BOEING SCIENTIFIC RESEARCH LABORATORIES DOCUMENT D1-82-0490

BOEING
SCIENTIFIC RESEARCH LABORATORIES

Monotonicity of Ratios of Means and Other Applications of Majorization

Albert W. Marshall
Ingram Olkin
Frank Proschan

MATHEMATICS RESEARCH DECEMBER 1965

D1-82-0490

MONOTONICITY OF RATIOS OF MEANS AND OTHER APPLICATIONS OF MAJORIZATION

by

Albert W. Marshall

Ingram Olkin
Stanford University

and

Frank Proschan

Mathematical Note No. 436

Mathematics Research Laboratory

BOEING SCIENTIFIC RESEARCH LABORATORIES

December 1965

The research of Professor Ingram Olkin was partially supported by National Science Foundation Grant GP-3837. This report is also being issued by the Department of Statistics, Stanford University, as Technical Report No. 20, December 1, 1965.

ABSTRACT

It is well known that if $a_1,\ldots,a_n$ are positive numbers, then the mean $(\sum_1^n a_i^r/n)^{1/r}$ is increasing in r. In this paper we obtain conditions for monotonicity of the ratio $(\Sigma a_i^r/\Sigma b_i^r)^{1/r}$ of means.

Monotonicity of a ratio can be viewed as a form of total positivity. The theory of total positivity is exploited to obtain more general results.

The proof of monotonicity is based on a theorem giving sufficient conditions for majorization. Several other applications of the majorization theorem are given. One application concerns a stochastic comparison between a function of order statistics from a distribution with increasing failure rate average and the same function of the order statistics from the exponential distribution. Another application is to a comparison between the condition number of a positive definite matrix and the condition number of a polynomial in the matrix.

BLANK PAGE

1. Introduction.

It is well known that if $a_1, \ldots, a_n$ are positive numbers, then the mean $\left(\sum_1^n a_i^r/n\right)^{1/r}$ is increasing in r. In this paper, we first obtain conditions for monotonicity of the ratio $\left(\sum_1^n a_i^r/\sum_1^n b_i^r\right)^{1/r}$ of means (and for its continuous version). The conditions for monotonicity permit $b_1 = \ldots = b_n = 1$, so that $\Sigma b_i^r = n$, and thus the classical result becomes a special case.

Since monotonicity of a ratio can be viewed as a form of total positivity, the theory of total positivity is exploited (Section 3) to obtain more general versions of the result above.

The proof of monotonicity is based on a theorem which gives sufficient conditions for majorization. Because majorization leads to many different forms of inequalities (see Hardy, Littlewood, and Pólya (1952), and Beckenbach and Bellman (1961)), it is not surprising that a diversity of applications of the majorization result are obtained.

One application concerns a stochastic comparison between a function of order statistics from a distribution with increasing failure rate average and the same function of the order statistics from the exponential distribution. The use of these stochastic comparisons in testing certain statistical hypotheses is pointed out (Section 5).

Marshall and Olkin (1965) compare the condition number of a matrix A with the condition number of its symmetrized version AA^*. The majorization result is applied (Section 6) to obtain a comparison between the condition number of a positive definite matrix and the condition number of a polynomial in the matrix.

In the final application, several inequalities concerning absolute deviations are obtained (Section 7) which generalize known results.

Remark. To avoid awkward notation we often omit subscripts in sums or in vectors, e.g., $\Sigma x \log x \equiv \sum_1^n x_i \log x_i$, and

$$\left(\frac{a^r}{\Sigma a^r}\right) \equiv \left(\frac{a_1^r}{\sum_1^n a_i^r}, \ldots, \frac{a_n^r}{\sum_1^n a_i^r}\right).$$

2. Majorization.

To determine conditions on $(a_1,\ldots,a_n)$ and $(b_1,\ldots,b_n)$ such that $g(r) \equiv (\Sigma a^r/\Sigma b^r)^{1/r}$ is nondecreasing in r, one might set $d \log g(r)/dr \geq 0$, or equivalently,

$$\sum\left(\frac{a^r}{\Sigma a^r}\right) \log \left(\frac{a^r}{\Sigma a^r}\right) \geq \sum\left(\frac{b^r}{\Sigma b^r}\right) \log \left(\frac{b^r}{\Sigma b^r}\right). \tag{2.1}$$

Inequalities of this type can be obtained under appropriate conditions using majorization.

