $p \in \underline{P}$

$$p = p_1 Q_1 + p_2 Q_2.$$

Thus any such point p in $\underline{P}$ can be represented by (p_1, p_2). It is now sufficient to show that the modified axioms 1-4 imply axioms 1-4 in terms of (p_1, p_2). This fact trivially follows for axioms 1, 2, and 4. To see axiom 2, take any two points u and v on $\underline{P}$. Let Q_1 and Q_2 have coordinates $(i_1, \ldots, i_m)$, and $(j_1, \ldots, j_m)$, respectively, and let

$$u = u_1 Q_1 + u_2 Q_2,$$

$$v = v_1 Q_1 + v_2 Q_2.$$

Suppose now $(u_1, u_2) \leq (v_1, v_2)$. Then $u \underset{\sim}{<} v$. Thus U is concave on the line joining u and v.

5. Measures of inequality

The following inequality measures preserve the partial order of matrix majorization:

(a) $$\eta(X) = \sum_{i=1}^{n} \bar{x}_i^R \, A(\bar{x}_i^R)',$$

where

$$\bar{x}_i^R = x_i^R \cdot \text{diag}(r_1, \ldots, r_m),$$

$$r_j = 1/\sum_{i=1}^{n} x_{ij},$$

and A is a positive semidefinite matrix.

(b) $$\eta(X) = h(\phi_1(\bar{X}\alpha_1), \ldots, \phi_k(\bar{X}\alpha_k)),$$

where $\alpha_i \in R^m$, $\alpha_i \geq 0$ for all i, h is an increasing function, ϕ_i's are univariate inequality measures which preserve the partial order of majorization, and $\bar{X}$ is the matrix with rows

R_i defined in (a).

All the above measures satisfy the condition of impartiality (i.e., invariant with respect to permutations of the rows of X), and scale invariance (with respect to each variate separately).

6. Unequal Populations

Consider two populations denoted by $X = (x_1, \ldots, x_r)$ and $Y = (y_1, \ldots, y_n)$ where $r \leq n$ and x_i's and y_i's are all column vectors in R^p. The following result due to Karlin and Rinott (1983) deals with the comparison between X and Y.

Theorem 6.1 : The following are equivalent.

(i) $\sum_{i=1}^{r} f(x_i) \leq \sum_{i=1}^{n} f(y_i)$ for every continuous convex non-negative (and coordinatewise increasing) functions f on R^p.

(ii) There exists a doubly stochastic matrix $M : n \times n$ such that

$$X = (\text{resp. } \leq) \, [YM]_r \, , \quad \text{where } [YM]_r$$

denotes the $p \times r$ matrix formed from the first r columns of the $p \times n$ matrix YM.

The above has been obtained from a general result of Fischer and Holbrook (1980).

Multivariate Stochastic Majorization

A random vector X ($1 \times k$) is said to be Schur-majorized by another $1 \times k$ random vector Y, written as $X \prec Y$, if

$$E\, h(x) \leq E\, h(y)$$

for all convex functions h for which the above expectations are defined.

It has been shown by Strassen (1965) that if $X \prec Y$ then there exists a probability space with $1 \times k$ random vectors U and V defined on it such that the distributions of U and V are respectively the same as those of X and Y, and $E(V|U) = U$ a.s.

In order to ensure $X = (X_1, \ldots, X_k) \prec Y = (Y_1, \ldots, Y_k)$ one needs much more stronger condition than $X_i \prec Y_i$ for $i = 1, \ldots, k$. However, when X_i's are independent and Y_j's are also independent, then $X_i \prec Y_i$ for $i = 1, \ldots, k$ is equivalent to $X \prec Y$. This result follows from the theorem below.

Theorem 7.1: Let $X_1, \ldots, X_n$ be a set of n independent random variables, and $Y_1, \ldots, Y_n$ be another set of n independent random variables. Suppose $X_i \prec Y_i$ for $i = 1, \ldots n$. Then

$$E\, h(X_1, \ldots, X_n) \leq E\, h(Y_1, \ldots, Y_n)$$

for every real-valued function h, separately convex in each argument.

Proof: Consider a function h as described above. For any fixed t, the function $h(X_1, \ldots, X_{n-1}, t)$ is separately convex in $X_1, \ldots, X_{n-1}$. We shall prove the theorem by induction on n. By the induction hypothesis

(7.1) $$E\, h(X_1, \ldots, X_{n-1}, t) \leq E\, h(Y_1, \ldots, Y_{n-1}, t).$$

Integrating both sides of (7.1) with respect to $d\,F_n(t)$, where F_n is the c.d.f. of X_n, we get

(7.2) $$E\, h(X_1, \ldots, X_{n-1}, X_n) \leq E\, h(Y_1, \ldots, Y_{n-1}, Z_n),$$

where Z_n is distributed as F_n, independently of $Y_1, \ldots, Y_{n-1}$.

Since $E\, h(Y_1, \ldots, Y_{n-1}, t)$ is convex in t, by definition

(7.3) $$E\, h(Y_1, \ldots, Y_{n-1}, Z_n) \leq E\, h(Y_1, \ldots, Y_{n-1}, Y_n).$$

Now we get the desired result from (7.2) and (7.3). A closely-related but different result is given in Ross (1983). It is interesting to note that under the assumptions of the above theorem

$$C_1 X_1 + \ldots + C_n X_n \prec C_1 Y_1 + \ldots + C_n Y_n$$

for any C_i's, and

$$X_1 \ldots X_n \prec Y_1 \ldots Y_n .$$

However, the above theorem does not yield $\max(X_i) \prec \max(Y_i)$. In fact, the following weaker result holds for the comparison between $\max(X_i)$ and $\max(Y_i)$.

Theorem 7.2: Let $(X_1, \ldots, X_n)$ and $(Y_1, \ldots, Y_n)$ be two sets of non-negative independent random variables such that $X_i \prec Y_i$, $i = 1, \ldots, n$. Then there exists a non-negative random variable Z with $E(\max Y_i) = E(Z)$ such that

$$\max(X_i) \overset{st}{\leq} Z, \quad Z \prec \max(Y_i).$$

Proof: It follows from Theorem 7.1 that

$$(7.4) \qquad E\, h(\max(X_i)) \leq E\, h(\max(Y_i))$$

for all convex and non-decreasing h.

From the equivalence of conditions G and F on weak stochastic majorization given earlier, (7.4) implies the desired result.

SECTION 1D : REFERENCES

[1] Atkinson, A.B. (1970) : On the measurement of inequality. J. Econom. Theory 5, 244 - 263.

[2] Bessler, S. and Veinott, T. (1966) : Optimal policy for a dynamic multi-echelon inventory model. Naval Res. Logist. Quart. 13, 355 - 389.

[3] Bhandari, S.K. and DasGupta, S. (1987) : Multivariate Majorization. Tech. Rep. No.12/87, Stat-Math. Division, Indian Statistical Institute, Calcutta.

[4] Brunk, H.D. (1956) : On an inequality for convex functions Proc. Amer. Math. Soc. 7, 817 - 824.

[5] Chandra, M. and Singpurwalla, N.D. (1981) : Relationship between some notions which are common to reliability theory and economics. Math. Operations Research 6, 113 - 121.

[6] Fischer, P. and Holbrook, J.A. (1980): Balayage defined by the non-negative convex functions. Proc. Amer. Math. Soc. 79, 445 - 448.

[7] Hardy, G.H., Littlewood, J.E. and Polya, G. (1929) : Some simple inequalities statisfied by convex functions. Messenger Math. 58, 145 - 152.

[8] Hardy, G.H., Littlewood, J.E. and Polya, G. (1934, 1952) : Inequalities, 1st ed., 2nd. ed. Cambridge Univ. Press, London and New York.

[9] Karamata, J. (1932) : Sur une inegalite relative aux fonctions convexes. Publ. Math. Univ. Belgrade 1, 145 - 148.

[10] Karlin, S. and Novikoff, A. (1963) : Generalized convex inequalities. Pacific J. Math. 13, 1251 - 1279.

[11] Karlin, S. and Rinott, Y. (1983) : Comparison of measures, multivariate majorization; and applications to statistics. Studies in Econometrics, Time Series and Multivariate Statistics. Ed. S. Karlin, T. Amemiya, Leo. A. Goodman. Academic Press, New York.

[12] Levin, V.I. and Steckin, S.B. (1948) : Inequalities. Amer. Math. Soc. Transl., Ser. 2, 14, 1 - 29.

[13] Lorenz, M.O. (1905) : Methods of measuring concentration of wealth. J. Amer. Statist. Assoc. 9, 209 - 219.

[14] Marshall, A.W. and Olkin, I. (1979): Inequalities. Theory of Majorization and its Applications. Academic Press, New York.

[15] Marshall, A.W. and Proschan, F. (1970): Mean life of series and parallel systems. J. of Appl. Probab. 7, 165 - 174.

[16] Meyer, P.A. (1966): Probability and Potentials. Ginn (Blaisdell), Boston, Massachusetts.

[17] Nygard, F. and Sandstrom, A. (1981) : Measuring income inequality. Almqvist and Wiksell International, Stockholm.

[18] Ross, S.M. (1983) : Stochastic Processes. John Wiley and Sons, New York.

[19] Rothschild, M. and Stiglitz, J.E. (1970) : Increasing risk. 1. A definition. J. Econ. Theory 6.

[20] Ryff, J.V. (1965) : Orbits of L_1 functions under doubly stochastic transformations. Trans. Amer. Math. Soc. 117, 92 - 100.

[21] Schur, I. (1923) : Uber eine Klasse von Mittelbildungen mit Anwendungen die Determinanten. Theorie Sitzungsber. Berlin. Math. Gesellschaft 22, 9 - 20.

[22] Sen, A. (1973) : On Economic Inequality. Oxford Univ. Press, London.

[23] Sen, A. (1976) : Real national income. Review of Economic Studies, Vol. 43.

[24] Stoyan, D. (1983) : Comparison Methods for Queues and Other Stochastic Models. Wiley, New York.

[25] Strassen, V. (1965) : The existence of probability measures with given marginals. Ann. Math. Statist. 36, 423 - 439.

[26] Taguchi, T. (1968) : Concentration - curve methods and structures of skew populations. Ann. Inst. Stat. Math., Vol. 20.

CHAPTER 2

SOME ASPECTS OF THE THEORY OF MAJORIZATION

SECTION 2A : INTRODUCTION

As can be seen in the book by Marshall and Olkin (1979), during the last few decades a lot of researchers have shown their interest in the development of the theory of majorization and different generalisations of it. In this chapter we have developed some relevant results in that direction.

An important tool in the theory of majorization is a theorem due to Hardy, Littlewood and Polya (1929) which says that for $x, y \in R^n$, $x \prec y$ if and only if $x = yP$, for some doubly stochastic matrix P. But similar results on weak supermajorization was unknown. Such a result has been developed in Section B. Section C is devoted to some problems in multivariate majorization which may be important for the concept of dilation in R^m, $m \geq 2$ and also for the theory of measuring inequality for multivariate distributions. The results in both the sections have originated from some open problems cited in Marshall and Olkin (1979).

SECTION 2B : DOUBLY SUPER-STOCHASTIC MATRICES AND WEAK SUPER-MAJORIZATION

1. Introduction

Recall the definition of majorization ($\prec$), weak submajorization ($\prec_w$) and weak supermajorization ($\prec^w$). We have the following two theorems on majorization and weak submajorization. Let $R_+^t = \left\{(x_1, \ldots, x_t) : x_i \geq 0 \text{ for all } i\right\}$.

Theorem 1.1 : A necessary and sufficient condition that $x \prec y$ on R_+^t is that there exists a doubly stochastic matrix P such that $x = yP$.

Theorem 1.2 : A necessary and sufficient condition that $x \prec_w y$ on R_+^t is that there exists a doubly substochastic matrix P such that $x = yP$.

Theorem 1.1 has been proved by Hardy, Littlewood and Polya (1929) and Theorem 1.2 can be proved by the results of Von Neumann (1953), as can be seen in Marshall and Olkin (1979).

But a similar results in this lines for weak supermajorization was not there in the literature. In this section we have made an attempt for the same. The results in particular solves some open questions in Marshall and Olkin (1979).

Marshall and Olkin (1979) called a $t \times t$ matrix P doubly superstochastic (d.s.s.) if there exists a doubly stochastic matrix (d.s.) D such that $P \geq D$, where $\geq$ signifies elementwise inequality. They stated two necessary conditions (to be mentioned as C2 and C3) in Proposition 2.D.3 (page 31) for a matrix P to be d.s.s., and posed the open question whether any of these conditions is sufficient for a matrix P to be d.s.s. A more interesting unsettled question stated in Marshall and Olkin (page 31) is whether $yP \prec^w y$ for all $y \in R_+^t$ implies that P is d.s.s.

In this section we have resolved both the problems stated above. We have introduced a condition (C1) which is shown to be equivalent to either C2 or C3, and proved that a non-negative matrix P (≥ 0) is d.s.s. iff it satisfies C1. Moreover, we have shown that $yP \prec^w y$ for all $y \in R_+^t$ iff P is d.s.s.

2. Preliminaries

(i) For a matrix $A = [a_{ij}]$ we define

(2.1) $$(A) = \sum_i \sum_j a_{ij}$$

and

(2.2) $$(A)_+ = \sum_i \sum_j \max(a_{ij}, 0)$$

(ii) Consider a 2×2 submatrix

(2.3) $$\begin{bmatrix} d_{ij} & d_{ik} \\ d_{sj} & d_{sk} \end{bmatrix}$$

of a $t \times t$ d.s. $D = [d_{ij}]$ such that $d_{ij} > 0$, and $d_{sk} > 0$. By δ-transform of D with respect to this submatrix we mean another doubly stochastic matrix which has all the elements the same as those in D except for the elements in the above submatrix which are transformed to

$$(2.4) \qquad \begin{bmatrix} d_{ij} - \delta & d_{ik} + \delta \\ d_{sj} + \delta & d_{sk} - \delta \end{bmatrix}$$

where $\delta > 0$.

Given a $t \times t$ matrix P we say that a δ-transformation of a d.s. matrix D is invariant with respect to P if $(D-P)_+$ remains unchanged when D is replaced by its δ-transform. For simplicity, we shall use the same notation for a d.s. matrix or any of its invariant δ-transform.

(iii) Consider the class $\mathscr{D}$ of all $t \times t$ d.s. matrices. It is easy to see that $\mathscr{D}$ is compact in R^{t^2}. Note that for any $t \times t$ matrix P, the function $(D-P)_+$ is continuous in the elements of D. Hence there exists a d.s. matrix D such that

$$(2.5) \qquad (D-P)_+ = \inf_{S \in \mathscr{D}} (S-P)_+ .$$

Such a matrix D will be called a minimizer with respect to P.

(iv) By a permutational transform of a $t \times t$ matrix P we mean the matrix P with some of its rows interchanged and/or some of its columns interchanged; i.e., P is transformed to

$\pi_1 P \pi_2$, where π_1 and π_2 are permutation matrices. Note that a permutational transformation of a matrix P keeps (P) or $(P)_+$ unchanged.

Suppose D is a minimizer with respect to P. Then $\pi_1 D \pi_2$ is a minimizer with respect to $\pi_1 P \pi_2$. Note that the double super-stochastic property of a matrix is invariant under permutational transformation. For simplicity, we shall use the same notation for a matrix and any of its permutational transform whenever any condition imposed on that matrix is also satisfied by any of its permutational transform.

3. The Basic Lemma and the Main Results

The following lemma is the key to all the results in this paper. Its proof will be given later.

Basic Lemma. Let P be a $t \times t$ matrix with all non-negative elements such that $\inf_{S \in \mathscr{D}} (S-P)_+ > 0$. There exists a minimizer $D \in \mathscr{D}$ with respect to P such that D and $A = D-P$, after being subjected to a suitable permutational transformation, can be partitioned as

$$(3.1) \quad D = \begin{array}{c} \left[\begin{array}{c|c|c} D_{11} & D_{12} & D_{13} \\ \hline D_{21} & D_{22} & D_{23} \\ \hline D_{31} & D_{32} & D_{33} \end{array}\right] \begin{array}{c} p \\ q \\ r \end{array} \\ \begin{array}{ccc} \ell & m & n \end{array} \end{array}, \quad A = \begin{array}{c} \left[\begin{array}{c|c|c} A_{11} & A_{12} & A_{13} \\ \hline A_{21} & A_{22} & A_{23} \\ \hline A_{31} & A_{32} & A_{33} \end{array}\right] \begin{array}{c} p \\ q \\ r \end{array} \\ \begin{array}{ccc} \ell & m & n \end{array} \end{array}$$

where $p\ell > 0$, and

(i) $A_{11} > 0$ and all other elements of A are non-positive,

(ii) $A_{12} = 0$, $A_{21} = 0$,

(iii) each row of A_{31} contains at least one negative element, and each column of A_{13} contains at least one negative element,

(iv) $D_{32} = 0$, $D_{33} = 0$.

<u>Note</u>. If any of q, r, m, n is zero, the corresponding row and/or column of both D and A will be absent in the above partitions.

<u>Theorem 3.1</u>: A $t \times t$ matrix $P \geq 0$ is d.s.s. iff it satisfies the following condition C1 :

<u>Condition C1</u>. For $1 \leq k, \ell \leq t$, and any $k \times \ell$ submatrix B of P,

$$(B) \geq k + \ell - t .$$

<u>Proof</u>. First note that P is d.s.s. iff any of its permutational transform is d.s.s. Moreover, P satisfies the condition C1 iff any of its permutational transform satisfies C1.

If $\inf_{S \in \mathcal{D}} (S-P)_+ = 0$, we are done. Next we apply the basic lemma to P. Suppose $\inf_{S \in \mathcal{D}} (S-P)_+ > 0$ and P satisfies the condition C1. Partition P as in the basic lemma. Then

$$\begin{pmatrix} P_{11} \\ P_{21} \end{pmatrix} < \begin{pmatrix} D_{11} \\ D_{21} \end{pmatrix} = \ell - (D_{31})$$
$$= \ell - r$$
$$= \ell - (t - p - q) = \ell + (p + q) - t .$$

This contradicts the condition C1. Hence P is d.s.s.

Suppose now P is d.s.s. Then there exists a d.s. matrix D such that $P \geq D$. Consider a $k \times \ell$ submatrix P_1 of P, and without loss of generality suppose

$$P = \left[\begin{array}{c|c} P_1 & P_2 \\ \hline P_3 & P_4 \end{array}\right] \begin{matrix} k \\ t-k \end{matrix}, \quad D = \left[\begin{array}{c|c} D_1 & D_2 \\ \hline D_3 & D_4 \end{array}\right] \begin{matrix} k \\ t-k \end{matrix} .$$
$$\qquad \ell \quad t-\ell \qquad\qquad\qquad \ell \quad t-\ell$$

Then

$$(P_1) \geq (D_1) = k - (D_2')$$
$$= k - [(t-\ell) - (D_4)]$$
$$= k + \ell - t + (D_4) \geq k + \ell - t .$$

Theorem 3.2: A non-negative matrix $P : t \times t$ is d.s.s. iff $yP \prec^w y$ for all $y \in R_+^t$.

Proof. Suppose P is d.s.s. Then there exists a d.s. matrix D such that $P \geq D$. Hence for any $y \in R_+^t$

$$yP \geq yD.$$

Thus

$$yP \prec^w yD$$

Since $yD \prec y$ we have $yP \prec^w y$.

Suppose $yP \prec^w y$ for all $y \in R_+^t$. We shall show that P satisfies the condition C1, and hence, by virtue of Theorem 3.1, P is d.s.s. It can be easily seen that if $yP \prec^w y$, for all $y \in R_+^t$, then the matrix P is non-negative.

Consider a $k \times \ell$ submatrix P_1 of P. Without any loss of generality, suppose

$$P = \begin{bmatrix} P_1 & P_2 \\ P_3 & P_4 \end{bmatrix} \begin{matrix} k \\ t-k \end{matrix}$$
$$\qquad\qquad \ell \qquad t-\ell$$

Let $\delta_k = (1_k, 0) : 1 \times t$, where 1_k is the $1 \times k$ vector with all elements equal to 1. Since $\delta_k P \prec^w \delta_k$, we have

$$\sum_{i=1}^{\ell} x_{(i)} \geq \max(0, \ell + k - t),$$

where $\delta_k P = (1_k P_1 \quad 1_k P_2) \equiv (x_1, \ldots, x_t)$.

But

$$(P_1) = (1_k P_1) \geq \sum_{i=1}^{\ell} x_{(i)}.$$

Hence

$$(P_1) \geq \max(0, k + \ell - t) \geq k + \ell - t.$$

Marshall and Olkin have shown (2.D.4, page 31) that if $P \geq 0$ is d.s.s. then P satisfies the following condition:

Condition C2. For $1 \leq k, \ell \leq t$, and $t \times t$ matrix P

(the sum of elements in any k columns of P) - k $\geq$

(the sum of elements in the intersection of the k columns and any ℓ rows of P) $- \ell$.