Definition 2.1. If $a_1 \geq \cdots \geq a_n$, $b_1 \geq \cdots \geq b_n$, $\sum_1^k a_j \geq \sum_1^k b_j$ for $k = 1, 2, \ldots, n-1$, and $\sum_1^n a_j = \sum_1^n b_j$, then $a = (a_1,\ldots,a_n)$ is said to majorize $b = (b_1,\ldots,b_n)$, written $a \succ b$.

Remark. An alternate definition of majorization is sometimes used:

A set $\{a\}$ is said to majorize a set $\{b\}$ if, possibly after ordering of the elements of each set, the conditions of Definition 2.1 are satisfied (Hardy, Littlewood, and Pólya (1952), page 45).

Definition 2.2. A real function φ of n real variables is said to be a Schur function if for every pair $i \neq j$, $(x_i - x_j)\left(\frac{\partial\varphi}{\partial x_i} - \frac{\partial\varphi}{\partial x_j}\right) \geq 0$.

These concepts are linked in the basic theorem:

Theorem 2.3. (Schur (1923), Ostrowski (1952)). Let $\varphi(x)$ be defined for $x_1 \geq \cdots \geq x_n$. $\varphi(a) \geq \varphi(b)$ for all $a \succ b$ if and only if φ is a Schur function.

Remark. If the alternate definition of majorization is used (see the remark after Definition 2.1), then the requirement that φ be symmetric must be imposed.

Now $\Sigma x \log x$ is a Schur function since all functions of the form $\sum_1^n g(x_i)$, g convex, are Schur functions. Thus to prove (2.1) it is sufficient that $(\alpha/\Sigma\alpha) \succ (\beta/\Sigma\beta)$, where $\alpha_i = a_i^r$ and $\beta_i = b_i^r$, $i = 1,\ldots,n$.

Theorem 2.4. If $\alpha_1 > 0, \ldots, \alpha_n > 0$, $\beta_1 \geq \cdots \geq \beta_n > 0$, and $\frac{\beta_1}{\alpha_1} \leq \cdots \leq \frac{\beta_n}{\alpha_n}$, then $(\frac{\alpha}{\Sigma\alpha}) \succ (\frac{\beta}{\Sigma\beta})$.

Proof. Note that the hypotheses imply $\alpha_1 \geq \cdots \geq \alpha_n > 0$. Thus, we must prove that for $k = 1, \ldots, n-1$,

$$\sum_1^k \alpha_j / \sum_1^n \alpha_\ell \geq \sum_1^k \beta_j / \sum_1^n \beta_\ell .$$

This follows from

$$\sum_1^k \alpha_j \sum_1^n \beta_\ell - \sum_1^k \beta_j \sum_1^n \alpha_\ell = \sum_1^k \alpha_j \sum_{k+1}^n \beta_\ell - \sum_1^k \beta_j \sum_{k+1}^n \alpha_\ell$$

$$= \sum_{j=1}^k \sum_{\ell=k+1}^n \alpha_j \alpha_\ell \left(\frac{\beta_\ell}{\alpha_\ell} - \frac{\beta_j}{\alpha_j}\right) \geq 0. \quad \|$$

Using Theorems 2.3 and 2.4 with (2.1), we have thus derived (for $r > 0$):

Theorem 2.5. If $a_1 > 0, \ldots, a_n > 0$, $b_1 \geq \cdots \geq b_n > 0$, $\frac{b_1}{a_1} \leq \cdots \leq \frac{b_n}{a_n}$, then $(\Sigma a^r/\Sigma b^r)^{1/r}$ is increasing in r.

A more direct, though possibly less clearly motivated, proof of Theorem 2.5 is based on an application of Theorem 2.3 to the Schur function $\varphi(x_1,\ldots,x_n) = \Sigma x^t$, $t \geq 1$. Again for $r > 0$, let $\alpha_i = a_i^r$ and $\beta_i = b_i^r$; for $r < 0$, let $\alpha_i = a_{n-i+1}^r$, $\beta_i = b_{n-i+1}^r$. Then by Theorems 2.3 and 2.4, $\Sigma(a^r/\Sigma a^r)^t \geq \Sigma(b^r/\Sigma b^r)^t$. The conclusion for $r > 0$ ($r < 0$) is a consequence of choosing $t = s/r$, $s \geq r$ ($s \leq r$). Finally, monotonicity in r follows from the continuity of $(\Sigma a^r/\Sigma b^r)^{1/r}$ at $r = 0$.

An important example in which the conditions of Theorem 2.4 are satisfied is obtained by choosing $\alpha_i = \psi(\beta_i)$, where ψ is a non-negative starshaped function. (A real function ψ defined on $[0,\infty)$ is said to be starshaped if $\psi(x)/x$ is increasing in x. An interesting example of starshapedness is ψ convex on $[0,\infty)$, $\psi(0) \leq 0$. Note that a non-negative starshaped function ψ must be increasing and must satisfy $\psi(0) = 0$. Such functions are discussed by Bruckner and Ostrow (1962).) With this choice, assuming $\beta_1 \geq \cdots \geq \beta_n > 0$, it follows that $\alpha_1 \geq \cdots \geq \alpha_n \geq 0$, so that by Theorem 2.4,

$$(2.) \qquad \left(\frac{\beta}{\Sigma\beta}\right) \prec \frac{\psi(\beta)}{\Sigma\psi(\beta)} = \left(\frac{\alpha}{\Sigma\alpha}\right) .$$

3. An Extension Using Total Positivity.

Theorems 2.4 and 2.5 can be viewed as theorems on total positivity and can be proved by the methods of total positivity. A matrix (t_{ij}) of non-negative numbers is said to be totally positive of order 2 (TP_2) if all its 2×2 minors are non-negative. Similarly, a non-negative function $t(x,y)$ of real variables x belonging to $\mathfrak{X}$ and y belonging to $\mathcal{Y}$ is said to be totally positive of order 2 if, for all $x_1 < x_2$ in $\mathfrak{X}$ and $y_1 < y_2$ in $\mathcal{Y}$, the determinant

$$\det\begin{pmatrix} t(x_1,y_1) & t(x_1,y_2) \\ t(x_2,y_1) & t(x_2,y_2) \end{pmatrix} \geq 0 .$$

Suppose the t_{ij} are non-negative; then (t_{ij}) is clearly TP_2 if and only if $t_{i+1,j}/t_{ij}$ is increasing in j for each i.

Let $t_{1j} = \alpha_j > 0$, $t_{2j} = \beta_j > 0$, $j = 1, 2, \ldots, n$; then β_i/α_i increasing in i is equivalent to (t_{ij}) being TP_2. Since it is easily concluded from Theorem 2.4 that $\sum_1^k \alpha_i / \sum_1^\ell \alpha_j \geq \sum_1^k \beta_i / \sum_1^\ell \beta_j$ for all $k \leq \ell$, Theorem 2.4 essentially states that if (t_{ij}) is TP_2 and $T_{ik} \equiv \sum_{j=1}^k t_{ij}$, then (T_{ij}) is TP_2. Write $T_{ik} = \sum_{j=1}^n t_{ij}H_{jk}$, where H_{jk} is 1 for $j \leq k$ and is 0 for $j > k$. Since (H_{jk}) is TP_2, Theorem 2.4 follows from a standard theorm of total positivity which asserts that

$$\varphi(x,z) \equiv \int \varphi_1(x,y)\, \varphi_2(y,z)\, d\mu(y)$$

is TP_2 provided $\varphi_1(x,y)$ and $\varphi_2(y,z)$ are each TP_2.

Because monotonicity of ratios can be interpreted as total positivity of order 2, Theorem 2.5 also is a statement of total positivity. This suggests the following generalization:

<u>Theorem 3.1</u>. Let $g(x,y)$ be TP_2 and decreasing in y for each x. Then

$$h(x,s) \equiv \left(\int g^{1/s}(x,y)\, d\mu(y)\right)^s$$

is TP_2.

Proof. Define

$$u(x,s) = \log h(x,s) = s \log \int g^{\frac{1}{s}}(x,y)\, d\mu(y) ,$$

then

$$-\, s^2\left[\int g^{\frac{1}{s}}\, d\mu\right]^2 \frac{\partial^2 u(x,s)}{\partial x\, \partial s}$$

$$= \int g^{\frac{1}{s}} d\mu \int g^{\frac{1}{s}-1} g_x \log g\, d\mu \;-\; \int g^{\frac{1}{s}-1} g_x d\mu \int g^{\frac{1}{s}} \log g\, d\mu$$