It is easily seen that the above condition is equivalent to the condition C1. In this connection, Marshall and Olkin (1979) have introduced the following condition which is also equivalent to the condition C1:

Condition C3. For a $t \times t$ matrix P and for $1 \leq k, \ell \leq t$ (the sum of elements in any ℓ rows of P) $- \ell$ $\geq$.

(the sum of elements in the intersection of the ℓ rows and any k columns) $- k$.

4. Proof of the Basic Lemma

Let D be a minimizer with respect to P. Then $(D-P)_+ > 0$. By suitable permutational transformation of $A = D-P$ it is possible to get a left-hand upper corner block of this matrix such that

(a) each row of this block has at least one positive element,

(b) each column of this block has at least one positive element, and

(c) all elements of the matrix outside this block are non-positive.

We shall show that all elements of this block are positive, or can be made to be positive by applying suitable invariant

δ-transformations on D. Such a block will be called the "positive block" A_{11}. It is clear that $p\ell > 0$.

If the above block has only one row and/or only one column (i.e., $p = 1$ and/or $\ell = 1$) the block is trivially the positive block. Otherwise, consider an element x of this block which is not positive. Then there exists a 2×2 submatrix of this block which can be expressed, after suitable rearrangements of its row and columns, as follows:

$$(4.1) \qquad B = \begin{bmatrix} u & v \\ x & w \end{bmatrix}, \quad u > 0, \quad w > 0.$$

Next we use a δ-transformation on D such that the above submatrix B of A is changed to

$$(4.2) \qquad B_\delta \equiv \begin{bmatrix} u-\delta & v+\delta \\ x+\delta & w-\delta \end{bmatrix},$$

while all other elements of A are unchanged.

If $x < 0$, then $\delta > 0$ can be suitably chosen so that $x + \delta < 0$, and

$$(4.3) \qquad (B_\delta)_+ < (B)_+ .$$

This contradicts the assumption that D is a minimizer. Thus $x \geq 0$. Similarly $v \geq 0$.

If $x = 0 = v$, δ can be so chosen that both $u-\delta$ and $w-\delta$ are positive, and

$$(B_\delta)_+ = (B)_+ .$$

Thus this δ-transformation is invariant, and the resulting d.s. matrix D is also a minimizer. In this way all non-positive elements of this block can be changed to positive elements. We shall denote such a block by A_{11}.

It is possible to partition A and correspondingly D, (by suitable permutational transformations, if necessary) as in the basic lemma, so that (i), (ii) and (iii) hold. If $r = 0$ there is nothing else to prove.

Next we shall show that $D_{33} = 0$ when $r > 0$, $n > 0$. Suppose there is an element d_{ij} of D lying in the block D_{33} which is not zero. Then there exist an element a_{ik} in A_{31} and an element a_{sj} in A_{13} such that both a_{ik} and a_{sj} are negative. Note that $a_{sk} > 0$ and $a_{ij} \leq 0$. Now consider the 2×2 submatrix

$$B = \begin{bmatrix} a_{sk} & a_{sj} \\ a_{ik} & a_{ij} \end{bmatrix}.$$

It is possible to find a δ-transformation on D such that the above submatrix of $A = D - P$ is changed to

$$B_\delta = \begin{bmatrix} a_{sk} - \delta & a_{sj} + \delta \\ a_{ik} + \delta & a_{ij} - \delta \end{bmatrix},$$

while the other elements of A remain unchanged. It is possible to choose $\delta > 0$ so that $a_{ik} + \delta < 0$, $a_{sj} + \delta < 0$ and $a_{sk} - \delta > 0$. Hence

$$(B_\delta)_+ < (B)_+ ,$$

which contradicts that D is a minimizer. Hence $D_{33} = 0$.

Suppose $r > 0$, $m > 0$, and there exists an element d_{ij} of D lying in the block D_{32} which is not zero. Then there exists a 2×2 submatrix of A given by

$$B = \begin{bmatrix} a_{sk} & a_{sj} \\ a_{ik} & a_{ij} \end{bmatrix}$$

such that $a_{ik} < 0$ is in A_{31}, $a_{sj} = 0$ is in A_{12}, and $a_{sk} > 0$ is in A_{11}. There exists a δ-transformation on D such that the above submatrix of $A = D - P$ is changed to

$$B_\delta = \begin{bmatrix} a_{sk} - \delta & a_{sj} + \delta \\ a_{ik} + \delta & a_{ij} - \delta \end{bmatrix}$$

while the other elements of A remain unchanged, and $a_{ik} + \delta < 0$, $a_{sk} - \delta > 0$. Then

(4.4) $$(B_\delta)_+ = (B)_+ .$$

Such a δ-transformation is invariant and it keeps the structure of A_{31} (satisfying (iii)) unchanged, while changing a_{sj} to a positive element. In this way, all the elements in A_{12} lying in the jth column can be changed to positive elements by suitable invariant δ-transformations. Then this entire column can be annexed to the block

$$\begin{bmatrix} A_{11} \\ A_{21} \\ A_{31} \end{bmatrix}$$

thereby extending the positive block A_{11} by one more column. This process is continued until the remaining elements of D_{32} are zero. This process leads to the following structure of **A** :

$$(4.5) \qquad A = \left[\begin{array}{cc|c|c} A_{11} & \boxed{\begin{array}{c}\text{annexed}\\ \text{columns}\end{array}} & A_{12} & A_{13} \\ \hline A_{21} & \boxed{\phantom{\text{columns}}} & A_{22} & A_{23} \\ \hline A_{31} & \boxed{\phantom{\text{columns}}} & A_{32} & A_{33} \end{array}\right]$$

Although the new A_{31} would satisfy (iii), the new A_{21} may not be 0. If there is any row in (new) A_{21} which contains at least one negative element, then that entire row of A would be annexed to $[A_{31} \;\; A_{32} \;\; A_{33}]$. We then would get the following structure of **A** :

$$(4.6) \qquad A = \left[\begin{array}{c|c|c} A_{11} & A_{12} & A_{13} \\ \hline A_{21} & A_{22} & A_{23} \\ \hline \boxed{\text{annexed rows}} & \boxed{(1)} & \boxed{(2)} \\ A_{31} & A_{32} & A_{33} \end{array}\right]$$

The new A_{31} would still satisfy (iii). Proceeding as before, it can be shown that the elements of D corresponding to the block (2), as indicated in (4.6), are all zero. However, the elements of D corresponding to the block (1), given in (4.6), may not be all zero.

At this stage, we repeat the above entire process until $D_{32} = 0$. This can be accomplished since the above process reduces the number of columns in A_{32} while possibly increasing its number of rows subsequently. But at some stage, there may not be any columns left in A_{32} so that no new rows may be annexed. The final partitions of A and D would then satisfy all the conditions (i), (ii), (iii) and (iv) in the basic lemma.

5. Remarks

Consider the matrix

$$P = \begin{bmatrix} 1 & 1 & 1 \\ 1 & 1 & 1 \\ 1 & 1 & -1 \end{bmatrix}$$

Although P satisfies Condition C1, P is not non-negative.

We gratefully acknowledge the following comments of Dr. Rahul Mukerjee. He pointed out that our basic lemma and consequently Theorem 3.1 could easily be generalized to the case when P and $D \geq 0$ are rectangular $m \times n$ matrices with specified row-sums and column-sums of D. In that case, the Condition C1 needs to be modified accordingly, and our proofs would go through except for trivial changes. Furthermore, he mentioned that the above problem could be seen as a problem in transportation theory, apparently unsolved.

After searching the existing literature we have found that the above problem (for rectangular matrices) was solved by Mirsky (1968) when the elements of P and D are all integers. This result has also been mentioned in the book by Mirsky (1971, p 205). Mirsky has proved this result using combinatorial arguments and with the help of four other major theorems. His complete proof would be exceedingly long, whereas our proof for the real matrices is not only short but also simple.

Mirsky (1971) has also pointed out in an exercise in his book (p.211) that his result for integral matrices could be extended to real matrices. However, as noted by Mirsky (1971, p.213), a slightly more general result was obtained by Kellerer (1961, 1964) from measure-theoretic viewpoint. Although our main result could also be obtained from the general result of Kellerer, our proof is much more direct and simple.

Theorem 3.1 again appears in a paper by Cruse (1975) where the author has also mentioned the generalisation to rectangular matrices. Our proof of Theorem 3.1 is entirely different from the proof given in Cruse (1975). Moreover, our Basic Lemma provides a new characterisation of matrices which are not d.s.s. Although Theorem 3.2 follows from Theorem 3.1, the statement of this theorem along with a proof is not available in the existing literature.

oooooooooooo

SECTION 2C : MULTIVARIATE MAJORIZATION AND DIRECTIONAL MAJORIZATION

1. Introduction

The definition of majorization $x \prec y$ is motivated to make precise formulation of the idea that the components of x are "less spread out" than the components of y. This basic idea makes sense whether the components of x and y are points on the real line or points in a more general linear space. An obvious generalisation to a Banach space may come from the concept of dilation of measures. But that may not be the only generalisation even when we consider R^m. Following Marshall and Olkin (1979) here we consider two generalisations of majorization in the context of vectors in R^m in the following two definitions.

Definition 1.1: For two matrices $X^{(m \times n)}$ and $Y^{(m \times n)}$, X is said to be majorized by Y, written as $X \prec Y$, if $X = YP$, for some $n \times n$ doubly stochastic matrix P.

Definition 1.2: X is said to be directionally majorized by Y, written as $X \prec_d Y$, if $aX \prec aY$ for all $a \in R^m$.

Note that an $m \times n$ matrix is an array of n vectors of R^m. Definition 1.1 comes from the concept of dilation. Also note that $X \prec Y$ implies $X \prec_d Y$.

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 section we give sufficient conditions under which directional majorization implies multivariate majorization.

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 hull 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 coplanner. Then $X \prec_d Y$ implies $X \prec Y$ for all X. Moreover, $AX \prec AY$ for all $A: k \times m$ implies $X \prec Y$.

3. Proof of the Main Results

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

Note that directional convex functions are convex functions.

<u>Lemma 3.1</u>: 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,

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

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

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

$$\sum_1^n g(x_i) \leq \sum_1^n g(y_i)$$

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

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

$X \prec_d Y$

$\iff \alpha X \prec \alpha Y$, for all $\alpha \in R^m$

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

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

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

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

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

$$C_{L,a,b,Z}(x) = a.d(L,x), \quad \text{if } \ell(x).\ell(Z) \geq 0$$
$$= b.d(L,x), \quad \text{if } \ell(x).\ell(Z) < 0,$$

where for $A \subset R^2$, $p \in R^2$

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

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

Lemma 3.2: For $m = 2$, $X \prec_d Y$ implies that the column vectors 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 $\xi = C_{L,1,0,x_i^c}$. Note that

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

since $\xi(x_i^c) > 0$. 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

$$(3.1)\qquad \sum_{i=1}^{n} \xi(x_i^c) \le \sum_{i=1}^{n} \xi(y_i^c),$$

for all convex functions $\xi : R^m \longrightarrow R$ [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 $A_1, A_2, \ldots, A_n$ in consecutive order. By Lemma 3.2, all x_i^c's are in C. Consider a convex function ξ on C. Define $\alpha_i = \xi(A_i)$, $\alpha = (\alpha_1, \ldots, \alpha_n)$, and

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

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

$$(3.3)\qquad \xi(x_i^c) \le F_\alpha(x_i^c), \quad \xi(y_i^c) = F_\alpha(y_i^c) \text{ for all } i.$$

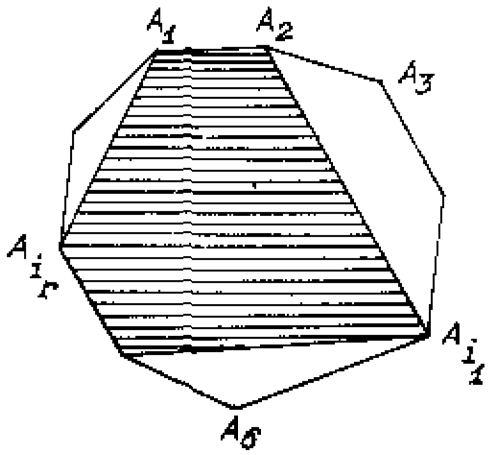
We can assume $\alpha_1 = \alpha_2 = 0$, since otherwise we can make $\alpha_1 = \alpha_2 = 0$ by adding a suitable affine function to ξ. We can furthermore assume that $\alpha_i \ge 0$ for all $i>2$, since this can be achieved by adding the affine function $C_{L,s,-s,A_i}$ for suitable large $s>0$, 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,O,A_3} for $t \geq 0$ and note that this function is affine on C. For $t = 0$

(3.4) $$C_{L,t,O,A_3}(A_i) \leq \alpha_i, \quad \text{for all } i.$$

Now as we increase t, at some point (say at $t = t_o$) at least one equality in (3.4) will be attained preserving the other inequalities. Let $A_{i_1}, A_{i_2}, \ldots, A_{i_r}$ $(r \leq n-2)$ be the vertices at which the equality in (3.4) is attained. If $r = n-2$, define $F_\alpha = C_{L,t_o,O,A_3}$ and we are done.

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



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

(3.5) $$\beta_i \geq 0, \ \beta_1 = \beta_2 = \beta_{i_1} = \quad = \beta_{i_r} = 0.$$

Consider now the polygon $A_1A_2A_{i_1}A_{i_2} \ldots A_{i_r}$ Note that $C - A_1A_2A_{i_1} \ldots A_{i_r}$ is the union of disjoint polygons. Because of (3.5) we can apply the above operation on each of these polygons, taking the two initial vertices so 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 $A_1A_2 \ldots A_n$, there exists a sub-polygon with vertices in $A_1, A_2, \ldots, A_n$ 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 $V_1, V_2, \ldots, V_k$, arranged in consecutive order, and B is a column vector of Y which lies on the segment V_1V_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_1V_2 is ≥ 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 $X \prec Y$ reduces to the corresponding problem in r-dimension.

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

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

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

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

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

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^c's are co-planner, i.e., they belong to a 2-dimensional affine space of R^m. Let these vectors be represented by the points $A_1, A_2, \ldots, A_{n-m+2}$ and their convex hull be a polygon denoted by $A_1A_2 \ldots A_{n-m+2}$, 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 $A_1 A_2 \cdots A_{n-m+2}$. 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

$$C_{H,t,0,A_3}(x) = td(H, x), \text{ if } \lambda(x)\lambda(A_3) \geq 0$$
$$= 0 \qquad , \text{ if } \lambda(x)\lambda(A_3) < 0$$

where $t \geq 0$ and $\lambda(x) = 0$ 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 ξ.

dddddddddddddd

SECTION 2D : REFERENCES

[1] Bhandari, S.K. (1984) : Multivariate Majorization and directional majorization. Tech. Report No.20/84, Stat-Math. Division, Indian Statistical Institute.

[2] Bhandari, S.K. (1988) : Multivariate majorization and directional majorization : Positive results. Sankhya 49, Series A, (to appear).

[3] Bhandari, S.K. and DasGupta, S. (1985) : Two characterizations of doubly super-stochastic matrices. Sankhya 47, Series A, Pt. 3, 357 - 365.

[4] Blackwell, D. (1953) : Equivalent comparisons of experiments. Ann. Math. Statist. 24, 265 - 272.

[5] Cruse, A.B. (1975) : A proof of Fulkerson's characterization of permutation matrices. Linear Algebra and its Applications 12, 21 - 28.

[6] Fischer, P. and Holbrook, J.A.R. (1977) : Matrices doubly stochastic by blocks. Can.J. Math. 29, 559 - 577.

[7] Hardy, G.H., Littlewood, J.E. and Polya, G. (1929) : Some simple inequalities satisfied by convex functions. Messenger Math. 58, 145 - 152.

[8] Hardy, G.H., Littlewood, J.E. and Polya, G. (1934) : Inequalities, Cambridge University Press.

[9] Kellerer, H.G. (1961) : Funktionen auf Produkträumen mit vorgegebenen Marginal-Funktionen, Math. Annalen 144 , 323 - 344.

[10] Kellerer, H.G. (1964) : Allegemeine Systeme von Reprasentanten, Z. Wahrsch. Theorie 2, 306 - 309.

[11] Marshall, A.W. and Olkin, I. (1979) : Inequalities : Theory of majorization and its applications, Academic Press, New York.

[12] Mirsky, L. (1968) : Combinatorial Theorems and Integral Matrices. J. Combinatorial Theory 5, 30 - 44.

[13] Mirsky, L. (1971) : Transversal Theory, Academic Press, New York.

[14] Rockafellar, R.T. (1970) : Convex Analysis, Princeton Univ. Press, Princeton.

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

[16] Von Neumann, J. (1953) : A certain zero-sum two-person game equivalent to the optimal assignment problem. Contributions to the Theory of Games, Vol. 2, pp 5 - 12. Princeton University Press, Princeton, New Jersey.

CHAPTER 3

MAJORIZATION AND KARLIN'S CONJECTURE FOR RANDOM REPLACEMENT SCHEMES

SECTION 3A : INTRODUCTION

A random replacement sampling plan $R(p_1, p_2, \ldots, p_{n-1})$ is a scheme for drawing a sample of n units from a population of N (distinct) units such that the ith unit is drawn at random from the remaining and the probabilities of replacing the ith unit (sampled) into the population is p_i.

Karlin (1974) conjectured that for all $N \geq n$ and for all ξ satisfying the following condition K

$$E_{R(\underline{p})}(\xi) \leq E_{R(\underline{p}')}(\xi) \qquad (*)$$

if, and only if, $p_i \leq p_i'$ for all i, where $\underline{p} = (p_1, \ldots, p_{n-1})$ and $\underline{p}' = (p_1', \ldots, p_{n-1}')$.

Condition K : A function ξ , $R^n \longrightarrow R$ is said to satisfy Condition K, if ξ is permutationally symmetric and

$$\xi(a,a,x_3,\ldots,x_n) + \xi(b,b,x_3,\ldots,x_n) \geq 2\,\xi(a,b,x_3,\ldots,x_n)$$

for all $a,b,x_3, \ldots,x_n$.

Definition 1.1 : Let $(x_1, \ldots, x_n)$ be a sample of n observations from a population of size N, and F be a symmetric function of $(x_1, \ldots, x_n)$. Define

$$\mu_F(k_1, k_2, \ldots, k_N) = F(x_1, \ldots, x_n)$$

where k_i is the number of times the ith unit appears in the sample. Note that μ_F is well-defined.

Definition 1.2 : Given a function μ_F, as defined in Definition 1.1, define

$$\phi_F(k_1, \ldots, k_N) = \frac{1}{N!} \sum_{\underline{i}} \mu_F(k_{i_1}, k_{i_2}, \ldots, k_{i_N}),$$

where the summation is taken over all permutations $\underline{i} = (i_1, \ldots, i_N)$ of $(1, 2, \ldots, N)$.

Note that ϕ_F is also the average of μ_F, over all distinguishable permutations.

The following theorem has been proved by Karlin (1974). A simpler and different proof can be found in Marshall and Olkin (1979).

Theorem 1.1 : If F satisfies Condition K, then ϕ_F is Schur-convex.

Theorem 1.1 shows that the class of Schur-convex functions are effectively more general than the class of functions that satisfy condition K.

Karlin's proof of Theorem 1.1 is very long and complicated, whereas Marshall and Olkin give a very short proof using some results of majorization.

Theorem 1.1 is the key result used by Karlin and other authors to derive inequalities related to the conjecture of Karlin. We shall use Theorem 1.1 and different consequences of it in the following sections.

Neither part of Karlin's conjecture is true; the relevant results are reported in the following sections.

In section B of this chapter, we give short and elegant proofs of some of the existing results. In section C, we analyse the problem from a different view point and give a large class of Schur-concave (convex) functions for which the conjecture holds. In section D, we develop a new concept namely "admissibility with respect to Schur-concavity" in the context of symmetric sampling plans and develop certain new inequalities. We believe that our results, particularly those in sectionsC and D, will stimulate some new research work on inequalities for symmetric sampling plans.

ππππππππππππ

SECTION 3B: SOME POSITIVE RESULTS FOR LARGE VALUES OF POPULATION SIZE

1. Introduction

Recall the inequality (*) in Section A. Karlin (1974) has shown that (*) holds if $\underline{p} = (0, 0, \ldots, 0)$, or if $p' = (1, 1, \ldots, 1)$ and

$$[N/(N-1)]^{n-1} \leq n/(n-3) \quad (1.1)$$

Krafft and Schaefer (1984) have shown that $\underline{p} \leq \underline{p}'$ implies (*) if $n \leq 7$, or if $n \geq 8$ and N is sufficiently large, Schaefer (1987) has shown that for $n \geq 8$, a sufficient condition is

$$N \geq N_o(n) = \min[N : n(N-1)^{n-1} \geq (n-3)N^{n-1}] \quad (1.2)$$

for $\underline{p} \leq \underline{p}'$ to imply (*). Furthermore, Schaefer (1987) has given an example with $n = N \geq 13$ for which $\underline{p} \leq \underline{p}'$ does not imply (*). It is easy to see that $\underline{p} \leq \underline{p}'$ does not necessarily follow from (*). [See Krafft and Schaefer (1984)].