$$= \det \begin{pmatrix} \int g^{\frac{1}{s}} d\mu & \int g^{\frac{1}{s}-1} g_x d\mu \\ \int g^{\frac{1}{s}} \log g\, d\mu & \int g^{\frac{1}{s}-1} g_x \log g\, d\mu \end{pmatrix}$$

$$= \int_{y_1<y_2} \det \begin{pmatrix} g^{\frac{1}{s}}(x,y_1) & g^{\frac{1}{s}-1}(x,y_1)\, g_x(x,y_1) \\ g^{\frac{1}{s}}(x,y_2) & g^{\frac{1}{s}-1}(x,y_2)\, g_x(x,y_2) \end{pmatrix} \det \begin{pmatrix} 1 & \log g(x,y_1) \\ 1 & \log g(x,y_2) \end{pmatrix} d\mu(y_1)\, d\mu(y_2)$$

from Problem 68, Pólya and Szegö (1925). The first determinant of the integrand is non-negative since $g(x,y)$ is TP_2. The second determinant is non-positive since $g(x,y)$ is decreasing in y for each x. The result follows from the fact that a non-negative function $h(x,s)$ is TP_2 if and only if $\partial^2 \log h(x,s)/\partial x\, \partial s \geq 0$. ||

To obtain Theorem 2.5 from Theorem 3.1, let y range over the values $1, 2, \ldots, n$ and choose $g(1,i) = a_i$, $g(2,i) = b_i$, and $\mu(i) = 1/n$. We can extend Theorem 2.5 slightly be choosing $\mu(i) = p_i > 0$, $\Sigma p_i = 1$.

Under the hypothesis of Theorem 2.5 we conclude that $(\Sigma pa^r/\Sigma pb^r)^{1/r}$ is increasing in r. A further extension is obtained in the next section.

4. Monotonicity of Ratios of Means.

With Theorem 2.5 in mind, consider right continuous distributions G and F of non-negative random variables approximated by step functions with jumps $1/n$ at a_i and b_i respectively, $a_1 \geq \cdots \geq a_n > 0$, $b_1 \geq \cdots \geq b_n > 0$. Then $\bar{G}(a_i) \approx i/n \approx \bar{F}(b_i)$ (where $\bar{F} \equiv 1 - F$, $\bar{G} \equiv 1 - G$), and the condition b_i/a_i increasing in i becomes $\bar{F}^{-1}(i/n)/\bar{G}^{-1}(i/n)$ increasing in i. In the limit, we obtain the condition $\bar{F}^{-1}(p)/\bar{G}^{-1}(p)$ increasing in p.

To avoid the necessity of assuming that the distribution functions are strictly increasing, define

$$\bar{H}^{-1}(p) = \inf\{x \geq 0: \bar{H}(x) \leq p\}, \quad H^{-1}(p) = \inf\{x \geq 0: H(x) \geq p\} ,$$

where H is a distribution function.

<u>Theorem 4.1</u>. If F and G are distribution functions, $F(0) = 0 = G(0)$, and $\bar{F}^{-1}(p)/\bar{G}^{-1}(p)$ is increasing in p, then

$$\left\{\frac{\int x^r dG(x)}{\int x^r dF(x)}\right\}^{1/r}$$

is increasing in r.

<u>Proof</u>. Using the approximations above, a limiting argument may be used to obtain the result from Theorem 2.5. Alternately, it may be obtained as a special case of Theorem 3.1 by choosing $g(1,y) = \bar{G}^{-1}(y)$, $g(2,y) = \bar{F}^{-1}(y)$ and taking μ to be uniform on $[0,1]$. ||

When $F(0\ ) = 0 = G(0\ )$, F and G are continuous and have continuous inverses, the hypothesis of Theorem 4.1 has the following equivalent formulations:

(i) $\bar{F}^{-1}(p)/\bar{G}^{-1}(p) \equiv F^{-1}(1-p)/G^{-1}(1-p)$ is increasing in p, $0 < p < 1$,

(ii) $\bar{G}^{-1}\,\bar{F}(x)/x \equiv G^{-1}F(x)/x$ is increasing in x in the support of F,

(iii) $F(x) = G(\psi(x))$ for some non-negative starshaped function ψ,

(iv) If X is a random variable with distribution F, then $\psi(X)$ has distribution G for some non-negative starshaped ψ.

We say that <u>F is starshaped with respect to G</u> if (i), (iii) or (iv) holds, because (i), (iii) and (iv) are equivalent even without the continuity restrictions on F and G. Since the equivalence of (iii) and (iv) is easily verified, we need show only that (i) and (iii) are equivalent. Let $\mathcal{S}_F$ be the set of all x such that $x = \inf\{z: \bar{H}(z) \leq p\}$ for some $p \in (0,1)$.

<u>Theorem 4.2</u>. If F and G are right continuous distribution functions such that $F(0\ ) = G(0\ ) = 0$, then $F(x) \equiv G(\varphi(x))$ for some strictly increasing function φ such that $\varphi(x)/x$ is increasing for all $x \in \mathcal{S}_F$ if and only if $\bar{F}^{-1}(p)/\bar{G}^{-1}(p)$ is increasing in p, $0 < p < 1$.

<u>Proof</u>. Suppose $F(x) = G(\varphi(x))$ for some non-negative strictly increasing φ such that $\varphi(x)/x$ is increasing in $x \in \mathcal{S}_F$. Then

$$\frac{\bar{F}^{-1}(p)}{\bar{G}^{-1}(p)} = \frac{\inf\{x: \bar{F}(x) \leq p\}}{\inf\{x: \bar{G}(x) \leq p\}} = \frac{\inf\{x: \bar{F}(x) \leq p\}}{\inf\{x: \bar{F}(\varphi^{-1}(x)) \leq p\}} = \frac{\inf\{y: \bar{F}(y) \leq p\}}{\inf\{\varphi(y): \bar{F}(y) \leq p\}}$$

$$= \frac{y_p}{\varphi(y_p)} .$$

Now $p < p^*$ implies $y_p \geq y_{p^*}$. Thus $p < p^*$ implies

$$y_p/\varphi(y_p) \leq y_{p^*}/\varphi(y_{p^*}), \text{ i.e., } \frac{\overline{F}^{-1}(p)}{\overline{G}^{-1}(p)} \leq \frac{\overline{F}^{-1}(p^*)}{\overline{G}^{-1}(p^*)} .$$

Next, suppose $\overline{F}^{-1}(p)/\overline{G}^{-1}(p)$ is increasing in p, let $y \in \mathcal{S}_F$ and define $\varphi(y) = \overline{G}^{-1}\overline{F}(y) = \inf\{z: \overline{G}(z) \leq \overline{F}(y)\}$. Then since

$$\frac{\overline{F}^{-1}(p)}{\overline{G}^{-1}(p)} = \frac{\inf\{z: \overline{F}(z) \leq p\}}{\inf\{z: \overline{G}(z) \leq p\}} = \frac{y_p}{\varphi(y_p)}$$

is increasing in p, it follows that $\varphi(y)/y$ is increasing in $y \in \mathcal{S}_F$. Also $\varphi(y)$ is strictly increasing because $\overline{F}(y)$ is strictly decreasing in $y \in \mathcal{S}_F$. ||

Some choices of F and G in Theorem 4.1 are of special interest.

I. Let F be degenerate at 1. Then $\overline{F}^{-1}(p)/\overline{G}^{-1}(p) = 1/\overline{G}^{-1}(p)$ is increasing for all G. In this case, Theorem 4.1 reduces to the well-known fact that $[\int x^r dG(x)]^{1/r}$ is increasing in r for every G.

II. Let $\overline{G}(x) = e^{-x}$. Then the condition $\overline{F}^{-1}(p)/\overline{G}^{-1}(p)$ increasing in p becomes $-x^{-1}\log \overline{F}(x)$ increasing in x. A distribution F with this property is said to have an <u>increasing failure (hazard) rate average</u> (IFRA), since in case F has a density f, $-\log \overline{F}(x) = \int_0^x q(u)du$, where $q(u) \equiv f(u)/\overline{F}(u)$ is the failure rate of F. Similarly, if $x^{-1}\log \overline{F}(x)$ is decreasing, we say F has a <u>decreasing failure (hazard) rate average</u> (DFRA). The class of IFRA (DFRA) distributions has applications in reliability theory (see Birnbaum, Esary, and Marshall (1965)). From Theorem 4.1 we see that if F is IFRA (DFRA), $[\int x^r dF(x)/\Gamma(r+1)]^{1/r}$ is decreasing (increasing) in $r \geq 0$. This conclusion was obtained for the smaller class of distributions with increasing (decreasing) failure rate by Barlow, Marshall, and Proschan (1963).