In this section, we have shown that $\underline{p} < \underline{p}'$ implies (*) if

$$N \geq n(n-1)/3 = C_1(n), \text{ say} \quad (1.3)$$

Clearly (1.3) implies (1.1) and Karlin's result follows from our result. Moreover, this implies the result of Krafft and Schaefer (1984) for $n \geq 8$ and provides an indication of how large N should be. With reference to Schaefer (1987), note that

$$N_o = \min \{N : (1 - \frac{1}{N})^{n-1} \geq 1 - \frac{3}{n}\} \quad (1.4)$$

Hence

$N_o = C_2(n)$, if $C_2(n)$ is an integer

$= [C_2(n)] + 1$, if $C_2(n)$ is not an integer

$$\text{where } C_2(n) = [1 - (1 - \frac{3}{n})^{\frac{1}{n-1}}]^{-1} \quad (1.5)$$

The values of $C_1(n)$ and $C_2(n)$ are tabulated below.

n	8	10	15	25	50	100
$C_1(n)$	18.6	30	70	200	816.7	3300
$C_2(n)$	15.4	25.7	63.2	188.2	792.4	3250.7
percent excess of $C_1(n)$ over $C_2(n)$	19.1	16.7	10.8	6.3	3.1	1.5

Although the above shows that our result is slightly weaker than that of Schaefer's (1987) the proof of our result is much simpler.

2. Proof of the Results

Theorem 2.1: For all $\underline{p}' = (1, \delta_2, \ldots, \delta_{n-1})$ and $\underline{p} = (0, \delta_2, \ldots, \delta_{n-1})$ with $\delta_i = 0$ or 1, and for all ξ satisfying Condition K,

$$E_{R(\underline{p}')}(\xi) \geq E_{R(\underline{p})}(\xi) \quad \text{provided}$$
$$N \geq n(n-1)/3.$$

Proof: Let $\delta_{\alpha_1}, \ldots, \delta_{\alpha_t}$ be the only 0's in δ_i's. Then

$$E_{R(\underline{p}')}(\xi) = \frac{1}{N^{\alpha_1} (N-1)^{\alpha_2-\alpha_1} \ldots (N-t)^{n-\alpha_t}} \Sigma' \xi(x_{i_1}, \ldots, x_{i_n}), \tag{2.1}$$

$$E_{R(\underline{p})}(\xi) = \frac{1}{N(N-1)^{\alpha_1-1} (N-2)^{\alpha_2-\alpha_1} \ldots (N-t-1)^{n-\alpha_t}} \Sigma\, \xi(x_{i_1}, \ldots, x_{i_n}), \tag{2.2}$$

where Σ' and Σ are the summations over the following sets C' and C, respectively:

$$C' = \left\{ (x_{i_1}, \ldots, x_{i_n}) \in \Omega^n : x_{i_\lambda} \neq x_{i_j} \text{ for } \lambda > j,\ j \in J' \right\},$$

$$C = \left\{ (x_{i_1}, \ldots, x_{i_n}) \in \Omega^n : x_{i_\lambda} \neq x_{i_j} \text{ for } \lambda > j,\ j \in J \right\},$$

where

$J' = \left\{ \alpha_1, \ldots, \alpha_t \right\}$, $J = \left\{ 1, \alpha_1, \ldots, \alpha_t \right\}$, and Ω is the set of population values.

The frequency distribution in a given n-tuple $(x_{i_1}, \ldots, x_{i_n})$ may be denoted by $(e\,;\,f) = (e_1, \ldots, e_s\,;\, f_1, \ldots, f_s)$, where exactly e_i number of x_j's occur with frequency f_i in the n-tuple, and $f_1 > \ldots > f_s$. Let

$$T(e\,;\,f) = \Sigma_{e\,;\,f}\, \xi(x_{i_1}, \ldots, x_{i_n}),$$

where $\Sigma_{e\,;\,f}$ is the sum over all selections of the n-tuple of the structure

$$\{\underbrace{x_1,\ldots,x_1}_{f_1 \text{ times}},\ldots,\underbrace{x_{e_1},\ldots,x_{e_1}}_{f_1 \text{ times}},\ \underbrace{x_{e_1+1},\ldots,x_{e_1+1}}_{f_2 \text{ times}},\ldots,$$

$$\underbrace{x_{e_1+\ldots+e_s},\ldots,x_{e_1+\ldots e_s}}_{f_s \text{ times}}\}$$

For a given set of $e_1+\ldots+e_s$ distinct x's, consider an n-tuple of the above structure. Let the number of possible arrangements of the elements of this n-tuple compatible with C' and C be C'(e ; f) and C(e , f), respectively.

It is clear that C(e , f) > 0 implies C'(e , f) > 0. Let $C'_1 \subset C'$ be the subset of these n-tuples which have a singleton in their first co-ordinate; define $C'_2 = C' - C'_1$. Consider an n-tuple in C corresponding to the frequency distribution (e ; f) with $f_s = 1$. Each such element of C would then generate $(n-e_s)$ elements in C'_2 by interchanging the first coordinate with any of the other coordinates which are not singletons. On the other hand, each element in C'_2 obtained in this way would be repeated e_s times in this process, since any of the e_s singletons can occupy the first position.

$$(n-e_s)\, C\,(e\,;\,f) \le e_s\, C'_2\,(e\,;\,f),$$

Also note that $C'_1\,(e\,;\,f) = C(e\,;\,f)$

$$C'(e\,;\,f) \geq (n/e_s)\,C\,(e\,;\,f) \tag{2.3}$$

We may write

$$E_{R(p')}(\xi) = \Sigma\, a'(e\,,\,f)\,T\,(e\,,\,f), \tag{2.4}$$

$$E_{R(p)}(\xi) = \Sigma\, a(e\,,\,f)\,T\,(e\,,\,f), \tag{2.5}$$

where the summation is over all frequency distributions (e , f).
Now note that

$$\left(\frac{N-1}{N}\right)^{\alpha_1-1}\left(\frac{N-2}{N-1}\right)^{\alpha_2-\alpha_1}\cdots\left(\frac{N-t-1}{N-t}\right)^{n-\alpha_t} \geq \frac{N-n+1}{N} \tag{2.6}$$

Hence, if $e_s \leq n-3$

$$\frac{a'(e,f)}{a(a,f)} = \left(\frac{N-1}{N}\right)^{\alpha_1-1} \cdots \left(\frac{N-t-1}{N-t}\right)^{n-\alpha_t} \cdot \frac{C'(e,f)}{C(e,f)} \tag{2.7}$$

$$\geq \frac{N-n+1}{N}\cdot\frac{n}{e_s} \geq 1.$$

Now an argument, similar to that given in the proof of Theorem 3.1 (pp. 1081-1082) of Karlin (1974) may be invoked to complete the proof of the theorem.

Note: The condition $N \geq n(n-1)/3$ implies that

$$(N-t') \geq (n-t)(n-t-1)/3$$

for $t \geq t'$, and $N-t' > 0$, $n-t > 0$.

Theorem 2.2: For $\underline{p} = (p_1, \ldots, p_{n-1})$ and $\underline{p}' = (p'_1, \ldots, p'_{n-1})$ with $0 \leq p_i \leq p'_i \leq 1$, $i = 1, \ldots, n-1$, and for all ξ satisfying

Condition K,

$$E_{R(\underline{p})}(\xi) \leq E_{R(\underline{p}')}(\xi), \tag{2.8}$$

provided $N \geq n(n-1)/3$.

Proof: It is sufficient to show (2.8) for $\underline{p} \leq \underline{p}'$, where $p_t < p'_t$ and $p_i = p'_i$ for all $i \neq t$. Now

$$E_{R(\underline{p})}(\xi) = \Sigma\, p_1(\delta_1) \ldots p_{n-1}(\delta_{n-1}) E_{R(\delta_1, \ldots, \delta_{n-1})}(\xi),$$

$$E_{R(\underline{p}')}(\xi) = \Sigma p'_1(\delta_1) \ldots p'_{n-1}(\delta_{n-1}) E_{R(\delta_1, \ldots, \delta_{n-1})}(\xi),$$

where $\delta_i = 0$ or 1, $p_k(1) = p_k$, $p_k(0) = 1 - p_k$, $p'_k(1) = p'_k$, $p'_k(0) = 1 - p'_k$, and the above summations are over all $\delta_1, \ldots, \delta_{n-1}$.

Thus (2.8) will follow, if

$$E_{R(\Delta)}(\xi) \leq E_{R(\Delta')}(\xi),$$

where $\Delta = (\delta_1, \ldots, \delta_{t-1}, 0, \delta_{t+1}, \ldots, \delta_{n-1})$,

and $\Delta' = (\delta_1, \ldots, \delta_{t-1}, 1, \delta_{t+1}, \ldots, \delta_{n-1})$.

Next, note that

$$E_{R(\Delta)}(\xi) = E_{R(\Delta)}[E_{R(\Delta)}\ \xi(X_1, \ldots, X_n) | X_1, \ldots, X_{t-1}]$$

$$E_{R(\Delta')}(\xi) = E_{R(\Delta')}[E_{R(\Delta')}\ \xi(X_1, \ldots, X_n) | X_1, \ldots, X_{t-1}].$$

Since the distribution of $X_1, \ldots, X_{t-1}$ is the same under $R(\Delta)$ and $R(\Delta')$, it is sufficient to prove

$$E_{R(\Delta)}(\xi(X_1,\ldots,X_n)|X_1,\ldots,X_{t-1}) \leq E_{R(\Delta')}(\xi(X_1,\ldots,X_n)|X_1,\ldots,X_{t-1})$$

for all values of $X_1, \ldots, X_{t-1}$. After drawing $X_1, \ldots, X_{t-1}$, the population size N is reduced to $N-t'$ (with $t' \leq t-1$), and the remaining sample size becomes $n-(t-1)$. The result (2.8) now follows from Theorem 2.1 and the note after that theorem.

SECTION 3C : POSITIVE RESULTS FOR SOME RESTRICTED CLASS OF SCHUR-CONVEX (CONCAVE) FUNCTIONS

1. Introduction

As we have pointed out in the last section, Karlin's conjecture does not hold for all values of population size N and sample size n. However, it may hold for all values of N and n, for some restricted class of functions.

Recall the inequality (*) in Section A. In this section we have considered some large subclasses of Schur-convex functions which satisfy (*), for all possible values of population and sample size and for all $\underline{p} \le \underline{p}'$. As the effective sample size ρ is Schur-concave, intuitively it is clear that if we multiply a general Schur-convex (concave) function by some suitable function of ρ, then the function may satisfy (*).

2. Main Results

Unless otherwise mentioned, ρ will denote the effective sample size i.e., $\rho(k_1, k_2, \ldots, k_N) = \sum_{i=1}^{N} \min\{1, k_i\}$.

The following notations will be used :

$$\binom{N}{k} = \frac{N!}{(n-k)!\ k!}, \quad (N)_k = \frac{N!}{(N-k)!}$$

The main theorems in this section are stated below; the proofs of these theorems are given in Section 4.

<u>Theorem 2.1</u>: For $\underline{p} \leq \underline{p}'$, $\emptyset$ Schur-concave, non-negative,

$$E_{R(\underline{p}')}\left[\frac{(N-1)^{\rho}}{\binom{N}{\rho}}\emptyset\right] \leq E_{R(\underline{p})}\left[\frac{(N-1)^{\rho}}{\binom{N}{\rho}}\emptyset\right]$$

<u>Theorem 2.2</u>: For $\underline{p} \leq \underline{p}'$, $\emptyset$ Schur-convex, non-negative,

$$E_{R(\underline{p}')}\left[\frac{1}{(2^{n-1}.(e-1))^{\rho-1}(N)_{\rho-1}}\emptyset\right]$$

$$\geq E_{R(\underline{p})}\left[\frac{1}{(2^{n-1}.(e-1))^{\rho-1}(N)_{\rho-1}}\emptyset\right],$$

provided $N \geq 2n$

<u>Theorem 2.3</u>: For $\underline{p} \leq \underline{p}'$, $\emptyset$ Schur-convex, non-negative,

$$E_{R(\underline{p}')}\left[\frac{1}{(2^{2n-2})^{\rho-1}(N)_{\rho-1}}\emptyset\right]$$

$$\geq E_{R(\underline{p})}\left[\frac{1}{(2^{2n-2})^{\rho-1}(N)_{\rho-1}}\emptyset\right]$$

<u>Theorem 2.4</u>: Let $\underline{p} \leq \underline{p}'$. Then

(i) $$E_{R(\underline{p}')}\left[\frac{(N-1)^{\rho}}{\binom{N}{\rho}}g(\rho)\right] \leq E_{R(\underline{p})}\left[\frac{(N-1)^{\rho}}{\binom{N}{\rho}}g(\rho)\right],$$

where $g : R \longrightarrow R$ is increasing and non-negative.

(ii) $$E_{R(\underline{p}')}\left[\frac{1}{(2^{n-1}.(e-1))^{\rho-1}\ (N)_{\rho-1}}\, g(\rho)\right]$$

$$\geq E_{R(\underline{p})}\left[\frac{1}{(2^{n-1}.(e-1))^{\rho-1}\ (N)_{\rho-1}}\, g(\rho)\right],$$

provided $N \geq 2n$, where $g : R \longrightarrow R$ is decreasing and non-negative.

(iii) $$E_{R(\underline{p}')}\left[\frac{1}{(2^{2n-2})^{\rho-1}\ (N)_{\rho-1}}\, g(\rho)\right]$$

$$\geq E_{R(\underline{p})}\left[\frac{1}{(2^{2n-2})^{\rho-1}\ (N)_{\rho-1}}\, g(\rho)\right],$$

where $g : R \longrightarrow R$ is decreasing and non-negative.

Remark : Instead of functions of ρ , it may be interesting to consider some other Schur-convex (concave) functions as multipliers to make a general Schur-convex (concave) function satisfy Karlin's conjecture. However we have not considered these.

3. Some Results on Sampling Plans Dominating Others

Let $\Omega = \left\{ Y_1, Y_2, \ldots, Y_N \right\}$ be the population, where Y_i's are distinct.

Let $\underline{K} = \left\{ \underline{k} : \underline{k} = (k_1, \ldots, k_N),\ k_i\text{'s non-negative integers,}\right.$

$$\sum_{i=1}^{N} k_i = n \Big\} \text{ and}$$

$$\underline{L} = \left\{ \underline{x} : \underline{x} = (x_1, \ldots, x_n) \in \left\{ Y_1, \ldots, Y_N \right\}^n = \Omega^n \right\}$$

We can represent a sampling scheme in two ways. In one, sample points belong to $\underline{K}$ and in the other sample points belong to $\underline{L}$.

A sampling plan over $\underline{K}$ is called symmetric if under it Prob.$(\underline{k})$ = Prob.$(\underline{k}\,\pi)$, for all parmutation matrix $\underset{N \times N}{\pi}$.

Definition 3.1 : For a symmetric sampling plan S over $\underline{L}$, we define corresponding sampling plan K(S) over $k = (k_1, \ldots, k_N) \in \underline{K}$ as

$$\text{Prob.}(\underline{k}) = \text{Prob.} \left\{ \underline{x} = (x_1, \ldots, x_n) : k_i \text{ is the number of times the ith unit occurs in } \underline{x} \right\}$$

We shall use the notation $|A|$, to denote the cardinality of the set A.

For $I \subset I_o = \left\{1, 2, \ldots, n-1\right\}$,
we define $B_I = \Big\{ \underline{x} = (x_1, x_2, \ldots, x_n)$:

$$\rho\,(\underline{x}) = |I| + 1,\ x_j \neq x_i \text{ for all } j > i \text{ and } i \in I \Big\}. \qquad (3.1)$$

$R_I = K(S_{B_I})$, where S_{B_I} is the equal distribution on B_I.

Let R and R^* be symmetric sampling plans. The sampling plan R^* is said to be an elementary dominant of R if for some $x, y \in \underline{K}$ such that $x \prec y$,

$$R^*(y) > R(y),\ R^*(x) < R(x),$$

$R^*(x) + R^*(y) = R(x) + R(y)$

$R^*(z) = R(z)$ if z is not a permutation of x or y.

Marshall and Olkin (1979) (p. 333) defines the dominance of symmetric sampling plans as the following:

Definition 3.2: R^* is said to dominate R if there exists a finite sequence $R = R_o, R_1, \ldots, R_r = R^*$ of symmetric sampling plans such that R_j is an elementary dominant of R_{j-1}, $j = 1, \ldots, r$.

Intuitively R^* dominates R if R^* attaches more probability to sample points with respect to majorization.

Lemma 3.1: Suppose $I, J \subset I_o$ are such that $|I| = |J| = k$ and

$$I = \{i_1, \ldots, i_{t-1}, i_t, i_{t+1}, \ldots, i_k\}$$

$$J = \{i_1, \ldots, i_{t-1}, i_t+1, i_{t+1}, \ldots, i_k\}$$

$$[i_{t+1} \neq i_t+1,\ i_j < i_{j+1} \text{ for all } j],$$

then R_I dominates R_J.

Proof: $\rho(\underline{x}) = |I| + 1 = k+1$, for all $\underline{x} \in B_I$.

For $a_1, a_2, \ldots, a_{k+1}$ fixed,

let $A_{I, a_1, \ldots, a_{k+1}} = \{ \underline{x} = (x_1, x_2, \ldots, x_n) \in B_I : x_n = a_{k+1}, x_{i_p} = a_p$, for all $p = 1, 2, \ldots, k \}$. Then B_I is given by the disjoint union

$$\bigcup_{a_1, \ldots, a_{k+1} \in \Omega} A_{I, a_1, \ldots, a_{k+1}}$$

$$B_I = \sum_{a_1, a_2, \ldots, a_{k+1} \in \Omega} (a_1) \times (a_2) \times \ldots \times (a_k) \times (a_{k+1})$$

$$\times \underbrace{\frac{1}{k+1}(a_1+a_2+\ldots+a_{k+1}) \times \ldots \times \frac{1}{k+1}(a_1+a_2+\ldots+a_{k+1})}_{(i_1-1)\text{ times}}$$

$$\times \underbrace{\frac{1}{k}(a_2+\ldots+a_{k+1}) \times \ldots \times \frac{1}{k}(a_2+\ldots+a_{k+1})}_{(i_2-i_1-1)\text{ times}}$$

$$\times \ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots$$

$$\times \ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots$$

$$\times \underbrace{\frac{1}{k-t+3}(a_{t-1}+\ldots+a_{k+1}) \times \ldots \times \frac{1}{k-t+3}(a_{t-1}+\ldots+a_{k+1})}_{(i_{t-1}-i_{t-2}-1)\text{ times}}$$

$$\times \underbrace{\frac{1}{k-t+2}(a_t+\ldots+a_{k+1}) \times \ldots \times \frac{1}{k-t+2}(a_t+\ldots+a_{k+1})}_{(i_t-i_{t-1}-1)\text{ times}}$$

$$\times \underbrace{\frac{1}{k-t+1}(a_{t+1}+\ldots+a_{k+1}) \times \ldots \times \frac{1}{k-t+1}(a_{t+1}+\ldots+a_{k+1})}_{(i_{t+1}-i_t-1)\text{ times}}$$

$$\times \ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots$$

$$\vdots$$

$$\times \ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots$$

$$\times \frac{1}{1}(a_{k+1}) \times \ldots \times \frac{1}{1}(a_{k+1})$$

$$\underbrace{\qquad\qquad\qquad}_{(n-i_k-1) \text{ times}}$$

In the above expression a_i's are to be considered as indeterminates. Then in the simplified expression sample points will come as terms with corresponding probabilities as coefficients.

B_J has the same expression with

$\frac{1}{k-t+2}(a_t + \ldots + a_{k+1})$, $(i_t - i_{t-1})$ times and

$\frac{1}{k-t+1}(a_{t+1} + \ldots + a_{k+1})$, $(i_{t+1} - i_t - 2)$ times.

So B_I is obtained from B_J, by replacing

one $\frac{1}{k-t+2}(a_t + a_{t+1} + \ldots + a_{k+1})$,

by one $\frac{1}{k-t+1}(a_{t+1} + \ldots + a_{k+1})$.

This will imply that B_I dominates B_J, by observing the following :

In the expression of B_I, in the place of one $\frac{1}{k-t+1}(a_{t+1} + \ldots + a_{k+1})$, if we put a_{t+1} (or, a_{t+2} or, ..., or, a_{k+1}), then that will dominate the scheme obtained by putting a_t in the same place. This dominance is obvious by observing transfer from higher to the lower coordinates.

Thus R_I dominates R_J. This completes the proof of Lemma 3.1.

__Lemma 3.2__: Suppose $I \subset I_o$, $1 \notin I$, then R_I dominates $R_{I \cup \{1\}}$.

__Proof__: Writing B_I and $B_{I \cup \{1\}}$, in the same way as in the proof of Lemma 3.1 and using similar argument, we can prove Lemma 3.2.

__Theorem 3.1__: Suppose $I, J \subset I_o$ are such that

$$I = \{i_1, i_2, \ldots, i_k\}$$
$$i_1 < i_2 < \ldots < i_k,$$
$$J = \{j_1, j_2, \ldots, j_\lambda\}$$
$$j_1 < j_2 < \ldots < j_\lambda,$$

$k \leq \lambda$ and $j_\lambda \geq i_k$, $j_{\lambda-1} \geq i_{k-1}, \ldots, j_{\lambda-k+1} \geq i_1$, than R_I dominates R_J.

__Proof__: By repeated applications of Lemma 3.1, we have R_I dominates

$$R_{J_1}, \text{ with } J_1 = \{j_{\lambda-k+1}, \ldots, j_{\lambda-1}, j_\lambda\}.$$

Let

$$J_2 = \{1, j_{\lambda-k+1}, \ldots, j_{\lambda-1}, j_\lambda\}$$
$$J_3 = \{j_{\lambda-k}, j_{\lambda-k+1}, \ldots, j_\lambda\}.$$

By Lemma 3.2 R_{J_1} dominates R_{J_2}, by Lemma 3.1 R_{J_2} dominates R_{J_3}.

By proceeding in this way we get $R_{J_4}, R_{J_5}, \ldots, R_{J_r} = R_J$, such that R_{J_k} dominates $R_{J_{k+1}}$, $k = 3, \ldots, r-1$.
This proves Theorem 3.1.

4. Proof of the Main Results

Theorem 4.1 : For a function $\emptyset$, $E_{R(\underline{p})}(\emptyset) \leq E_{R(\underline{p}')}(\emptyset)$, for all $\underline{p} \leq \underline{p}'$ if and only if

$E_{R(\Delta)}(\emptyset) \leq E_{R(\Delta')}(\emptyset)$ for all pair (Δ, Δ') of the form

$$\Delta = (\delta_1, \delta_2, \ldots, \delta_{t-1}\ 0, \delta_{t+1}, \ldots, \delta_{n-1})$$
$$\Delta' = (\delta_1, \delta_2, \ldots, \delta_{t-1}, 1, \delta_{t+1}, \ldots, \delta_{n-1})$$

where δ_i's are 0 or 1.

The proof of the above theorem is easy and can be found in Krafft and Schaefer (1984). The proof of the following theorem can be found in Marshall and Olkin (1979).

Theorem 4.2 : The symmetric sampling plan R^* dominates R if and only if $E_R(\xi) \leq E_{R^*}(\xi)$, for all Schur-convex ξ on $\underline{K}$.

Proof of Theorem 2.1 :

In view of Theorem 4.1 it is sufficient to show for the pairs

$$\underline{p} = (\delta_1, \delta_2, \ldots, \delta_{t-1}, 0, \delta_{t+1}, \ldots, \delta_{n-1})$$
$$\underline{p}' = (\delta_1, \delta_2, \ldots, \delta_{t-1}, 1, \delta_{t+1}, \ldots, \delta_{n-1}),$$

where δ_i's are 0 or 1, for all i.

$$\text{Let } I = \left\{ i : \delta_i = 0,\ 1 \leq i \leq t-1,\ t+1 \leq i \leq n-1 \right\}$$
$$= \left\{ \alpha_1, \ldots, \alpha_k \right\} \cup \left\{ \alpha_{k+1}, \ldots, \alpha_\ell \right\} \text{ (say)}$$
$$\text{with } \alpha_1 < \ldots < \alpha_k < t < \alpha_{k+1} < \ldots < \alpha_\ell \tag{4.1}$$

Note that $R(\underline{p})$ gives equal probabilities to all elements of C and $R(\underline{p}')$ gives equal weights to all elements of the set $C \cup D$, where C and D are disjoint and

$$C = \cup \left\{ B_J : J \supset I \cup \{t\} \right\}$$
$$D = \cup \left\{ B_J : t \notin J,\ J \supset I \right\}$$, where B_J is defined in (3.1).

Let us define E_C and E_D as expectations under the equal probability distributions on the elements of C and D respectively.

Hence it is sufficient to show that for Schur-concave non-negative ξ,

$$E_D\left[\frac{(N-1)^\rho}{\binom{N}{\rho}}\xi\right] \leq E_C\left[\frac{(N-1)^\rho}{\binom{N}{\rho}}\xi\right] \tag{4.2}$$

We have,

$$E_D\left[\frac{(N-1)^\rho}{\binom{N}{\rho}}\xi\right] = \sum_{B_J \text{ in } D} |B_J|\, E_{B_J}\left[\frac{(N-1)^\rho}{\binom{N}{\rho}}\xi\right]/|D| \tag{4.3}$$

$$E_C\left[\frac{(N-1)^\rho}{\binom{N}{\rho}}\xi\right] = \sum_{B_J \text{ in } D} |B_{J\cup\{t\}}| . E_{J\cup\{t\}}\left[\frac{(N-1)^\rho}{\binom{N}{\rho}}\xi\right]/|C| \tag{4.4}$$

$$\frac{|B_{J\cup\{t\}}|}{|B_J|} = \frac{(N)_{\rho(B_{J\cup\{t\}})}}{(N)_{\rho(B_J)}} \cdot \frac{1}{\delta} \geq \frac{N-\rho(B_J)}{\rho(B_J)}, \tag{4.5}$$

where $\rho(A)$ denotes the effective sample size of the elements of A, which are necessarily equal, and $0 < \delta \leq \rho(B_J)$.

By Theorem 3.1, we have that B_J dominates $B_{JU\{t\}}$. Again for ξ non-negative and Schur-concave and as $\rho(B_{JU\{t\}}) = \rho(B_J) + 1$ using Theorem 4.2, we have

$$E_{B_{JU\{t\}}}\left[\frac{(N-1)^{\rho}}{\binom{N}{\rho}}\xi\right] \geq \frac{(N-1)(\rho(B_J)+1)}{N-\rho(B_J)} \cdot E_{B_J}\left[\frac{(N-1)^{\rho}}{\binom{N}{\rho}}\xi\right] \geq 0 \qquad (4.6)$$

$$\frac{|D|}{|C|} + 1 = \frac{|C| + |D|}{|C|} = \frac{|C \cup D|}{|C|}$$

$$= \frac{N^{\alpha_1}(N-1)^{\alpha_2-\alpha_1}\ldots(N-k+1)^{\alpha_k-\alpha_{k-1}}(N-k)^{\alpha_{k+1}-\alpha_k}\cdot(N-k-1)^{\alpha_{k+2}-\alpha_{k+1}}\ldots}{N^{\alpha_1}(N-1)^{\alpha_2-\alpha_1}\ldots(N-k+1)^{\alpha_k-\alpha_{k-1}}\cdot(N-k)^{t-\alpha_k}\cdot(N-k-1)^{\alpha_{k+1}-t}}$$

$$\frac{\ldots(N-\ell+1)^{\alpha_\ell-\alpha_{\ell-1}}\cdot(N-\ell)^{n-\alpha_\ell}}{(N-k-2)^{\alpha_{k+2}-\alpha_{k+1}}\ldots(N-\ell)^{\alpha_\ell-\alpha_{\ell-1}}\cdot(N-\ell-1)^{n-\alpha_\ell}}$$

[Using (4.1)]

Hence $\frac{N}{N-1} \leq \frac{|D|}{|C|} + 1 \leq 2^{n-1}$

$$\text{or,} \quad \frac{|D|}{2^{n-1}} \leq |C| \leq (N-1)|D|. \qquad (4.7)$$

Again we have,

$$\frac{|D|}{|C|} + 1 \leq \left(\frac{N-\ell}{N-\ell-1}\right)^{n-1} \leq \left(1 + \frac{1}{n-1}\right)^{n-1} \leq e,$$

[Provided $N-\lambda-1 \geq n-1$, i.e. provided $N \geq 2n$]

or, $|D| \leq (e-1)|C|$, provided $N \geq 2n$ (4.8)

Combining (4.4), (4.5), (4.6) and (4.7),

$$E_C\left[\frac{(N-1)^{\rho}}{\binom{N}{\rho}}\,\xi\right] \geq \frac{\sum\limits_{B_J \text{ in } D} |B_J|\,\frac{N-\rho(B_J)}{\rho(B_J)}\,(N-1)\,.\,\frac{\rho(B_J)+1}{N-\rho(B_J)}\,.E_{B_J}\left[\frac{(N-1)^{\rho}}{\binom{N}{\rho}}\,\xi\right]}{(N-1)\,.\,|D|}$$

$$\geq E_D\left(\frac{(N-1)^{\rho}}{\binom{N}{\rho}}\,\xi\right) \quad [\text{by } (4.3)].$$

Hence (4.2) is proved. Thus the proof of Theorem 2.1 is complete.

<u>Proof of Theorem 2.2</u>

From the first part of the proof of theorem 2.1, it is sufficient to show

$$E_D\left[\frac{1}{(2^{n-1}.(e-1))^{\rho-1}.(N)_{\rho-1}}\,\emptyset\right] \geq E_C\left[\frac{1}{(2^{n-1}.(e-1))^{\rho-1}.(N)_{\rho-1}}\,\emptyset\right],$$

(4.9)

provided $N \geq 2n$.

We have $\frac{|B_J \cup \{t\}|}{|B_J|} \leq (N-\rho(B_J)).(2^{n-1})$ (4.10)

Analogous to (4.6), we get

$$\therefore\ 0 \leq E_{B_{J\cup\{t\}}}\left[\frac{1}{(2^{n-1}.(e-1))^{\rho-1}.(N)_{\rho-1}}\,\emptyset\right]$$

$$\leq \frac{1}{2^{n-1}.(e-1)(N-\rho(B_J)+1)}\,E_{B_J}\left[\frac{1}{(2^{n-1}.(e-1))^{\rho-1}.(N)_{\rho-1}}\,\emptyset\right] \qquad (4.11)$$

Noting that we can get similar equations to (4.3) and (4.4), for this case and using (4.8), (4.10), (4.11), it is easy to see that we can prove (4.9), provided $N \geq 2n$.

Thus proof of Theorem 2.2 is completed.

Noting $|C| \geq \frac{|D|}{2^{n-1}}$ in equation (4.7), it is easy to prove Theorem 2.3 in a way analogous to the proof of Theorem 2.2.

From Theorems 2.1, 2.2, and 2.3, proof of Theorem 2.4 is immediate, noting that increasing functions of ρ are Schur-concave and decreasing functions of ρ are Schur-convex.

αααααααααααααααα

SECTION 3D : COMPARISON OF SYMMETRIC SAMPLING PLANS AND SOME INEQUALITIES IN RANDOM REPLACEMENT SCHEMES

1. Introduction

Though the inequality (*) in Section A does not hold for all Schur-convex function, it has been found in Section C that (*) holds for all functions of the form $\xi = F.G$, where F is any non-negative Schur-convex function and G is a suitable fixed function. It is quite natural to ask for the best G which serves this purpose. To meet this objective, we have developed a new concept in this section, namely "admissibility with respect to Schur-concavity", and, as a consequence, we have derived certain new inequalities for random replacement schemes.

2. Main Results

We shall define ordering of positive functions with respect to Schur-concavity and then using that we shall consider admissibility of such functions with respect to Schur-concavity for comparing two different sampling plans.

Definition 2.1 : Let ξ_1, ξ_2 be two positive functions. Define ξ_2 to be greater than or, equal to ξ_1 with respect to Schur-concavity, written as $\xi_1 \leq_S \xi_2$, if ξ_2/ξ_1 is Schur-concave.

Let R_1, R_2 be two different sampling plans. Let

$\underline{T}_\rho = \{ \mu : \mu > 0, \mu$ is a function of ρ and $E_{R_1}(\mu(\rho).\xi) \leq E_{R_2}(\mu(\rho).\xi)$, for all non-negative Schur-concave function $\xi \}$.

Definition 2.2 : μ_o is called ρ-admissible with respect to R_1, R_2 if $\mu_o \in \underline{T}_\rho$ and there does not exists $\mu \in \underline{T}_\rho$ such that $\mu \neq \mu_o$ and $\mu \leq_S \mu_o$.

In the following theorem we obtained a ρ-admissible function for $R_1 = R(1, 1, \ldots, 1)$, $R_2 = R(0, 1, \ldots, 1)$.

Theorem 2.1 : A ρ-admissible function μ_o for comparing $R(1, 1, \ldots, 1)$ and $R(0, 1, \ldots, 1)$ is given by

$$\frac{\mu_o(\rho + 1)}{\mu_o(\rho)} = \frac{\rho}{N - \rho} \cdot \frac{1}{(F_{N,n} - 1)} ,$$

where $F_{N,n} = (\frac{N}{N-1})^{n-1}$. Hence a form of μ_o is given by

$$\mu_o(\rho + 1) = \frac{1}{\binom{N-1}{\rho}} \cdot \frac{1}{(F_{N,n}-1)^\rho}$$

Remark 2.1 : Proof of Theorem 2.1 will be given in the next subsection. In a way similar to this proof, ρ-admissible functions for comparing other replacement schemes can be derived. Those and Theorem 2.1 give examples of series of inequalities which cannot be improved.

Remark 2.2 : Instead of functions of ρ it may be interesting to consider functions of some other Schur-concave function ξ as multiplier to make a general Schur-concave function satisfy Karlin's conjecture. Then an analogous concept, ξ-admissibility of such functions, could also be defined.

The following inequality will be derived in the next subsection using the methods in the proof of Theorem 2.1.

Theorem 2.2: For all non-negative Schur-concave function ξ and all $\underline{p} \leq \underline{p}'$,

$$(1 - \frac{n-1}{N})E_{R(\underline{p})}(\xi) \leq E_{R(\underline{p}')}(\xi) \leq \frac{1}{1-(\frac{n-1}{N})^2}E_{R(\underline{p})}(\xi) .$$

3. Proof of the Main Results

We use results of Section C for the following proof; in particular, we use Theorem C.3.1 and C.4.2.

Proof of Theorem 2.1

$$\text{Let } C = \bigcup_{1 \in J} B_J$$

$$D = \bigcup_{1 \notin J} B_J$$

$R(1, \ldots, 1)$ gives equal weights to all elements of $C \cup D$. $R(0, 1, \ldots, 1)$ gives equal weights to all elements of C. $C \cap D$ is empty. Let us define E_C and E_D as expectations under the equal probability distributions on the elements of C and D respectively. Hence

$$E_{R(1,\ldots,1)}(\xi) \leq E_{R(0,1,\ldots,1)}(\xi)$$

if and only if $E_D(\xi) \leq E_C(\xi)$ (3.1)

$$E_D(\xi) = \sum_{B_J \text{ in } D} |B_J| E_{B_J}(\xi)/|D| \quad (3.2)$$

$$E_C(\xi) = \sum_{B_J \text{ in } D} |B_J \cup \{1\}| E_{B_J \cup \{1\}}(\xi)/|C| \quad (3.3)$$

$$\frac{|D|}{|C|} = (F_{N,n} - 1) \quad (3.4)$$

$$\frac{|B_J \cup \{1\}|}{|B_J|} = \frac{N - \rho(B_J)}{\rho(B_J)}, \quad (3.5)$$

where $\rho(B_J)$ = effective sample size of the elements of B_J.

We have $E_{B_J}(\mu(\rho).\xi) = \mu(\rho(B_J)).E_{B_J}(\xi)$ (3.6)

Again from theorem C.3.1 and C.4.2,

$$\frac{E_{B_J \cup \{1\}}(\xi)}{E_{B_J}(\xi)} \geq 1,$$ for all positive Schur-concave

function ξ. Also note that this ratio can be made arbitrarily close to 1, by taking ξ suitable increasing function of ρ. This along with (3.1), ..., (3.6), shows that a ρ-admissible function is given by μ_o and proves Theorem 2.1.

Proof of Theorem 2.2

By induction and conditioning suitably [by an analogous argument, as can be seen in the proof of Theorem 3.2.2], it can be seen that it is sufficient to prove the Theorem for

$$\underline{p} = (0, \delta_1, \ldots, \delta_{n-1})$$

$$\underline{p}' = (1, \delta_1, \ldots, \delta_{n-1}), \text{ for } \delta_i = 0 \text{ or } 1, \forall i \qquad (3.7)$$

Then it is easy to see that

$$|R(\underline{p})| E_{R(\underline{p})}(\xi) \leq |R(\underline{p}')| . E_{R(\underline{p}')}(\xi),$$

where $|R(\underline{p})|$ is the number of points in the scheme $R(\underline{p})$. This proves the left-hand inequality by noting that

$$\frac{|R(\underline{p})|}{|R(\underline{p}')|} \geq \frac{N-n+1}{N} .$$

Recall (3.7). Let $I = \left\{ i : \delta_i = 0 \right\}$.

Let $$C = \bigcup_{\substack{I \subset J \\ 1 \notin J}} B_J \cup \{1\}, \quad D = \bigcup_{\substack{I \subset J \\ 1 \notin J}} B_J$$

$R(\underline{p})$ gives equal weights to all elements of C. $R(\underline{p}')$ gives equal weights to all elements of $C \cup D$. $C \cap D$ is empty.

$$E_{R(\underline{p})}(\xi) = \sum_{B_J \text{ in } D} |B_J \cup \{1\}| E_{B_J \cup \{1\}}(\xi) / |C|$$

$$E_{R(\underline{p}')}(\xi) = \sum_{B_J \text{ in } D} (|B_J| + |B_J \cup \{1\}|) E_{B_J \cup B_J \cup \{1\}}(\xi) / (|C| + |D|) .$$

The proof of Theorem 2.2 follows by noting that

$$\frac{|B_J| + |B_J \cup \{1\}|}{|B_J \cup \{1\}|} \cdot \frac{|C|}{|C| + |D|} \leq \frac{N}{N - \rho(B_J)} \cdot \frac{1}{(1 + \frac{1}{N-1})^{n-1}}$$

$$\leq \frac{N}{N-n+1} \cdot \frac{1}{1 + \frac{n-1}{N-1}} \leq \frac{1}{1 - (\frac{n-1}{N})^2}$$

Remark 3.1 : Theorem 2.2 says that for large values of the population size (N), all replacement schemes give almost same values of the expectations of non-negative Schur-concave functions. This also gives bounds for small values of N.

Remark 3.2 : An interesting (but seemingly hard) problem is to construct an Ω-admissible function, where Ω refers to some moment of the sample mean. Using this, it seems that the bounds given by Theorem 2.2 can be made more stringent.

....
ooooooooooooooo

SECTION 3E : REFERENCES

[1] Bhandari, S.K. (1986) : Comparison of symmetric sampling plans and some concepts of admissibility with respect to Schur-concavity. Tech. Report No.1/86, Stat-Math. Division, Indian Statistical Institute, Calcutta.

[2] Bhandari, S.K. (1987a) : On a conjecture of Karlin in sampling theory. Sankhya, Vol.49, Series A, Pt.1.

[3] Bhandari, S.K. (1987b) : On some families of Schur-convex functions those satisfy Karlin's conjecture in random replacement schemes. Proceedings of the International Conference on Multivariate Analysis, Springer-Verlag (to appear).

[4] Karlin, S. (1974) : Inequalities for Symmetric Sampling Plans I. Ann. Statist. 2, 1065 - 1094.

[5] Krafft, O. and Schaefer, M. (1984) : On Karlin's conjecture for random replacement sampling plans. Ann. Statist. 12, 1528 - 1535.

[6] Marshall, A.W. and Olkin, I. (1979) : Inequalities : Theory of Majorization and Its Applications. Academic Press, New York.

[7] Schaefer, M. (1987) : A counterexample to Karlin's conjecture for random replacement sampling plans. Sankhya, Vol.49, Series A, Pt. 1.

CHAPTER 4

MAJORIZATION AND RANKING MULTINOMIAL CELL PROBABILITIES

SECTION 4A : INTRODUCTION

The problem of selecting the most (or, least) likely event in a multinomial population has drawn the attention of many researchers in recent years [e.g., Chen and Hwang (1984), Chen (1986), Bhandari and Bose (1987) etc.]. Using the indifference zone approach, an obvious procedure is to select the cell corresponding to the highest (or, lowest) observed frequency, with ties broken by randomisation. It can be seen that the above rule is the uniformly best procedure in the class of rules invariant under the permutation group.

Suppose we have a sample of size N from a multinomial population with k cells. The probability vector is denoted by $\theta = (\theta_1, \theta_2, \ldots, \theta_k)$, where θ is assumed to satisfy some constraints. Denote the probability of correct selection (PCS) by $\xi(\theta)$. The least favourable configuration (L.F.C) is defined to be the configuration $\theta^*(N)$ such that $\inf_{\theta} \xi(\theta) = \xi(\theta^*(N))$. The derivation of L.F.C. is essential for calculating the sample size needed to make the PCS greater than or equal to a given value. Throughout this chapter

$$\theta_{(1)} \leq \theta_{(2)} \leq \cdots \leq \theta_{(k)}$$

will denote the ordered values of the probabilities of different cells.

Usually the problem of deriving L.F.C. is hard and cumbersome. Marshall and Olkin (1979) have introduced some applications of majorization to tackle such problems. We have applied the majorization concept in these problems following the works of Marshall and Olkin (1979). In particular, we have used the following theorem due to Kemperman (see Marshall and Olkin (1979), p.132 for a proof).