Properties of order statistics from IFRA (DFRA) distributions are developed in Section 5.

Choose $\overline{F}(x) = e^{-x}$ and G a DFRA distribution for which $\int x^r dG(x) = r^r, r>0$. (A proof that such a distribution exists has been obtained by Herman Rubin and is presented in the appendix.) From Theorem 4.1 it follows that $[r^r/\Gamma(r+1)]^{1/r}$ is an increasing function of $r>0$, a result obtained by Minc and Sathre (1964).

III. Let $F(x) = G(x^t)$, $t \geq 1$. Then $\overline{G}^{-1}(p) = [\overline{F}^{-1}(p)]^t$ so that $\overline{F}^{-1}(p)/\overline{G}^{-1}(p) = [\overline{F}^{-1}(p)]^{1-t}$ is increasing in p. Thus, the conditions of Theorem 4.1 are satisfied, and so

$$\left[\frac{\int x^r dG(x)}{\int x^r dF(x)}\right]^{1/r} = \left[\frac{\int x^{rt} dF(x)}{\int x^r \, dF(x)}\right]^{1/r}$$

is increasing in r. More generally, if X has distribution F and $Y = \psi(X)$ has distribution G where $\psi \geq 0$ is starshaped, then

$$\{\int [\psi(x)]^r \, dF(x)/\int x^r dF(x)\}^{1/r}$$

is increasing in r.

5. Statistical Applications.

In the present section we make some comparisons involving the order statistics from distributions F and G, where F is starshaped with respect to G.

Let $X_1 \geq \cdots \geq X_n$ be order statistics from F. Then from (2.2) we have for any starshaped $\psi \geq 0$ that

$$\left(\frac{X_1}{\Sigma X}, \ldots, \frac{X_n}{\Sigma X}\right) \prec \left(\frac{\psi(X_1)}{\Sigma\psi(X)}, \ldots, \frac{\psi(X_n)}{\Sigma\psi(X)}\right) .$$

Consequently, if F is starshaped with respect to the distribution G, and $X_1 \geq \cdots \geq X_n$ are order statistics from F, $Y_1 \geq \cdots \geq Y_n$ are order statistics from G, then

$$(5.1)\qquad \left(\frac{X_1}{\Sigma X}, \ldots, \frac{X_n}{\Sigma X}\right) \overset{st}{\prec} \left(\frac{Y_1}{\Sigma Y}, \ldots, \frac{Y_n}{\Sigma Y}\right),$$

i.e., $\sum_1^k X_i / \sum_1^n X_i$ is stochastically less than or equal to $\sum_1^k Y_i / \sum_1^n Y_i$ for $k = 1, \ldots, n$. (A random variable U is stochastically greater than or equal to a random variable V, denoted by $U \overset{st}{\geq} V$, if $P[U \geq a] \geq P[V \geq a]$ for all a.)

With (5.1), Theorem 2.3 yields some interesting applications. For example, choosing the Schur function $\varphi(z_1,\ldots,z_n) = n^{-1}\Sigma z^2 - 1$, we obtain

$$(5.2)\qquad \Sigma(X_i-\overline{X})^2/\,\overline{X}^2 \overset{st}{\leq} \Sigma(Y_i-\overline{Y})^2/\,\overline{Y}^2 \ .$$

Choosing the Schur function $\varphi(z_1,\ldots,z_n) = n\Sigma az/\Sigma z$, $a_1 \geq \cdots \geq a_n$, $z_1 \geq \cdots \geq z_n$, we obtain

$$(5.3)\qquad \Sigma a_i X_i/\overline{X} \overset{st}{\leq} \Sigma a_i Y_i/\overline{Y} \ ,$$

where $\overline{X} = \Sigma X/n$, $\overline{Y} = \Sigma Y/n$.

An important special case where F is starshaped with respect to G (introduced in Section 4) is obtained by choosing $\overline{G}(x) = e^{-x}$ and F IFRA. We discuss the statistical applications of (5.2) and (5.3) in this case, although the results hold more generally.

I. Consider first the problem of testing the hypothesis that F is exponential versus the alternative that F is IFRA. Because of the stochastic ordering given by (5.2), a test based on the statistic $R(X) = \Sigma(X_i-\overline{X})^2/\overline{X}^2$ is unbiased. (By using different Schur functions, alternative statistics could be used, yielding unbiased tests.) To carry out the test at the level of significance α, determine k_α such that $P(R(X) \leq k_\alpha) = \alpha$ when F is exponential, and reject the hypothesis of exponentiality when $R(X) \leq k_\alpha$.