<u>Theorem 1.1</u>: Suppose that $m \leq x_i \leq M$, $i = 1, \ldots, n$. Then there exists a unique $\theta \in [m, M)$ and a unique integer $\ell \in \{0, 1, \ldots, n$ such that

$$x \prec (\underbrace{M, \ldots, M}_{\ell}, \theta, \underbrace{m, \ldots, m}_{n-\ell-1}) \text{ and } \sum_{i=1}^{n} x_i = (n-\ell-1)m + \theta + \ell M.$$

Section 4B deals with the problem of selecting the cell associated with the largest probability. We have assumed the following constraint:

$$\theta_{(k)} \geq a\, \theta_{(k-1)} + b,$$

and studied the problem of deriving the L.F.C. for different possible values of a and b. In particular, we have disproved a conjecture of Marshall and Olkin (1979) on the form of the L.F.C. for a = 1. Moreover, our results provide partial answers to all the four conjectures of Chen and Hwang (1984) on the form of L.F.C.

Section 4C deals with the problem of selecting the cell associated with the smallest probability. In that context, we have assumed the following constraint:

$$\theta_{(1)} \leq a\, \theta_{(2)} - c$$

and studied the problem of deriving the L.F.C. for different possible values of a and c. In particular, we have derived certain known results through simpler and more elegant proofs.

<<<<<<<<<<<<<>>>>>>>

SECTION 4B: SELECTING THE MOST PROBABLE CATEGORY

1. Introduction

In this section we consider the problem of selecting the most likely event in multinomial population and study the procedure which selects the cell corresponding to the highest observed frequency with ties broken by randomisation.

Recall that $\theta_{(k)}$ and $\theta_{(k-1)}$ denote the highest (unique) and the next highest of the θ_i's respectively. Kesten and Morse (1959) showed that under the following constraint:

$$\theta_{(k)} \geq a\theta_{(k-1)}, \quad a > 1 \quad \text{(given)} \tag{1.1}$$

the L.F.C. is independent of N and is given by

$$\theta^* = (\frac{a}{a+k-1}, \frac{1}{a+k-1}, \ldots, \frac{1}{a+k-1}) \tag{1.2}$$

Consider now the analogous "location" problem where the constraint is of the form:

$$\theta_{(k)} \geq \theta_{(k-1)} + b \tag{1.3}$$

Marshall and Olkin (1979, page 399) conjectured that in this situation the L.F.C. would be the following:

$$\theta^* = (\frac{1+(k-1)b}{k}, \frac{1-b}{k}, \ldots, \frac{1-b}{k}) \tag{1.4}$$

However, the conjecture is not true in general as shown by examples of Chen and Hwang (1984). We show that, when N is sufficiently large, the L.F.C. is given by

$$\theta^*(N) = ((1+b)/2,\ (1-b)/2,\ 0\ldots0). \qquad (1.5)$$

Thus here we have an interesting situation where the L.F.C. depends upon N.

Chen and Hwang (1984) deals with the problem of choosing the t best cells and state four conjectures on the existence and nature of the L.F.C. We show that all of these (the first one, only for large N and the second one, for small b) are valid for t = 1. (See subsection 4).

We also deal with the general constraint of the following form

$$\theta_{(k)} \geq a\theta_{(k-1)} + b, \qquad (1.6)$$

and study the various possible cases. The results of Kesten and Morse (1959) (the case $a \geq 1$, $b = 0$) of course follow from our results.

2. Preliminaries

In a random sample of size N from a k-cell multinomial population, let X_{iN}, $i = 0, 1, \ldots, k-1$ denote the number of observations in the ith cell. For simplicity, we drop the suffix N. In the subsequent calculations, we assume w.l.g. that $\theta_{(k)} = \theta_1$, $\theta_{(k-1)} = \theta_2$.

Note that

$$\xi(\theta_1, \ldots, \theta_k) = P(X_0 > X_\alpha \ \forall\ \alpha \neq 0)$$

$$+ \frac{1}{2} \sum_{j \neq 0} P(X_0 = X_j, X_0 > X_\alpha \ \forall\ \alpha \neq j)$$

$$+ \frac{1}{3} \sum_{\substack{j_1 \neq 0 \\ j_2 \neq 0}} P(X_0 = X_{j_1} = X_{j_2},$$

$$X_0 > X_\alpha \ \forall\ \alpha \neq j_1, j_2) + \ldots\ldots$$

$$+ \frac{1}{k} P(X_0 = X_1 = \ldots = X_{k-1})$$

The following result is due to Marshall and Olkin (1979)(p. 398).

<u>Theorem 2.1</u> : Given θ_1, ξ is a Schur-concave function of $(\theta_2, \ldots, \theta_k)$.

<u>Remark 2.1</u> : Theorems A.1.1 and 2.1 together imply that the search for the L.F.C. $\theta^*(N)$ can be confined to probability vectors of the form

$$(a\delta + b, \underbrace{\delta, \ldots, \delta}_{q}, \mu, 0, \ldots, 0) \text{ where } 0 \leq \mu \leq \delta \qquad (2.1)$$

For the case $b = 0$, $a > 1$, Kesten and Morse (1959) show that $\xi(a\delta, \delta, \ldots, \delta, \mu, 0, \ldots, 0)$ increases with δ and hence achieves its minimum at $\delta = 1/(a + k - 1)$ yielding the L.F.C. (1.2).

For $a + bk > 1$, $a > 1$, $b \leq 0$ we show that the above monotonicity holds, yielding the L.F.C. as

$$\left(\frac{a+b(k-1)}{a+k-1}, \frac{1-b}{a+k-1}, \ldots, \frac{1-b}{a+k-1}\right) \tag{2.2}$$

However, when $a = 1$, $b > 0$, the above monotonicity does not hold for large N. Indeed, just the opposite happens, as shown in Theorem 3.1.

3. The Main Results

In the following theorem, we study the cases $a > 1$, $b \leq 0$, $a + bk > 1$ and $a = 1$, $b > 0$ (for large N). The case $a \neq 1$, $b > 0$ is discussed in Remark 3.3.

Theorem 3.1: (i) When $a > 1$, $b \leq 0$, $a + bk > 1$, the L.F.C. is given by (2.2).

(ii) For $a = 1$, $b > 0$ and for sufficiently large N (depending on b and k) the L.F.C. is given by (1.5).

Remark 3.1: (a) The condition $a+bk > 1$ in (i) is essential to ensure that the cell having the largest probability is unique.

(b) The result of Kesten and Morse (1959) follows from (i) when $b = 0$.

Proof of (i): Consider any arbitrary probability vector of the form $(\alpha, \delta_1, \ldots, \delta_q, \delta_{q+1}, 0, \ldots, 0)$ where $\delta_{q+1} \leq \delta_i$, $\alpha \geq a\delta_i + b$ $\forall$ $i = 1, \ldots, (q+1)$. Observe that to prove (i) it is sufficient to prove that the derivative of ξ in the direction $(-a, \underbrace{-1, \ldots, -1}_{q}, a+q, \ldots, 0)$ is negative at all points

$(a\delta+b, \underbrace{\delta, \ldots, \delta}_{q}, \delta_{q+1}, 0, \ldots, 0)$ where $0 \leq \delta_{q+1} \leq \delta$, $a\delta+b \geq \delta$.

$$\text{i.e.,} \quad -a\frac{\partial \xi}{\partial \alpha} - \sum_{i=1}^{q} \frac{\partial \xi}{\partial \delta_i} + (a+q)\frac{\partial \xi}{\partial \delta_{q+1}} < 0 \qquad (3.1)$$

We define the following quantities, whose relevance is self evident. $N_o, N_1, \ldots, N_{q+1}$ denote *non-negative integers*. The counting of "equalities" below always includes N_o, unless otherwise stated.

$$a(\underline{N}) = \frac{N!}{\prod_{i=0}^{q+1} N_i!}, \; N_i \leq N_o, \; i = 1, \ldots, (q+1).$$

$$L(\underline{N}, \alpha, \underline{\delta}) = a(\underline{N})\alpha^{N_o} \prod_{i=1}^{q+1} \delta_i^{N_i} = L, \quad \text{say}.$$

$$Q_s = \left\{ L : \sum_{i=0}^{q+1} N_i = N-1 \text{ and } s \text{ of the } N_i\text{'s} = N_o \right\}, \; Q = \bigcup_{s=1}^{q+2} Q_s.$$

$$Q_{os} = \left\{ L : \sum_{i=0}^{q+1} N_i = N \text{ and } s \text{ of the } N_i\text{'s} = N_o \right\}, \; Q_o = \bigcup_{s=1}^{q+2} Q_{os}.$$

$$A_1 = \left\{ L : L \in Q_1 \right\}, \quad A_2 = \bigcup_{s=2}^{q+2} \left\{ L/s : L \in Q_s \right\}$$

$$A_{i1} = \bigcup_{s=1}^{q+1} \left\{ L\delta_i/(N_i+1)s : L \in Q_s, \; N_o > N_i+1 \right\}, \quad i = 1, \ldots, (q+1)$$

$$A_{i2} = \bigcup_{s=1}^{q+1} \left\{ L\delta_i/(N_i+1)(s+1) : L \in Q_s, \; N_o = N_i+1 \right\}, \quad i = 1, \ldots, (q+1)$$

$$A_{\alpha 1} = \left\{ L\alpha/(N_o + 1) : L \varepsilon \Omega_1 \right\}$$

$$A_{\alpha 2} = \bigcup_{s=2}^{q+2} \left\{ L\alpha/(N_o + 1) : L \varepsilon \Omega_s \right\}$$

$$A_{\alpha 3} = \bigcup_{s=1}^{q+1} \bigcup_{M \varepsilon \Omega_{o(s+1)}} \left\{ M/(s+1) \right\} = \bigcup_{i=1}^{q+1} \bigcup_{s=1}^{q+1} \left\{ L\delta_i/(N_i+1)(s+1) : L \varepsilon \Omega_s ,\ N_o = N_i + 1 \right\}.$$

$$A_\alpha = \bigcup_{j=1}^{3} A_{\alpha j}.$$

Let $V = \sum_{s=1}^{q+2} \sum_{L \varepsilon \Omega_s} L/s = \sum_{L \varepsilon \Omega_1} L + \sum_{s=2}^{q+2} \sum_{L \varepsilon \Omega_s} L/s = I_1 + I_2$ (say).

Note that V can also be written as

$$V = \sum_{s=1}^{q+1} \sum_{\substack{L \varepsilon \Omega_s \\ N_o > N_i + 1}} L/s + \sum_{s=1}^{q+1} \sum_{\substack{L \varepsilon \Omega_s \\ N_o = N_i + 1}} L/s + \sum_{s=2}^{q+2} \sum_{\substack{L \varepsilon \Omega_s \\ N_o = N_i}} L/s \quad \text{for any } i = 1, \ldots, (q+1).$$

$$= K_{1i} + K_{2i} + K_{3i} \quad \text{(say)}.$$

$$\frac{\partial \xi}{\partial \alpha} = \sum_{x \varepsilon A_\alpha} \frac{\partial x}{\partial \alpha} = \sum_{x \varepsilon A_{\alpha 1}} \frac{\partial x}{\partial \alpha} + \sum_{x \varepsilon A_{\alpha 2}} \frac{\partial x}{\partial \alpha} + \sum_{x \varepsilon A_{\alpha 3}} \frac{\partial x}{\partial \alpha}$$

$$= \sum_{L \varepsilon \Omega_1} L + \sum_{s=2}^{q+2} \sum_{L \varepsilon \Omega_s} L + \sum_{s=1}^{q+1} \sum_{M \varepsilon \Omega_{o(s+1)}} N_o\, M/\alpha(s+1)$$

$$= J_1 + J_2 + J_3 \quad \text{(say)}.$$

For $1 \leq i \leq q+1$, $\frac{\partial \xi}{\partial \delta_i} = \sum_{x \in A_{i1} \cup A_{i2}} \frac{\partial x}{\partial \delta_i}$

$$= \sum_{s=1}^{q+1} \sum_{\substack{L \in \Omega_s \\ N_o > N_i+1}} L/s + \sum_{s=1}^{q+1} \sum_{\substack{L \in \Omega_s \\ N_o = N_i+1}} L/(s+1)$$

$$= L_{1i} + L_{2i} \quad \text{(say)}.$$

Note that $V - \frac{\partial \xi}{\partial \delta_{q+1}} = \sum_{s=1}^{q+1} \sum_{\substack{L \in \Omega \\ N_o = N_{q+1}+1}} L/s(s+1) + \sum_{s=2}^{q+2} \sum_{\substack{L \in \Omega_s \\ N_o = N_{q+1}}} L/s > 0$

Thus to prove (3.1), it suffices to show that

$$a(V - \frac{\partial \xi}{\partial \alpha}) + \sum_{i=1}^{q} (V - \frac{\partial \xi}{\partial \delta_i}) \leq 0 .$$

The following relations are immediately observed.

$I_1 = J_1$, $K_{1i} = L_{1i}$, $i = 1, \ldots, (q+1)$.

$J_3 = \sum_{i=1}^{q+1} \sum_{s=1}^{q+1} \Sigma \left\{ L\delta_i/\alpha s(s+1) : N_o = N_i+1 \text{ and } (s+1) N_j\text{'s} = N_o \text{ in } L\delta_i \right\}$

$K_{2i} - L_{2i} = \sum_{s=1}^{q+1} \sum_{\substack{L \in \Omega_s \\ N_o = N_i+1}} L/s(s+1)$, $i = 1, \ldots, (q+1)$.

$\sum_{i=1}^{q+1} K_{3i} = \sum_{s=2}^{q+2} \sum_{L \in \Omega_s} (s-1)L/s$

$$a(I_2 - J_2) = -a \sum_{s=2}^{q+2} \sum_{L \in \Omega_s} (s-1)L/s = -a \sum_{i=1}^{q+1} K_{3i}$$

Thus $\quad a(V - \frac{\partial \xi}{\partial \alpha}) + \sum_{i=1}^{q} (V - \frac{\partial \xi}{\partial \delta_i})$

$$= a(I_2 - J_2 - J_3) + \sum_{i=1}^{q} (K_{2i} + K_{3i} - L_{2i})$$

$$\leq \sum_{i=1}^{q+1} K_{3i} + a(I_2 - J_2) + \sum_{i=1}^{q} (K_{2i} - L_{2i}) - aJ_3$$

$$\leq -(a-1) \sum_{s=2}^{q+2} \sum_{L \in \Omega_s} (s-1)L/s + \sum_{i=1}^{q} \sum_{s=1}^{q+1} \sum_{\substack{L \in \Omega_s \\ N_o = N_i + 1}} (1 - a\delta_i/\alpha)L/s(s+1).$$

Note that $a > 1$ and at the point $(a\delta + b, \underbrace{\delta, \ldots, \delta}_{q}, \delta_{q+1}, 0, \ldots, 0)$, $1 - a\delta/\alpha = b/\alpha \leq 0$. This proves (i).

<u>Proof of (ii)</u>: In this case, the expression (3.1) is written as (note that $a = 1$)

$$V - \frac{\partial \xi}{\partial \alpha} + \sum_{i=1}^{q+1} (V - \frac{\partial \xi}{\partial \delta_i}) - (q+2)(V - \frac{\partial \xi}{\partial \delta_{q+1}})$$

$$= (I_2 - J_2 - J_3) + \sum_{i=1}^{q+1} (K_{2i} + K_{3i} - L_{2i}) - (q+2)(K_{2(q+1)} + K_{3(q+1)} - L_{2(q+1)})$$

$$= \sum_{i=1}^{q+1} (K_{2i} - L_{2i}) - J_3 - (q+2)(K_{2(q+1)} + K_{3(q+1)} - L_{2(q+1)})$$

$$= \sum_{i=1}^{q+1} \sum_{s=1}^{q+1} \sum_{\substack{L \in \Omega_s \\ N_o = N_i + 1}} (1 - \delta_i/\alpha)L/s(s+1) - (2+q)(K_{2(q+1)} + K_{3(q+1)} - L_{2(q+1)})$$

$$\geq \sum_{i=1}^{q} \sum_{s=1}^{q+1} \sum_{\substack{L \varepsilon \Omega_s \\ N_o=N_1+1}} (1-\delta_i/\alpha)L/s(s+1)$$

$$- (q+2)\Big[\sum_{s=1}^{q+1} \sum_{\substack{L \varepsilon \Omega_s \\ N_o=N_{q+1}+1}} L/s(s+1) + \sum_{s=2}^{q+2} \sum_{\substack{L \varepsilon \Omega_s \\ N_o=N_{q+1}}} L/s\Big] = D_1 - (q+2)D_2 \quad \text{(say)}.$$

Thus to arrive at the configuration (1.5), it suffices to show that for sufficiently large N,

$$\int_{\eta_1}^{\eta_2} D_1(\delta_{q+1})d\delta_{q+1} > (q+2)\int_{\eta_1}^{\eta_2} D_2(\delta_{q+1})d\delta_{q+1} \;\forall\; \eta_1 < \eta_2 \leq \frac{1-b}{q+2}.$$

This follows from the following basic lemma, proving the theorem completely.

<u>Lemma 3.3</u>: As $N \longrightarrow \infty$,

$$\int_{\eta_1}^{\eta_2} D_1(\delta_{q+1})d\delta_{q+1} \Big/ \int_{\eta_1}^{\eta_2} D_2(\delta_{q+1})d\delta_{q+1} \longrightarrow \infty \quad \text{uniformly in}$$

$$\ast \; 0 \leq \eta_1 < \eta_2 \leq \frac{1-b}{q+2}.$$

<u>Proof</u>: Assume $\eta_1 = 0$. For $\eta_1 > 0$, the proof is similar.

Throughout the proof, c shall denote a constant independent of η and N. Note that $1 - \delta_i/\alpha = b/\alpha \;\forall\; i = 1, \ldots, q$.

Thus $$D_1(\delta_{q+1}) = \frac{bq}{\alpha} \sum_{s=1}^{q+1} \sum_{\substack{L \varepsilon \Omega_s \\ N_o=N_1+1}} L/s(s+1)$$

$$= \frac{bq}{\alpha}\, \Sigma_1 P(X_o = N_o,\ X_1 = N_o - 1/X_o + X_1 = 2N_o - 1)\, P\,(\sum_{j=2}^{q+1} X_j = N - (2N_o - 1))$$

$$\times\ \Sigma_2 P(X_j = N_j,\ j = 2, \ldots, (q+1))\, s(N_o, N_2, \ldots, N_{q+1}) / P(\sum_{j=2}^{q+1} X_j = N - (2N_o - 1)) \qquad (5.2)$$

[where Σ_1 denotes summation over all N_o, $N/(q+2) < N_o \leq (N+1)/2$ and, for fixed N_o, Σ_2 denotes summation over all N_j's such that $\sum_{j=2}^{q+1} N_j = N - (2N_o - 1)$.

$$s(N_o, N_2, \ldots, N_{q+1}) = 1/s(s+1) \text{ if exactly } (s-1) \text{ of the } N_j\text{'s equal } N_o \text{ and } N_j \leq N_o\ \forall\ j$$

$$= 0 \quad \text{if } N_j > N_o \text{ for some } j.]$$

$$\geq \frac{bq}{\alpha} \sum^{*}_{\frac{1}{(q+2)} < r \leq \frac{1}{2} - \varepsilon} P(X_o = Nr,\ X_1 = Nr - 1/X_o + X_1 = 2Nr - 1)\, P(\sum_{j=2}^{q+1} X_j = N - (2Nr - 1))\, f(r, \delta_{q+1}, N) \qquad \text{(say)}$$

where (*) indicates that the summation is over all r's such that Nr is an integer and ε is a fixed small pre-assigned number. From the relations $\alpha + q\delta + \delta_{q+1} = 1$, $\alpha = b + \delta$, we have

$$\delta = (1 - \delta_{q+1} - b)/(q+1),\ \alpha = (1 + qb - \delta_{q+1})/(q+1),\ 1 - (\alpha + \delta) = ((q-1)(1-b) + 2\delta_{q+1})/(q+1)$$

Note that for $q > 1$, $1 - (\alpha + \delta)$ is bounded away from zero. Thus using Stirling's approximation for $q > 1$,

$$P(X_o = Nr,\ X_1 = Nr-1/X_o+X_1 = 2Nr-1)\ P\left(\sum_{j=2}^{q+1} X_j = N-(2Nr-1)\right)$$

$$\geq \frac{c(1+qb-\delta_{q+1})^{Nr}(1-b-\delta_{q+1})^{Nr}((q-1)(1-b)+2\delta_{q+1})^{N(1-2r)}}{N(1-2r)^{1/2}r(r^{2r}(1-2r)^{1-2r})^N(q+1)^N}$$

Writing $f_1(\delta_{q+1}, r) = \dfrac{(1+qb-\delta_{q+1})^r(1-b-\delta_{q+1})^r((q-1)(1-b)+2\delta_{q+1})^{1-2r}}{r^{2r}(1-2r)^{1-2r}}$,

$$\left(\int_o^{\eta} D_1(\delta_{q+1})d\delta_{q+1}\right)^{1/N} \geq \frac{c^{1/N}}{(q+1)}\left[\int_o^{\eta} \sum^*_{\frac{1}{(q+2)} < r \leq \frac{1}{2}-\varepsilon} f(r,\delta_{q+1},N)\frac{f_1^N(\delta_{q+1},r)}{N}\right]^{1/N} \tag{3.3}$$

We shall show later that in a small neighbourhood of a point of maximum of $f_1(\delta_{q+1}, r)$, $f(r, \delta_{q+1}, N)$ is bounded away from zero uniformly in r. (3.4)

From this it follows that (note that the summation in (*) is a Riemann sum) for $q > 1$,

$$\lim_{N\to\infty}\left[\int_o^{\eta} D_1(\delta_{q+1})d\delta_{q+1}\right]^{1/N} \geq \frac{1}{(q+1)} \max f_1(\delta_{q+1}, r) \tag{3.5}$$

where the max is over all (r,δ) such that $\frac{1}{(q+2)} \leq r \leq \frac{1}{2} - \varepsilon$, $0 \leq \delta_{q+1} \leq \eta$. It is easy to see that (3.5) holds also for $q = 1$.