For a test of the hypothesis that F is exponential versus the alternative that F has increasing failure rate, i.e., $\log \overline{F}(x)$ is concave where finite, see Proschan and Pyke (1965).

II. Next, we discuss testing for outliers when the distribution F is known to be IFRA. Suppose that $X_1 \geq \cdots \geq X_n$ are order statistics from F unless possibly that X_1 is an "outlier," i.e., X_1 does not arise as an observation from F, but from $F_1 \leq F$, $F_1 \neq F$.

A natural test of the hypothesis that $X_1 \geq \cdots \geq X_n$ all come from F is to reject the hypothesis if $X_1/\overline{X}$ is too large. If F is unknown, the distribution of this statistic is unavailable, but since F is IFRA, it follows from (5.3) with $a_1 = 1$, $a_2 = \cdots = a_n = 0$ that $X_1/\overline{X} \overset{st}{\leq} Y_1/\overline{Y}$, where $Y_1 \geq \cdots \geq Y_n$ are the order statistics from an exponential distribution. To control the type I error, note that $P(X_1/\overline{X} > k_\alpha) \leq P(Y_1/\overline{Y} > k_\alpha) = \alpha$.

Although the type II error cannot be determined without knowing F_1, we can assert that the type II error is smaller under F_2 than under F_1 whenever F_2 is starshaped with respect to F_1.

Similarly, a test can be obtained when X_n is suspected of being an outlier; here the test statistic is $X_n/\overline{X}$, and we use (5.3) with $a_1 = \cdots = a_{n-1} = 0$, $a_n = -1$ to control the type I error as above. More generally, a test against the possibility that $X_1, \ldots, X_r$ and $X_s, \ldots, X_n$ are all outliers makes use of (5.3) with $a_1 = \cdots = a_r = 1$, $a_{r+1} = \cdots = a_{s-1} = 0$, $a_s = \cdots = a_n = -1$. The test statistic is

$(X_s + \dots + X_n - X_1 - \dots - X_r)/\bar{X}$.

6. Condition Numbers.

A commonly used measure of the difficulty of numerically inverting a non-singular matrix A is its condition number $c_\varphi(A) = \varphi(A)\,\varphi(A^{-1})$, where ordinarily φ is a norm (i.e., $\varphi(A) > 0$ when $A \neq 0$, $\varphi(\gamma A) = |\gamma|\,\varphi(A)$ for complex γ, $\varphi(A+B) \leq \varphi(A) + \varphi(B)$). Marshall and Olkin (1965) show that

$$(6.1) \qquad c_\varphi(A) \leq c_\varphi(AA^*)$$

when φ is a unitarily invariant norm (i.e., $\varphi(A) = \varphi(UA) = \varphi(AU)$ for all unitary matrices U). The proof is based in part on a result of von Neumann (1937) that φ is a unitarily invariant norm if and only if there exists a symmetric gauge function Φ such that $\varphi(A) = \Phi(\alpha)$ for all A, where $\alpha_1^2, \dots, \alpha_n^2$ are the characteristic roots of AA^*. (A function Φ on a complex vector space is called a symmetric gauge function (SGF) if $\Phi(u) > 0$ when $u \neq 0$, $\Phi(\gamma u) = |\gamma|\,\varphi(u)$ for complex γ, $\Phi(u+v) \leq \Phi(u) + \Phi(v)$, and $\Phi(u_1,\dots,u_n) = \Phi(\epsilon_1 u_{i_1},\dots,\epsilon_n u_{i_n})$ whenever $\epsilon_j = \pm 1$ and $(i_1, \dots, i_n)$ is a permutation of $(1,\dots,n)$.) Using this, a more general result than (6.1) is also obtained by Marshall and Olkin (1965), namely, that if Φ is a SGF and $\alpha_i > 0$, then

$$(6.2) \qquad \Phi(\alpha_1^r,\dots,\alpha_n^r)\,\Phi(\alpha_1^{-r},\dots,\alpha_n^{-r}) \text{ is increasing in } r > 0 .$$

The proof of (6.2) rests on the following

<u>Lemma 6.1</u>. If $u \prec v$, Φ is a SGF, and g is a non-negative convex function, then

$$(6.3) \qquad \Phi(g(u)) \leq \Phi(g(v)) .$$

The proof of (6.3) was given by Fan (1951) in case (i) $g(x) = x$, and by Marshall and Olkin (1965) in case (ii) $g(x) = x^{-1}$. The proof of Lemma 6.1 parallels that of (ii) where the only properties of x^{-1} used are its non-negativity and convexity.