Let us now study the behaviour of D_2.

$$D_2 = \sum_{s=1}^{q+1} \sum_{\substack{L\in\Omega_s \\ N_o=N_{q+1}+1}} L/s(s+1) + \sum_{s=2}^{q+2} \sum_{\substack{L\varepsilon\Omega_s \\ N_o=N_{q+1}}} L/s$$

$= D_2(A) + D_2(B)$ (say).

Clearly $D_2(B) \leq c(q)D_2(A)$ (Compare each term and use $\delta_{q+1} \leq \delta$).

Hence $\int_0^\eta D_2(\delta_{q+1})d\delta_{q+1} \leq c(q) \int_0^\eta D_2(A)(\delta_{q+1})d\delta_{q+1}$

Define $f_2(\delta_{q+1}, r) = \dfrac{(1+qb-\delta_{q+1})^r(q(1-b-\delta_{q+1}))^{1-2r}((q+1)\delta_{q+1})^r}{r^{2r}(1-2r)^{1-2r}}$

Proceeding as in D_1,

$$\overline{\lim_{N\to\infty}}\left(\int_0^\eta D_2(\delta_{q+1})d\delta_{q+1}\right)^{1/N} \leq \lim_{N\to\infty} \frac{c^{1/N}}{(q+1)}\left[\int_{1/(q+2)}^{1/2}\int_0^\eta \frac{1}{r(1-2r)^{1/2}}drd\delta_{q+1}\right]^{1/N}.$$

$$\max f_2(\delta_{q+1}, r)$$

$\leq \dfrac{1}{(q+1)} \max f_2(\delta_{q+1}, r)$ where the maximum is over all (r,δ) such that $\dfrac{1}{(q+2)} \leq r \leq \dfrac{1}{2}$, $0 \leq \delta_{q+1} \leq \eta$.

Thus it suffices to show that $\forall\, \eta \leq (1-b)/(q+2)$,

$$\max_{\substack{\frac{1}{(q+2)} \leq r \leq \frac{1}{2} - \varepsilon \\ 0 \leq \delta_{q+1} \leq \eta}} f_1(\delta_{q+1}, r) > p_1 > p_2 > \max_{\substack{\frac{1}{(q+2)} \leq r \leq \frac{1}{2} \\ 0 \leq \delta_{q+1} \leq \eta}} f_2(\delta_{q+1}, r) \qquad (3.6)$$

First assume that $q > 1$. Suppose $\eta < (1-b)/(q+2)$.

Let $x = \dfrac{1-b-\delta_{q+1}}{(q+1)\delta_{q+1}}$. Note that $x > 1$ and x decreases as

δ_{q+1} increases. It is easily checked that

$$g(\delta_{q+1},r) = f_1(\delta_{q+1},r)/f_2(\delta_{q+1},r) = A_r(x), \text{ where}$$

$$A_r(x) = x^r\left(\frac{q-1+\frac{1}{x}}{q}\right)^{1-2r}$$

$$\frac{\partial \log A_r(x)}{\partial x} = \frac{r}{x} + (1-2r)\frac{1}{q-1+\frac{1}{x}}(-1/x^2)$$

$$= \frac{r}{x} - \frac{1-2r}{x+(q-1)x^2} > 0, \quad \text{since} \quad r \geq \frac{1}{(q+2)}.$$

Thus $g(\delta_{q+1},r)$ is a strictly decreasing function of δ_{q+1} (3.7)

Also note that $g((1-b)/(q+2), r) = 1 \; \forall \; r$ (3.8)

Hence $\max\limits_{0 \leq \delta_{q+1} \leq \eta} g(\delta_{q+1}, r) > t(\eta) > 1 \; \forall \; \eta < (1-b)/(q+2)$ (3.9)

where $t(\eta)$ is a decreasing function of η.

(3.6) would follow if we could show that $f_1(\delta_{q+1}, r)$ attains maximum at some $r \leq \frac{1}{2} - \varepsilon$ and $\delta_{q+1} < (1-b)/(q+2)$ (3.10)

This is because f_2 attains its maximum at $\delta_{q+1} = (1-b)/(q+2)$ and for some r. Hence (3.6) immediately follows from (3.8), (3.9) and (3.10). (That f_2 attains its maximum at $\delta_{q+1} = (1-b)/(q+2)$ can be shown by showing that

$$\frac{\partial^2 \log f_2}{\partial \delta_{q+1}^2} < 0 \quad \text{and} \quad \frac{\partial \log f_2}{\partial \delta_{q+1}}\left(\frac{1-b}{q+2}, r\right) > 0).$$

Thus it remains to show (3.10) (and of course (3.4)).

Proof of (3.10): Suppose we prove the following

$$\frac{\partial^2 \log f_1}{\partial r^2}\left(\frac{1-b}{q+2}, r\right) < 0 \qquad (3.11)$$

$\frac{\partial \log f_1}{\partial r}\left(\frac{1-b}{q+2}, r_1\right) = 0$, and maximum of f_1 on the line

$\delta_{q+1} = (1-b)/(q+2)$ occurs at r_1 (3.12)

$$\frac{\partial \log f_1}{\partial \delta_{q+1}}\left(\frac{1-b}{q+2}, r\right) \text{ is strictly decreasing in } r \qquad (3.13)$$

$$\frac{\partial \log f_1(\delta_{q+1}, r)}{\partial \delta_{q+1}} \geq 0 \text{ at } \delta_{q+1} = (1-b)/(q+2),\ r = 1/(q+2) \qquad (3.14)$$

If r_o solves $\frac{\partial \log f_1}{\partial \delta_{q+1}}\left(\frac{1-b}{q+2}, r\right) = 0$ then $r_o < r_1$ (3.15)

(3.11) and (3.12) show that maximum of $f_1\left(\frac{1-b}{q+2}, r\right)$ is attained at r_1. (3.13), (3.14) and (3.15) show that

$$\frac{\partial \log f_1}{\partial \delta_{q+1}}\left(\frac{1-b}{q+2}, r_1\right) < 0, \text{ i.e., for sufficiently small } \beta,$$

$$f_1\left(\frac{1-b}{q+2} - \beta, r_1\right) > f_1\left(\frac{1-b}{q+2}, r_1\right) \text{ which gives (3.10).}$$

We thus have to verify (3.11) - (3.15).

$$\frac{\partial \log f_1}{\partial r} = \log(1-b-\delta_{q+1}) + \log(1+qb-\delta_{q+1}) - 2\log((q-1)(1-b) + 2\delta_{q+1}) - [2\log r + 2 - 2(\log(1-2r) + 1)] \qquad (3.16)$$

$$\frac{\partial \log f_1}{\partial r}\left(\frac{1-b}{q+2}, r\right) = -\log \frac{r^2}{(1-2r)^2} + \log\left(\frac{1+b(q+1)}{q^2(1-b)}\right) \text{ which gives}$$

$$r_1 = \frac{1}{(2+qx_o)} \text{ where } x_o = \left[\frac{1-b}{1+b(q+1)}\right]^{1/2}.$$

From (3.16),

$$\frac{\partial^2 \log f_1}{\partial r^2} = -\left(\frac{2}{r} + \frac{4}{1-2r}\right) < 0 \text{ proving (3.11) and (3.12).}$$

$$\frac{\partial \log f_1}{\partial \delta_{q+1}} = \frac{-r}{1-b-\delta_{q+1}} - \frac{r}{1+qb-\delta_{q+1}} + \frac{2(1-2r)}{(q-1)(1-b)+2\delta_{q+1}} \tag{3.17}$$

Hence $\frac{\partial \log f_1}{\partial \delta_{q+1}}\left(\frac{1-b}{q+2}, r\right)$

$$= \frac{-r(q+2)}{(1-b)(q+1)} - \frac{r(q+2)}{(q+1)(1+b\ (q+1))} + \frac{2(1-2r)(q+2)}{q(q+1)(1-b)} \tag{3.18}$$

(3.13) immediately follows from this.

(3.14) can be verified easily from (3.18).

$$\frac{\partial \log f_1}{\partial \delta_{q+1}}\left(\frac{1-b}{q+2}, r\right) = \frac{q+2}{q+1}\left(\frac{-r}{1-b} - \frac{r}{1+b(q+1)} + \frac{2(1-2r)}{q(1-b)}\right).$$

Thus $r_o = \dfrac{1}{2+q(\frac{1}{2} + \frac{1}{2}x_o^2)} < r_1$ proving (3.15).

Now suppose that $q = 1$. Note that the relations (3.11) - (3.15) still hold. However (3.9) does not hold (since (3.7) is no longer true). Instead we have

$$g(\delta_{q+1}, r) = \left(\frac{1-b-\delta_{q+1}}{2\delta_{q+1}}\right)^{3r-1} > 1 \quad \text{if } r \text{ and } \delta_{q+1} \text{ are bounded}$$

away respectively from 1/3 and (1-b)/3.

From (3.16), $\forall\ \delta_{q+1}, r \leq r_1$, f_1 is an increasing function of r. (3.6) follows from the above observations.

Proof of (3.4): Suppose first that $q > 1$. By the strong law of large numbers, it is sufficient to show that a maximizing point of f_1 is contained in the region (for some fixed $0 < \beta < 1/2$).

$$\left\{(\delta_{q+1}, r) : EX_j \leq (r-\beta)N,\ j = 2, \ldots, (q+1) \text{ and } X_j\text{'s as in (3.2)}\right\}$$

$$= \left\{ (\delta_{q+1}, r) : \frac{\delta(1-2r)N}{(q-1)\delta+\delta_{q+1}} \leq (r-\beta)N,\ \frac{\delta_{q+1}(1-2r)N}{(q-1)\delta+\delta_{q+1}} \leq (r-\beta)N \right\}$$

$$= \left\{ (\delta_{q+1}, r) : \frac{\delta}{(q-1)\delta+\delta_{q+1}} \leq \frac{r-\beta}{1-2r} \right\} (\text{since } \delta_{q+1} \leq \delta)$$

$$= \left\{ (\delta_{q+1}, r) : \frac{\delta}{(q+1)\delta+\delta_{q+1}} \leq \frac{r-\beta}{1-2\beta} \right\}$$

$$\supseteq \left\{ (\delta_{q+1}, r) : \frac{\delta}{(q+1)\delta+\delta_{q+1}} \leq r-\beta \right\} = A \quad \text{(say)}.$$

Suffices to show that

$$\frac{\partial \log f_1}{\partial r} > 0 \quad \forall \quad (\delta_{q+1}, r) \in A^c.$$

As $\dfrac{\partial^2 \log f_1}{\partial r^2} < 0$, it suffices to show that

$$\frac{\partial \log f_1}{\partial r} \quad \text{at} \quad r = \frac{\delta}{(q+1)\delta+\delta_{q+1}} + \beta \qquad \text{is positive.}$$

The above expression is positive at $\beta = 0$ (follows from 3.16). By continuity it is positive if β is sufficiently small. This proves (3.4) and hence the lemma for $q > 1$. For $q = 1$, (3.4) is trivially satisfied. Thus the lemma is proved.

Remark 3.2: An approximate value of N in Theorem 3.1(ii) can be obtained by a more detailed analysis of the behaviour of f_1 and f_2. However, the details will be quite messy.

Remark 3.3(a): The case $a > 1$, $b > 0$. Assume N is large. In this case (as in the first step of (ii))

$$- a \frac{\partial \xi}{\partial \alpha} - \sum_{i=1}^{q} \frac{\partial \xi}{\partial \delta_i} + (a+q) \frac{\partial \xi}{\partial \delta_{q+1}}$$

$$= \sum_{i=1}^{q+1} \sum_{s=1}^{q+1} \sum_{\substack{L \in \Omega_s \\ N_o = N_i + 1}} (1 - a\delta_i/\alpha) L/s(s+1)$$

$$- (a-1) \sum_{s=2}^{q+2} \sum_{L \in \Omega_s} \frac{(s-1)L}{s} + (a+q+1)\left(\frac{\partial \xi}{\delta_{q+1}} - V\right)$$

$$= T_1 - T_2 + T_3 \quad \text{(say)} \qquad (3.19)$$

We have expressions $\bar{f}_1$ (corresponding to T_1 and T_2) and $\bar{f}_2$

(corresponding to T_3) similar to f_1 and f_2.

A similar argument shows that

$$\frac{\max \bar{f}_1}{\max \bar{f}_2} > 1, \quad \text{if} \quad a < 1.$$

Hence for large N, T_3 is negligible (w.r.t. T_1 and T_2) if $a < 1$. (3.20)

We show that (for certain values of a) T_2 dominates T_1. We are interested in the behaviour of T_1/T_2 only at the point of maximum of $\bar{f}_1$.

Note that $\bar{f}_1$ does not attain its maximum at $r = \frac{1}{(q+2)}$ or $\frac{1}{2}$. Thus at its maximum,

$$\frac{\partial \log \bar{f}_1}{\partial r} = 0, \text{ which gives}$$

$$\left(\frac{1}{r} - 2\right)^2 \frac{\alpha \delta}{((q-1)\delta + \delta_{q+1})^2} = 1 \qquad (3.21)$$

At the point $(a\delta + b, \delta, \ldots, \delta, \delta_{q+1}\ 0, \ldots, 0)$, the leading terms of T_1 and T_2 are respectively,

$$T_{1L} = \frac{b}{2\alpha} \sum_{i=1}^{q} \sum_{L \varepsilon \Omega_1,\, N_o = N_i + 1} L$$

and $$T_{2L} = \frac{(a-1)}{2} \sum_{i=1}^{q} \sum_{\substack{L \varepsilon \Omega_2 \\ N_o = N_i}} L \quad (\text{note that } \alpha = a\delta_i + b \ \forall\, i = 1, \ldots, q)$$

Hence $$T_{1L}/T_{2L} \leq \frac{r}{1-2r} \frac{(q-1)\delta + \delta_{q+1}}{\delta} \frac{b}{\alpha(a-1)}.$$

Using equation (3.21),

$$T_{1L}/T_{2L} \leq \max (\alpha/\delta)^{1/2} \max \frac{b}{\alpha(a-1)},$$

where the first max is over all maximizing points of $\bar{f}_1$. Observe that $(\alpha/\delta)^{1/2}$ is a decreasing function of α. (Recall $\alpha = a\delta + b$).

Thus $T_{1L}/T_{2L} \leq \frac{b}{a-1} \left[\frac{(a+b(k-1))(1-b)}{(a+k-1)^2} \right]^{-1/2}$. T_3 is negative.

Hence for large N, the L.F.C. will be given by (2.2) if

$$\frac{b}{a-1} \left[\frac{(a+b(k-1))(1-b)}{(a+k-1)^2} \right]^{-1/2} < 1.$$

<u>Remark 3.3(b)</u>: <u>The case $a < 1$, $b > 0$</u>. As in Remark 3.1(a), the condition $a + bk > 1$ is essential. In this case T_2 is negative in equation (3.19). By (3.20), T_3 is negligible w.r.t. T_1 and T_2. Thus $T_1 - T_2 + T_3 > 0$ for large N. Thus the L.F.C. (for large N) is given by $(a\delta + b, \delta, 0, \ldots, 0)$ i.e.,

$$\left(\frac{a+b}{a+1}, \frac{1-b}{a+1}, 0, \ldots, 0 \right).$$

4. <u>DISCUSSIONS</u>

Chen and Hwang (1984) consider the problem of choosing the t best cells $(t \geq 1)$ and state four conjectures on the nature of the L.F.C. In this section we provide partial answers to all the four conjectures for $t = 1$. For $t = 1$, the conjectures essentially takes the following forms with the constraint given by (1.3).

Conjecture I : For any N, the L.F.C. $\theta^*(N)$ is given by (2.1) with $\mu = 0$ and some $1 \leq q \leq k-1$.

Conjecture II : For any sample size $N \geq N(b)$ and any b, there exists a k_o such that the slippage configuration (given by (1.4)) is not a least favourable configuration if $k \geq k_o$.

Conjecture III : For any N, any k, there exists a δ_o such that the slippage configuration is a least favourable configuration if $0 < \delta < \delta_o$.

Conjecture IV : For each $k \geq 3$ and any δ, there exists an N_o such that the slippage configuration is not a least favourable configuration if the sample size $N \geq N_o$.

Conjecture I (for large N) and Conjecture IV follow directly from Theorem 3.1.

Look at the configuration (1.4) and assume that $k \to \infty$. Among the probabilities of different configurations, it is easy to see the dominating terms are the probabilities of the configurations

$$(N, 0, \ldots, 0)$$
$$(N-1, 1, \ldots, 0) \ldots \quad \ldots$$

... ... (1, 1, ..., 1, 0...0) and any permutation of these, keeping the first cell fixed. As $k \to \infty$, the other terms $\to 0$.

Thus $1 - PCS \simeq \frac{N-1}{N} \alpha \, \delta^{n-1} \binom{k-1}{N-1} N!$

$$= (N-1)\alpha \, \delta^{n-1} (k-1)!/(k-N)!$$

Using $\alpha = \dfrac{1+(k-1)b}{k}$, $\delta = \dfrac{1-b}{k}$, in this case

$$1\text{-PCS} \longrightarrow (N-1)(1-b)^{N-1} \; b = a_1 \text{ , say.} \qquad (4.1)$$

Using normal approximation, it is easy to see that for the configuration (1.5),

$$1\text{-PCS} \simeq P(N(0,1) \leq - N^{1/2} \, b(1-b^2)^{-1/2}) = a_2 \text{, say.} \qquad (4.2)$$

Thus $a_2 > a_1$ if

$$\exp(-N(b^2/(1-b^2) + \log(1-b))) > (2\pi N)^{1/2} (N-1) b^2 (1-b)^{-1/2} (1-b^2)^{-1/2} .$$

Let $g(b) = b^2/(1-b^2) + \log(1-b)$.

It can be checked that $g(0) = 0$ and $g'(0) < 0$. Thus $g(b) < 0$ for some b. Thus $a_2 > a_1$ for all large N. This shows that for large k, the "slippage configuration" (1.4) is not the LFC (for large N). Thus Conjecture II of Chen and Hwang is settled in the affirmative, for small b.

From the proof of Theorem 3.1 (ii), it can be seen that at the point $(\alpha, \delta, \ldots, \delta_{q+1})$ with $b = 0$ (i.e., $\alpha = \delta$),

$$\xi(\alpha, b) = - \frac{\partial \xi}{\partial \alpha} - \sum_{i=1}^{q} \frac{\partial \xi}{\partial \delta_i} + (q+2) \frac{\partial \xi}{\partial \delta_{q+1}}$$

$$= \sum_{s=1}^{q+1} \sum_{\substack{L \in \Omega_s \\ N_o = N_{q+1}+1}} (1-\delta_{q+1}/\alpha) L/s(s+1) - (q+2)K_{3(q+1)}$$

$$- (2+q) \sum_{s=1}^{q+1} \sum_{\substack{L \in \Omega_s \\ N_o = N_{q+1}+1}} L/s(s+1) < 0$$

Thus, using continuity, there exists a b_o such that $\inf_\alpha \xi(\alpha, b) < 0$ for all $b \leq b_o$. This shows that for small b, (1.4) is the L.F.C. This settles Conjecture III of Chen and Hwang.

+‡‡‡‡‡‡‡‡‡‡‡‡+

SECTION 4C : SELECTING THE LEAST PROBABLE CATEGORY

1. Introduction

In this section we consider the problem of selecting the least likely event in a multinomial population and use the procedure which selects the cell corresponding to the highest observed frequency with ties broken by randomisation.

Recall that $\theta_{(1)}$ and $\theta_{(2)}$ denote the lowest and next lowest cell probabilities, respectively. In this section we derive the L.F.C. with the preference zone given by

$$\theta_{(1)} \leq a\,\theta_{(2)} - c \qquad (1.1)$$

for different values of a, c > 0. The particular case with a = 1 and c > 0 was considered by Alam and Thompson (1972) and they found the L.F.C. to be the so called slippage configuration given by

$$\theta^* = \left(\frac{a-c(k-1)}{a+k-1}, \frac{1+c}{a+k-1}, \ldots, \frac{1+c}{a+k-1}\right) \qquad (1.2)$$

Alam and Thompson's proof is long and cumbersome, whereas our proof for this particular case is comparatively short and simple.

Actually for $a < 1$, $c > 0$, the L.F.C. $\theta^*(N)$ depends on N. The limiting value of the L.F.C. is derived for this in Theorem 3.1. It is found that the L.F.C is not the slippage configuration for certain combinations of a and c.

2. Preliminaries

In a random sample of size N from a k-cell multinomial population, let $X_{iN}(=X_i)$, $i = 0, 1, \ldots, (k-1)$ denote the frequency of the ith cell. For calculation of the PCS we assume w.l.g. that θ_1 and θ_2 are the lowest and the next lowest probabilities (i.e., $\theta_1 = \theta_{(1)}$ and $\theta_2 = \theta_{(2)}$). Note that

$$\begin{aligned}\xi(\theta_1, \ldots, \theta_k) &= P(X_o < X_\alpha \;\forall\; \alpha \neq 0) \\ &+ \frac{1}{2} \sum_{j>0} P(X_o = X_j, X_o < X_\alpha \;\forall\; \alpha \neq j, \alpha > 0) \\ &+ \ldots \\ &+ \frac{1}{k} P(X_o = X_1 = \ldots = X_{k-1}).\end{aligned}$$

The following result is due to Marshall and Olkin (1979) (Proposition C.2.b, page 400).

Theorem 2.1 : For fixed θ_1, ξ is a Schur-concave function of $(\theta_2, \ldots, \theta_k)$.

Remark 2.1 : Theorem A.1.1 and Theorem 2.1 together imply that the search for L.F.C. $\theta^*(N)$ can be confined to probability vectors of the form $(a\delta - c, \underbrace{\delta, \ldots, \delta}_{q}, \mu)$ where $a\delta - c \leq \delta \leq \mu$ (2.1)

3. The Main Results

Theorem 3.1 : Assume that $a - c(k-1) > 0$.

(i) If $a \geq 1$, $c > 0$, $a - ck < 1$, then for every N, the L.F.C. is given by (1.2).

(ii) If $a < 1$, $c > 0$ and $\frac{a-(k-1)c}{a+k-1} \le \frac{ac}{1-a}$ then, as $N \to \infty$, the L.F.C. is given by (1.2).

(iii) If $a < 1$, $c > 0$ and $\frac{a-(k-1)c}{a+k-1} \ge \frac{ac}{1-a}$, then, as $N \to \infty$, the L.F.C. is given by

$$\theta^*(N) = \left(\frac{ac}{1-a}, \frac{c}{a(1-a)}, \ldots, \frac{c}{a(1-a)}, 1 - \frac{c(a^2+k-2)}{a(1-a)}\right) \tag{3.1}$$

Remark 3.1(a) : The condition $a-c(k-1) > 0$ is natural. The condition $a-ck < 1$ in (i) is necessary to ensure that the cell with the lowest probability is unique.

(b) The result of Alam and Thompson (1972) is Theorem 3.1 (i) with $a = 1$.

Proof of theorem 3.1 : Consider any arbitrary probability vector of the form $(\alpha, \delta_1, \ldots, \delta_{q+1})$ where $\delta_i \le \delta_{q+1}$ and $\alpha \le a\delta_i - c \quad \forall \; i = 1, \ldots, q$. (3.2)

The derivative of ξ along the direction $(a, 1, \ldots, 1, -(a+q))$ is given by

$$a\frac{\partial \xi}{\partial \alpha} + \sum_{i=1}^{q} \frac{\partial \xi}{\partial \delta_i} - (a+q) \frac{\partial \xi}{\partial \delta_{q+1}} \tag{3.3}$$

$$= a\left(\frac{\partial \xi}{\partial \alpha} - V\right) + \sum_{i=1}^{q} \left(\frac{\partial \xi}{\partial \delta_i} - V\right) - (a+q)\left(\frac{\partial \xi}{\partial \delta_{q+1}} - V\right) \tag{3.4}$$

where V is defined later.

We now define the following quantities, whose relevance is self-evident. $N_o, N_1, \ldots, N_{q+1}$ shall denote non-negative integers.

$$L(\underset{\sim}{N},\alpha,\underset{\sim}{\delta}) = \frac{N!}{\prod_{i=0}^{q+1} N_i!} \alpha^{N_o} \prod_{i=1}^{q+1} \delta_i^{N_i} = L, \quad \text{say,} \quad (\text{Here } \sum_{i=0}^{q+1} N_i = N-1)$$

Let $M(\underset{\sim}{N})$ = Second lowest in $\{N_i,\ i = 0, 1, \ldots, q+1\}$, ignoring ties

= M, say.

$s(\underset{\sim}{N})$ = # N_i's which are equal to M (excluding the lowest and ignoring ties) = s, say.

$t(\underset{\sim}{N})$ = # N_i's equal to N_o (including N_o) = t, say.

$$A_{\alpha 1} = \bigcup_{s=1}^{k-1} \left\{ \frac{L\alpha}{(s+1)(N_o+1)} : N_o = M-1 \right\}$$

$$A_{\alpha 2} = \left\{ \frac{L\alpha}{N_o+1} : N_o < M-1 \right\}, \quad A_\alpha = A_{\alpha 1} \cup A_{\alpha 2}$$

$$A_{i1} = \bigcup_{s=1}^{k-1} \left\{ \frac{L\delta_i}{(N_i+1)(s+1)} : N_o = N_i+1,\ N_j \geq N_o \ \forall\ j \neq i \right\}$$

$$A_{i2} = \bigcup_{s=1}^{k-1} \left\{ \frac{L\delta_i}{(N_i+1)s} : N_o = N_i,\ N_j \geq N_o \ \forall\ j \neq i \right\}$$

$$A_{i3} = \bigcup_{t=1}^{k-1} \left\{ \frac{L\delta_i}{t(N_i+1)} : N_o < N_i,\ N_j \geq N_o \ \forall\ j \neq i \right\}$$

$$A_i = \bigcup_{j=1}^{3} A_{ij}.$$

$$V = \sum_{s=1}^{k-1} \sum_{N_o=M} L/(s+1) + \sum_{N_o \leq M-1} L \tag{3.5}$$

Note that V can also be written as (for every i)

$$V = \sum_{s} \sum_{\substack{N_o=N_i \\ N_j \geq N_o \ \forall j \neq i}} L/(s+1) + \sum_{t} \sum_{\substack{N_o<N_i \\ N_o \leq N_j \ \forall j \neq i}} L/t \tag{3.6}$$

$$\frac{\partial \xi}{\partial \alpha} = \sum_{x \in A_\alpha} \frac{\partial x}{\delta \alpha} = \sum_{x \in A_{\alpha 1}} \frac{\partial x}{\delta \alpha} + \sum_{x \in A_{\alpha 2}} \frac{\partial x}{\delta \alpha}$$

$$= \sum_{s} \sum_{N_o=M-1} L/(s+1) + \sum_{N_o<M-1} L$$

$$= B_1 + B_2, \text{ say.} \tag{3.7}$$

$$\frac{\partial \xi}{\partial \delta_i} = \sum_{x \in A_i} \frac{\partial x}{\partial \delta_i} = \sum_{x \in A_{i1}} \frac{\partial x}{\partial \delta_i} + \sum_{x \in A_{i2}} \frac{\partial x}{\partial \delta_i} + \sum_{x \in A_{i3}} \frac{\partial x}{\partial \delta_i}$$

$$= \sum_{s} \sum_{\substack{N_o=N_i+1 \\ N_j \geq N_o \ \forall j \neq i}} L/(s+1) + \sum_{s} \sum_{\substack{N_o=N_i \\ N_j \geq N_o \ \forall j \neq i}} L/s + \sum_{t,} \sum_{N_o<N_i, N_o \leq N_j \ \forall j \neq i} L/t$$

$$= C_{1i} + C_{2i} + C_{3i}, \text{ say.} \tag{3.8}$$

Thus using equations (3.5) and (3.7),

$$\frac{\partial \xi}{\delta \alpha} - V = -\sum_{s} \sum_{N_o=M} L/(s+1) - \sum_{s} \sum_{N_o=M-1} Ls/(s+1)$$

$$= I_1 + I_2, \text{ say.}$$

Using equations (3.6) and (3.8),

$$\frac{\partial \xi}{\partial \delta_i} - V = \sum_s \sum_{\substack{N_o = N_i \\ N_j \geq N_o \,\forall\, j}} L/s(s+1) + \sum_s \sum_{\substack{N_o = N_i + 1 \\ N_j \geq N_o \,\forall\, j \neq i}} L/(s+1)$$

$$= J_{1i} + J_{2i}, \text{ say.}$$

Similarly $\dfrac{\partial \xi}{\partial \delta_{q+1}} - V = K_1 + K_2$, say.

Thus the expression in (3.4) is

$$a(I_1 + I_2) + \sum_{i=1}^{q} (J_{1i} + J_{2i}) - (a+q)(K_1 + K_2).$$

$$aI_2 + \sum_{i=1}^{q} J_{2i} = -a \sum_{i=1}^{q+1} \sum_s \sum_{\substack{N_o = N_i - 1 \\ N_j \geq N_i = M \,\forall\, j \neq i}} L/(s+1)$$

$$+ \sum_{i=1}^{q} \sum_s \sum_{\substack{N_i = N_o - 1 \\ N_j \geq N_o \,\forall\, j \neq i}} L'/(s+1)$$

Consider two typical terms of the summation when $i \neq q+1$. Note that L' is obtained from L by interchanging the role of N_o and N_i (i.e., α and δ_i). Thus at any point satisfying (2.1)

$$L' - aL = (\alpha/\delta_i - a)L = -\frac{c}{\delta} L.$$

Thus $\quad aI_2 + \sum_{i=1}^{q} J_{2i} = -\dfrac{c}{\delta} \sum_{i=1}^{q} \sum_s \sum_{N_o = N_i - 1 = M-1} L/(s+1) - \dfrac{a}{\alpha} \delta_{q+1} K_2$

$$= -\frac{c}{\delta} q L_1 - \frac{a}{\alpha} \delta_{q+1} K_2, \text{ say.} \tag{3.9}$$

Further note that $I_1 = -(\sum_{i=1}^{q} J_{1i} + K_1)$. Thus

$$aI_1 + \sum_{i=1}^{q} J_{1i} = (1-a) \sum_{i=1}^{q} J_{1i} - aK_1$$

$$= (1-a)qJ_{11} - aK_1 \tag{3.10}$$

Thus the expression (3.4) is

$$E = -\frac{c}{\delta} qL_1 + (1-a)qJ_{11} - (\frac{a\delta_{q+1}}{\alpha} + a+q)K_2 - (2a+q)K_1 \tag{3.11}$$

Note that $L_1, K_1, K_2 \geq 0$. Thus if $a \geq 1$, $c > 0$ then $(3.11) < 0$. This proves (i) of theorem 3.1.

Note that

$$J_{11} = \Sigma\, P(X_o = X_1 = N_o / X_o + X_1 = 2N_o)\, P(\sum_{j=2}^{q+1} X_j = N-2N_o) \times$$

$$\Sigma\, P(X_j = N_j,\ j = 2, \ldots, (q+1))\, s(N_o, N_2, \ldots, N_{q+1}) / P(\sum_{j=2}^{q+1} X_j = N-2N_o)$$

where the first summation is over all N_o, $N_o \leq N/(q+2)$ and, for fixed N_o, the second summation is over all N_j's such that

$\sum_{j=2}^{q+1} N_j = N-2N_o$, $N_j \geq N_o \ \forall\ j$. $s(N_o, N_2, \ldots, N_{q+1}) = 1/s(s+1)$

if exactly $(s-1)$ of the N_j's $= N_o$. Thus

$$J_{11} = \sum_{0 \le r \le 1/(q+2)}^{*} P(X_o = X_1 = Nr / X_o + X_1 = 2Nr) P(\sum_{j=2}^{q+1} X_j = N-2Nr).$$

$f(r, \alpha, N)$, say.

where (*) indicates that the summation is over all r's such that Nr is an integer.

From the relations $\alpha + q\delta + \delta_{q+1} = 1$, $\alpha = a\delta - c$, we have

$$\delta = (\alpha + c)/a, \; \delta_{q+1} = (a - (q+a)\alpha - qc)/a, \; (q-1)\delta + \delta_{q+1} = (a - (1+a)\alpha - c)/a$$

Thus
$$J_{11} = \sum_r \frac{N!\; f(r, \alpha, N)}{(Nr)!(Nr)!((1-2r)N)!} (\frac{a\alpha}{a})^{Nr} (\frac{\alpha+c}{a})^{Nr} (\frac{a-(1+a)\alpha-c}{a})^{N(1-2r)}$$

$= \Sigma_1 + \Sigma_2$ (say) where Σ_1 denotes summation over $r > \varepsilon$ (ε fixed) and Σ_2 denotes the remaining part.

Note that Σ_2 is negligible (as compared to Σ_1) as $N \to \infty$.

By Stirling's formula,

$$\Sigma_1 \simeq \Sigma_1 \frac{f(r,\alpha,N)[r^{-2r}(1-2r)^{-(1-2r)}]^N [(a\alpha)^r(\alpha+c)^r(a-(1+a)\alpha-c)^{1-2r}]^N}{2\pi a^N \, Nr(1-2r)^{1/2}}$$

$$= \Sigma_1 \frac{f(r, \alpha, N)}{2\pi a^N \, Nr(1-2r)^{1/2}} f_1^N(\alpha, r), \quad \text{say}$$

$$= \Sigma_1 \; t(r, \alpha, N) f_1^N(\alpha, r), \quad \text{say} \qquad (3.12)$$

In an exactly similar way it follows that

$$\Sigma_1 \simeq \Sigma_1 \frac{\bar{f}(r,\alpha,N) f_1^N(\alpha,r)(1-2r)\delta}{2\pi a^N N r(1-2r)^{1/2} r((q-1)\delta+\delta_{q+1})} \text{ and } \Sigma_2 \text{ is negligible} \tag{3.13}$$

We now find the point at which f_1 attains its maximum. First fix α.

$$\frac{\partial \log f_1}{\partial r} = \log\left(\frac{1}{r} - 2\right)^2 + \log \frac{a\alpha(\alpha+c)}{(a-(a+1)\alpha-c)^2} \tag{3.14}$$

Easy calculations give

$$\frac{\partial^2 \log f_1}{\partial r^2} \leq 0$$

$$\frac{\partial \log f_1}{\partial r} < 0 \quad \text{at} \quad r = 1/(q+2) \cdot \frac{\partial \log f_1}{\partial r} > 0 \quad \text{at } r \text{ around zero.}$$

Thus for fixed α, maximum of f_1 w.r.t. r is attained at

$$r_o (= r_o(\alpha)) \quad \text{where} \quad \left(\frac{1}{r_o} - 2\right)^2 \frac{\alpha\delta}{((q-1)\delta+\delta_{q+1})^2} = 1 \tag{3.15}$$

It easily follows that

$$\log f_1(\alpha, r_o) = \log \frac{a-(1+a)\alpha-c}{1-2r_o} = b(\alpha), \quad \text{say.}$$

Using (3.15) it can be easily checked that

$$b(\alpha) = \log(c(\alpha)-A) \quad \text{where}$$

$$c(\alpha) = 2[a\alpha(\alpha+c)]^{1/2} - (a+1)\alpha \quad \text{and} \quad A = c-a.$$

Thus $c''(\alpha) < 0$.

Further $c'(\alpha) = 0$ implies

$$[\alpha/(\alpha+c)]^{1/2} + [(\alpha+c)/\alpha]^{1/2} = a^{1/2} + a^{-1/2}.$$

Recall that $a < 1$ and $c > 0$. Thus the solution for α is given by

$$c/\alpha = (1-a)/a, \quad \text{i.e.} \quad \alpha_o = ac/(1-a) \tag{3.16}$$

Hence $a = (\alpha_o/\delta_o)^{1/2}$ (3.17)

Note that the range of possible values of α is $[0, \frac{a-(q+1)c}{a+q+1}]$.

When $\frac{ac}{1-a} < \frac{a-(q+1)c}{a+q+1}$, the maxima of f_1 is attained at (α_o, r_o) where r_o solves (3.15) with $\alpha = \alpha_o$.

When $\frac{ac}{1-a} \geq \frac{a-(q+1)c}{a+q+1}$, the maximum of f_1 is attained at $(\frac{a-(q+1)c}{a+q+1}, r_o)$ where r_o solves (3.15) with $\alpha = \frac{a-(q+1)c}{a+q+1}$.

By strong law of large numbers, in the following region R, $f(r, \alpha, N)$ and $\bar{f}(r, \alpha, N) \rightarrow \frac{1}{2}$.

$$R = \left\{(\delta, r) : \frac{\delta}{(q-1)\delta+\delta_{q+1}} N(1-2r) \geq Nr\right\} \tag{3.18}$$

$$= \left\{(\delta, r) : (\frac{1}{r} - 2) \geq \frac{a-(1+a)\alpha-c}{\alpha+c}\right\}$$

Note that from (3.14) $\frac{\partial \log f_1}{\partial r}$ is a decreasing function of r and at

$$\frac{1}{r} - 2 = \frac{a-(1-a)\alpha-c}{\alpha+c}, \quad \frac{\partial \log f_1}{\partial r} = \log \frac{ac}{\alpha+c} < 0 .$$

Hence the point of maximum of f_1 (for each fixed α) lies in R.

From (3.12) and (3.13) we have, for large N,

$$- \frac{c}{\delta}L_1 + (1-a)J_{11} \simeq \Sigma_1 [- \frac{c}{\delta} \frac{1-2r}{r} \frac{\delta}{(q-1)\delta+\delta_{q+1}} + (1-a)] t(r,\alpha,N) f_1^N(\alpha,r)$$

$$\simeq \sum_{r_o-\varepsilon \le r \le r_o+\varepsilon} m(r,\alpha) t(r,\alpha,N) f_1^N(\alpha,r) \qquad (3.19)$$

Suppose now that $\frac{a-(k-1)c}{a+k-1} < \frac{ac}{1-a}$.

Note that $m(r_o(\alpha),\alpha) = - \frac{c}{\delta} \frac{(1-2r_o)}{r_o} \frac{\delta}{(q-1)\delta+\delta_{q+1}} + (1-a)$

$$= - \frac{c}{(\alpha\delta)^{1/2}} + (1-a) < 0 \ \forall \ \alpha < \frac{ac}{1-a} \qquad (3.20)$$

Hence for sufficiently small $\varepsilon > 0$, in the region $r_o(\alpha) - \varepsilon \le r \le r_o(\alpha) + \varepsilon$, m is uniformly (over α) negative.

Noting that $K_1, K_2 \ge 0$, it follows therefore that (3.4) < 0. This proves (ii).

We now proceed to prove the last part of the theorem. So we assume

$$\frac{a-(k-1)c}{a+k-1} \ge \frac{ac}{1-a} .$$

We now compare $\int_{\eta_1}^{\eta_2} (- \frac{c}{\delta} L_1 + (1-a)J_{11}) d\alpha$, $\int_{\eta_1}^{\eta_2} \frac{\delta_{q+1}}{\alpha} K_2(\alpha) d\alpha$ and

$$\int_{\eta_1}^{\eta_2} K_1(\alpha) d\alpha . \qquad (3.21)$$

Note that $\forall\ \eta_1 < \eta_2 < \frac{ac}{1-a}$, the first integral is negative for large N (use equation (3.19) and (3.20)).

$\forall\ \frac{ac}{1-a} < \eta_1 < \eta_2 < \frac{a-(k-1)c}{a+k-1}$, the first integral is by the same reasoning, positive.

So if we could show that in this case the first integral dominates the other two then it would be proved that (3.1) is the L.F.C. We proceed to show this now. We tackle only

$\int_{\eta_1}^{\eta_2} K_1(\alpha)d\alpha$. $\int_{\eta_1}^{\eta_2} \frac{\delta_{q+1}}{\alpha} K_2(\alpha)d\alpha$ can be tackled along

similar lines (note that $\alpha \geq \eta_1 \geq \frac{ac}{1-a}$). Analogous to L_1 and J_{11}, $K_1 \leq \sum_{r>\varepsilon} t_1(r,\alpha,N) f_2^N(\alpha,r)$ where

$$f_2(\alpha,r) = r^{-2r}(1-2r)^{-(1-2r)}(a\alpha)^r(a-(q+a)\alpha-qc)^r(q(\alpha+c))^{1-2r}$$

and $\quad t_1(r,\alpha,N) = \dfrac{1}{2\pi a^N Nr(1-2r)^{1/2}}$.

$$\text{Thus}\quad f_1/f_2 = \left(\frac{\delta}{\delta_{q+1}}\right)^r [((q-1)\delta + \delta_{q+1})/q\delta]^{1-2r}$$

$$= x^r\left[\frac{(q-1)+1/x}{q}\right]^{1-2r} \quad \text{where } x = \delta/\delta_{q+1}.$$

$$= g(x), \quad \text{say.}$$

Note that x is an increasing function of α and increases to 1. $g(1) = 1$.

g is a strictly decreasing function of x for $x < 1$.

Hence $\dfrac{f_1(\alpha,r)}{f_2(\alpha,r)} \geq t(\alpha)$ for some strictly decreasing t

$$\text{where} \quad t(\alpha) > 1 \quad \forall \quad \alpha < \frac{a-(k-1)c}{a+k-1} \qquad (3.22)$$

Note that $\displaystyle\max_{\substack{0 \leq \alpha \leq \frac{a-(k-1)c}{a+k-1} \\ \varepsilon \leq r \leq 1/(q+2)}} f_1$ is attained in the interior (follows from the study of f_1 given before).

Hence in this case, by (3.22)

$$\max_{\substack{\varepsilon \leq r \leq 1/(q+2) \\ \eta_1 \leq \alpha \leq \eta_2}} f_1 > \max_{\substack{\varepsilon \leq r \leq 1/(q+2) \\ \eta_1 \leq \varepsilon \leq \eta_2}} f_2 \quad \forall \quad \eta_1 < \eta_2 \leq \frac{a-(k-1)c}{a+k+1}.$$

That $\int_{\eta_1}^{\eta_2} K_1(\alpha)d\alpha$ is negligible w.r.t. the first integral of (3.21) follows now from the following lemma, proving the theorem completely.

<u>Lemma</u>: $\displaystyle\lim_{N \to \infty} \Big[\int_a^b \sum_{\varepsilon < r \leq 1/(q+2)} \frac{1}{N} t(r,\alpha,N)m(r,\alpha)u^N(r,\alpha)\Big]^{1/N} = u(r^*,\alpha^*)$

where r^*, α^* is the maximizer of $u(r,\alpha)$ in the region $(a,b) \times (\varepsilon, 1/(q+2))$, $t(r,\alpha,N)$ is bounded away from 0 in a region containing (r^*, α^*) and m is positive and continuous.

We omit the proof of this lemma, since this is an extension of the well known result of convergence of L^p norms (as $p \to \infty$) to the L^∞ norm.

SECTION 4D : REFERENCES

[1] Alam, K. and Thompson, J.R. (1972): On selecting the least probable multinomial event. Ann. Math. Statist. 43, 1981 - 1990.

[2] Bhandari, S.K. and Bose, A. (1985): On selecting the least likely event. Tech. Rep. No.34/85, Stat-Math. Division, Indian Statistical Institute, Calcutta.

[3] Bhandari, S.K. and Bose, A. (1987): On selecting the most likely event. Journal of Statistical Planning and Inference (to appear).

[4] Bhandari, S.K. and Bose, A. (1987): On selecting the t-best cells in a multinomial distribution. Tech. Rep. No.87 - 19, Dept. of Stat., Purdue Univ., U.S.A.

[5] Bhandari, S.K. and Ray Chaudhuri, A. (1988): On two conjectures about two-stage selection procedures. Sankhya, Ser. A, Vol. 50. (to appear).

[6] Bechhofer, R.E., Elmaghraby, S. and Morse, N. (1959): A single-sample multiple-decision procedure for selecting the multinomial event which has the highest probability. Ann. Math. Stat. 30, 102 - 119.

[7] Chen, P. (1986): On the least favourable configuration in multinomial selection problems. Commun. Statist. - Theor. Math. 15(2), 367 - 385.

[8] Chen, R.W. and Hwang, F.K. (1984): Some theorems, counter examples, and conjectures in multinomial selection theory. Commun. Statist. - Theor. Meth. 13(10), 1289 -1298.

[9] Kesten, H. and Morse, N. (1959): A property of the multinomial distribution. Ann. Math. Stat. 30, 120 -127.

[10] Marshall, A.W. and Olkin, I. (1979): Inequalities : Theory of majorization and its applications. Academic Press.

CHAPTER 5

LORENZ - DOMINANCE AND MEASURING INCOME INEQUALITY

SECTION 5A : INTRODUCTION

It has already been defined in the introduction that a non-negative random variable X is said to be Lorenz - dominated by another non-negative random variable Y, written as $X \prec_L Y$, if

$$\frac{1}{E(X)} \int_0^{M_{p,F}} x\, dF(x) \geq \frac{1}{E(Y)} \int_0^{M_{p,G}} x\, dG(x),$$

for all $0 < p < 1$, where F and G are the distribution functions of X and Y, respectively, and $M_{p,F}$ is the pth fractile of F.

In the non-stochastic set-up the above definition is equivalent to $x \prec y$, where $x = (x_1, \ldots, x_n)$, $y = (y_1, \ldots, y_n)$ and $\sum_{i=1}^{n} x_i = \sum_{i=1}^{n} y_i$.

Different measures of inequality have been suggested in the literature [see Nygard and Sandstrom (1981), Ord et al (1983), Marshall and Olkin (1979), Sen (1973), Bhandari (1986)] which preserve the partial order of Lorenz - dominance.

Section B generalises some already existing scattered results on characterisation of the parent distribution by inequality measures on its truncations. Section C develops some inequalities among different Schur-convex functions which are mostly used as quantitative measures of income inequality.

SECTION 5B: APPLICATIONS IN CHARACTERISATION OF THE PARENT DISTRIBUTION BY INEQUALITY MEASURES ON ITS TRUNCATIONS

1. Introduction

It is shown in this section that if for some measures of inequality (in income), the upper α-truncated distributions corresponding to two income distributions F and G have the same inequality measure for every α in (0,1), then F and G are equal except for possible change in scale. The specific inequality measures considered in this paper are Gini-index, coefficient of variation, measures derived from Mellin transform, and Dalton's measure. We use the concept of Lorenz - dominance to prove this.

Bhattacharya (1963) and Schelling (1934) [see Piesch (1975)] have proved independently that a necessary and sufficient condition for an arbitrary lower truncation to leave the Lorenz curve unchanged, is that the continuous density function has the Pareto form with index greater than 1. Ord et al (1983) have shown that if the Gini-index or H-index (based on Mellin's transform) is invariant for all upper truncations, then the parent distribution is Pareto. This generalizes the result of Schelling (1934) and Bhattacharya (1963).

Our results thus generalize the results of Ord et al (1983), as well as those of Bhattacharya (1963) and Schelling (1934).

2. The Main Result

Let F be the distribution function of a non-negative random variable X, and F_α be the distribution function of X, given $X \geq Z_\alpha(F)$, where $Z_\alpha(F)$ is the upper α-quantile $(0 < \alpha < 1)$ of F. Let $\mu(F) > 0$ be the mean of the distribution F. We shall assume throughout that the inequality measure $I(F)$ for any distribution F is scale-invariant. For the following theorem we have considered $I(F)$ to be any one of the following : Coefficient of variation, Gini-index, measures derived from Mellin transform, and Dalton's measure. [see Nygard and Sandstrom (1981), Ord et al (1983), Marshall and Olkin (1979).] Note that the above inequality indices are special cases of the following general functional form, or related to this form by one-to-one correspondence :

(2.1) $$I(F) = \int_0^\infty S_F(x)\, dF(x)/T[\mu(F)],$$

where $S_F(x)$ is either $\int_x^\infty t\, dF(t)$, or a suitably chosen strictly convex function S of x, and T is some suitable function.

Theorem 2.1 : If for any two distribution functions F and G on $(0, \infty)$

(2.2) $$I(F_\alpha) = I(G_\alpha)$$

for all $0 < \alpha < 1$, then G is a scale-transform of F. Conversely, if G is a scale-transform of F, then $I(F_\alpha) = I(G_\alpha)$ for all $0 < \alpha < 1$

Proof: Suppose, for distributions F and G with $\mu(F) > 0$ and $\mu(G) > 0$, we have $I(F_\alpha) = I(G_\alpha)$ for all $0 < \alpha < 1$, but F is not a scale-transform of G. Without any loss of generality, we may assume that $\mu(F) = \mu(G) = 1$, since $I(F_\alpha)$ is not affected by a scale transformation of F.

Let

$$(2.3) \qquad L_F(\alpha) = \int_0^{Z_\alpha(F)} t \, dF(t)/\mu(F).$$

Since F is not a scale-transform of G by our assumption, the Lorenz curves corresponding to F and G will be different. Note that the set of all points $\alpha \varepsilon [0, 1]$ for which $L_F(\alpha) = L_G(\alpha)$ is closed. Since $L_F(\alpha)$ is continuous in α, we can get α_1 and α_2, $0 \leq \alpha_1 < \alpha_2 \leq 1$ such that

$$(2.4) \qquad L_F(\alpha) \neq L_G(\alpha) \quad \text{for all} \quad \alpha \varepsilon (\alpha_1, \alpha_2),$$

$$L_F(\alpha_i) = L_G(\alpha_i) \quad \text{for} \quad i = 1, 2.$$

Without loss of generality let us assume

$$(2.5) \qquad L_F(\alpha) < L_G(\alpha), \text{ for all } \alpha \varepsilon (\alpha_1, \alpha_2)$$

since $L_F(\alpha) - L_G(\alpha)$ has the same sign in (α_1, α_2).

Note that (2.4) implies

$$(2.6) \qquad \mu(F_{\alpha_i}) = \mu(G_{\alpha_i}), \quad \text{for} \quad i = 1, 2.$$

Thus, we must have

$$(2.7) \qquad \int_{Z_{\alpha_i}(F)}^{\infty} S_F(x)\ dF(x) = \int_{Z_{\alpha_i}(G)}^{\infty} S_G(x)\ dG(x),\ i = 1, 2.$$

The relation (2.7) implies

$$(2.8) \qquad \int_{Z_{\alpha_2}(F)}^{Z_{\alpha_1}(F)} S_F(x)\ dF(x) = \int_{Z_{\alpha_2}(F)}^{Z_{\alpha_1}(G)} S_G(x)\ dG(x).$$

Case I: Suppose now

$$(2.9) \qquad S_F(x) = \int_x^{\infty} t\ dF(t).$$

Then assuming F and G to be continuous, (2.5) contradicts (2.8); note that the Lorenz curves corresponding to F and G cannot both become straight lines in (α_1, α_2).

Case II: Suppose $S_F(x)$ is a strictly convex function S of x.

Let X and Y denote random variables with distribution functions F and G, respectively. Now note that the conditional distribution of X, given $Z_{\alpha_2}(F) \leq X \leq Z_{\alpha_1}(F)$, Lorenz-dominates the conditional distribution of Y, given $Z_{\alpha_2}(G) \leq Y \leq Z_{\alpha_1}(G)$. This follows from (2.4) and (2.5); as a matter of fact, (2.4) implies that the above conditional distributions have the same mean. Now, it follows that (see Chapter 1):

$$(2.10) \qquad \int_{Z_{\alpha_2}(F)}^{Z_{\alpha_1}(F)} S(x)\ dF(x) > \int_{Z_{\alpha_2}(G)}^{Z_{\alpha_1}(G)} S(x)\ dG(x),$$

which contradicts (2.8) with $S_F = S_G = S$.

Remark 2.1: It follows from the above proof that an analogous result holds also for lower truncations.

Remark 2.2: Consider the following density on (0 , c) :

(2.11) $$f(x) = \frac{\beta+1}{1+\beta} x^{\beta}, \quad -1 < \beta < \infty .$$

Note that all the indices considered above for this density are invariant with respect to lower truncations. Hence, any continuous distribution for which the above inequality indices are invariant with respect to lower truncation has the density of form (2.11).

Remark 2.3: It is an interesting problem to find a necessary and sufficient condition for C_α, $0 < \alpha < 1$ to correspond to $I(F_\alpha)$, $0 < \alpha < 1$ for some given inequality measure I.

++++++++++++++

SECTION 5C : SOME RELATIONS AMONG INEQUALITY MEASURES

1. Introduction

This section develops a number of inequalities among different Schur-convex functions which are mostly used as quantitative measures of income inequality. In most cases the inequalities obtained can not be made more stringent. The results are expected to be helpful in exploring possible inter-relationships among the measures.

With non-negative observations $x_1, x_2, \ldots, x_n$, having positive arithmatic mean $\bar{x}$, the following inequality measures (cf. Bhandari (1986), Marshall and Olkin (1979), Ord at. al. (1983), Sen (1973) among others) have been considered in this paper. Note that these measures are all normalized i.e., They are equal to zero when $x_1, x_2, \ldots, x_n$ are all equal.

(i) Gini's coefficient:

$$G = G(x_1, \ldots, x_n) = \sum_{i<j=1}\sum |x_i - x_j|/(n^2\bar{x}),$$

(ii) Coefficient of variation:

$$C = C(x_1, \ldots, x_n) = [\sum_{i=1}^{n} (x_i - \bar{x})^2/(n\bar{x}^2)]^{1/2},$$

(iii) Measure derived from Mellin transformation:

$$H_\lambda = H_\lambda(x_1, \ldots, x_n) = \sum_{i=1}^{n} (x_i/\bar{x})^\lambda/n \quad - 1 (\lambda > 1),$$

(iv) Theil's entropy measure :

$$T = T(x_1, \ldots, x_n) = [\sum_{i=1}^{n} (x_i/\bar{x}) \log(x_i/\bar{x})]/n .$$

For the sake of simplicity in presentation, we have taken a form of H_λ in (iii) above, which is a strictly increasing function of the conventional form (Ord et. al. (1983)).

2. Some lemmas

The distributions under consideration are non-degenerate.

Lemma 2.1: $(n-1) \sum\sum_{i<j=1}^{n} (x_i - x_j)^2 \leq (\sum\sum_{i<j=1}^{n} |x_i - x_j|)^2$

$$\leq \frac{1}{3}(n^2 - 1) \sum\sum_{i<j=1}^{n} (x_i - x_j)^2 .$$

Proof: Without loss of generality, let $x_1 \leq x_2 \leq \cdots \leq x_n$. Let $t = n-1$, $\alpha_i = x_{i+1} - x_i (1 \leq i \leq t)$, $\underline{\alpha} = (\alpha_1, \ldots, \alpha_t)'$. Then it can be seen that

$$\sum\sum_{i<j=1}^{n} (x_i - x_j)^2 = \underline{\alpha}'M\underline{\alpha}, \quad \sum\sum_{i<j=1}^{n} |x_i - x_j| = \underline{\ell}'\underline{\alpha}, \qquad (2.1)$$

where $\underline{\ell} = (\ell_1, \ldots, \ell_t)'$, $\ell_i = i(n-i)(1 \leq i \leq t)$ and $M^{(t \times t)} = ((m_{ij}))$ is a symmetric matrix with $m_{ij} = i(n-j)(1 \leq i \leq j \leq t)$. One may check that M is positive definite (p.d.) and M^{-1} has entries $\frac{2}{n}$ along the principal diagonal, $-\frac{1}{n}$ just below and above the principal diagonal and 0 elsewhere.

Since M is p.d., by (2.1) and Cauchy-Schwarz inequality (Rao (1973, p.60)),

$$\left(\sum_{i<j=1}^{n}\sum |x_i - x_j|\right)^2 = (\ell'\alpha)^2 \leq (\ell' M^{-1}\ell)(\underline{\alpha}' M \underline{\alpha})$$

$$= (\ell' M^{-1}\ell) \sum_{i<j=1}^{n}\sum (x_i - x_j)^2 .$$

After considerable algebra, $\ell' M^{-1}\ell = \frac{1}{3}(n^2-1)$ and the right-hand inequality follows. Also noting that

$$\ell_i \ell_j \geq (n-1) m_{ij} \quad (1 \leq i, j \leq t),$$

the left-hand inequality follows immediately from (2.1) and the fact that $\underline{\alpha} \geq \underline{0}$.

Remark: The right-hand inequality in Lemma 2.1 attains equality iff $\alpha_1 = \alpha_2 = \ldots = \alpha_t$ i.e., if $x_1, x_2, \ldots, x_n$ are equispaced which means

$$x_i = x_1 + (i-1)(x_n - x_1)/(n-1) \quad (i = 1, 2, \ldots, n).$$

The left-hand inequality attains equality iff $x_1 = x_2 = \ldots = x_{n-1} \leq x_n$.

Lemma 2.2: Let $S = \left\{(x_1, \ldots, x_n) : x_i \geq 0 \ (1 \leq i \leq n), \ \bar{x} > 0\right\}$. Then

$$1 \leq \left(\sum_{i=1}^{n} (x_i/\bar{x})^{\mu}\right) / \left(\sum_{i=1}^{n} (x_i/\bar{x})^{\lambda}\right) \leq n^{\mu-\lambda},$$

provided $(x_1, \ldots, x_n) \in S$ and $1 < \lambda < \mu$.

Proof: The right-hand inequality follows trivially as

$(x_i/\bar{x})^{\mu} \leq n^{\mu-\lambda}(x_i/\bar{x})^{\lambda}$ $(1 \leq i \leq n)$. To prove the left-hand inequality, let $f(\xi) = \Sigma_{i=1}^{n}(x_i/\bar{x})^{\xi}$. Observe that for $(x_1, \ldots, x_n) \in S$, $f'(1) = \Sigma_{i=1}^{n}(x_i/\bar{x}) \log (x_i/\bar{x}) \geq 0$ (by Jensen's inequality) and $f''(\xi) \geq 0$ for $\xi \geq 1$. Hence $f(\xi)$ is non-decreasing in ξ for $\xi \geq 1$, so that $f(\mu) \geq f(\lambda)$, completing the proof.

Remark: The right-hand inequality in Lemma 2.2 attains equality iff among $x_1, \ldots, x_n$ exactly one is positive while the rest equal 0. The left-hand inequality attains equality iff $x_1, \ldots, x_n$ are all equal.

3. Main results

The above lemmas will be applied in this section to derive inequalities among the measures considered in Section 1. The notation is as before and $x_1, \ldots, x_n$ are non-negative observations with $\bar{x} > 0$.

Theorem 3.1: $(n-1)^{1/2}n^{-1}C \leq G \leq ((n^2-1)/3)^{1/2}n^{-1}C$.

Theorem 3.2: For $1 < \lambda < \mu$, $H_\lambda \leq H_\mu \leq n^{\mu-\lambda}(H_\lambda + 1) - 1$.

Since $H_2 = C^2$, the following corollary holds:

Corollary 3.1: (i) For $\lambda > 2$, $n^{2-\lambda}(H_\lambda+1) - 1 \leq C^2 \leq H_\lambda$;

(ii) For $1 < \lambda < 2$, $H_\lambda \leq C^2 \leq n^{2-\lambda}(H_\lambda + 1) - 1$.

Theorems 3.1 and 3.2 are immediate consequences of Lemmas 2.1 and 2.2. In particular, the proof of Theorem 3.1 utilizes the fact that

$$\sum_{i=1}^{n} (x_i - \bar{x})^2 = n^{-1} \sum_{i<j=1}^{n} (x_i - x_j)^2 .$$

The case of equality in these theorems may be obtained from the remarks following the respective lemmas. As for Theil's measure, one has the following result.

<u>Theorem 3.3</u>: $T \leq e^{-1}(H_2 + 1) = e^{-1}(C^2 + 1)$.

<u>Proof</u>: Note that $\max\limits_{y>0} y^{-1} \log y = e^{-1}$ and hence $y \log y \leq e^{-1} y^2$ for $y \geq 0$. Therefore, defining $y_i = x_i/\bar{x}$ $(1 \leq i \leq n)$,

$$T/(H_2 + 1) = (\sum_{i=1}^{n} y_i \log y_i)/(\sum_{i=1}^{n} y_i^2) \leq e^{-1},$$

completing the proof.

<u>Remark</u>: In Theorem 3.3, equality holds iff every non-zero x_i equals $\bar{x}e$ which is, however, impossible since e is irrational. However, it may be checked that $T/(H_2 + 1)$ can be made arbitrarily close to e^{-1} for sufficiently large n provided $x_1, \ldots, x_n$ are suitably chosen.

SECTION 5D : REFERENCES

[1] Bhandari, S.K. and Mukerjee, R. (1986a) : Some relations among inequality measures. Sankhya, Ser. B, Vol.48, Pt. 2, 258 - 261.

[2] Bhandari, S.K. (1986b) : Characterisation of the Parent distribution by inequality measures on its truncations. Sankhya, Ser. B, Vol. 48, Pt. 3, 297 - 300.

[3] Bhattacharya, N. (1963) : A property of the pareto distribution. Sankhya, Ser. B, Vol. 25, 195 - 196.

[4] Marshall, A.W. and Olkin, I. (1979) : Inequalities : Theory of Majorization and its applications. Academic Press, New York.

[5] Nygard, F. and Sandstrom, A. (1981) : Measuring income inequality. Almqvist and Wiksell International, Stockholm.

[6] Ord, J.K., Patil, G.P. and Taillie, C. (1983) : Truncated distributions and measures of income inequality. Sankhya, Ser. B, Vol. 45, 413 - 430.

[7] Piesch, W. (1975) : Statistische Konzentrationsmasse. J.C.B. Mohr (Paul Sieback), Tübinsen.

[8] Rao, C.R. (1973) : Linear Statistical Inference and its Applications, 2nd ed. John Wiley and sons, Inc., New York.

[9] Sen, A. (1973) : On income inequality. Oxford University Press, London.