We note that every SGF is a Schur function. The result (6.3) does not hold for Schur functions Φ without the additional condition that g is increasing.

In Section 2 we have shown that if $\psi \geq 0$ is starshaped then $(\psi(\beta)/\Sigma\psi(\beta)) \succ (\beta/\Sigma\beta)$. In the following, we consider non-negative starshaped functions of the form $\psi(\beta) = \Sigma c_j \beta^j$ where $c_j \geq 0$. From (6.3) with $g(x) = x$ we obtain that $\Phi(\beta_1/\Sigma\beta, \ldots, \beta_n/\Sigma\beta) \leq \Phi(\psi(\beta_1)/\Sigma\psi(\beta), \ldots, \psi(\beta_n)/\Sigma\psi(\beta))$. From (6.3) with $g(x) = x^{-1}$ we obtain that $\Phi(\Sigma\beta/\beta_1, \ldots, \Sigma\beta/\beta_n) \leq \Phi(\Sigma\psi(\beta)/\psi(\beta_1), \ldots, \Sigma\psi(\beta)/\psi(\beta_n))$. By multiplying these inequalities and using $\Phi(\gamma x) = |\gamma| \, \Phi(x)$, we obtain

$$(6.4) \quad \Phi(\beta_1, \ldots, \beta_n) \, \Phi\left(\frac{1}{\beta_1}, \ldots, \frac{1}{\beta_n}\right) \leq \Phi(\psi(\beta_1), \ldots, \psi(\beta_n)) \, \Phi\left(\frac{1}{\psi(\beta_1)}, \ldots, \frac{1}{\psi(\beta_n)}\right) .$$

The special case $\psi(\beta) = \beta^{s/r}, 0<r<s$, and $\beta_i = \alpha_i^r$, $i = 1, 2, \ldots, n$, yields (6.2).

If B is a positive definite matrix with characteristic roots $\beta_1, \ldots, \beta_n$, then the characteristic roots of $\psi(B) = \Sigma c_j B^j$ are $\psi(\beta_i) = \Sigma c_j \beta_i^j$. (For any positive definite matrix C, the square root of the characteristic roots of CC^* are the characteristic roots of C.) Hence, in terms of condition numbers, (6.4) becomes

$$(6.5) \qquad c_\varphi(B) \leq c_\varphi(\psi(B))$$

for any positive definite matrix B.

For any matrix A and any unitarily invariant norm, $c_\varphi(A) = c_\varphi((AA^*)^{\frac{1}{2}})$. With $B = (AA^*)^{\frac{1}{2}}$ and $\psi(B) = \Sigma c_j B^j$, it follows from (6.5) that

$$c_\varphi(A) \le c_\varphi(\psi(AA^*)^{\frac{1}{2}})) \quad (6.6)$$

This reduces to (6.1) when $\psi(B) = B^2$.

7. Inequalities for Absolute Deviations.

If $c_1, \ldots, c_n$ are real numbers satisfying $\sum_1^n c_i = 0$, then

$$\tfrac{1}{2}\Sigma|c|^s \le (\tfrac{1}{2}\Sigma|c|)^s, \qquad s \ge 1 . \quad (7.1)$$

This inequality was proved by Gatti (1956) for $s = 2$, and in its present form by Birnbaum (1958). Generalizations of (7.1) can be obtained using Theorems 2.3 and 2.4.

Lemma 7.1. If $c_1, \ldots, c_n$ satisfy $\sum_1^n c_i = 0$, $u_i \equiv |c_i|$, and $u_1 \ge \cdots \ge u_n$, then

$$\left(\frac{u_1}{\Sigma u}, \ldots, \frac{u_n}{\Sigma u}\right) \prec (\tfrac{1}{2}, \tfrac{1}{2}, 0, \ldots, 0) . \quad (7.2)$$

Proof. From $c_1 = -\sum_2^n c_i$ we have $u_1 = |\sum_2^n c_i| \le \sum_2^n u_i$, so that $2u_1 \le \sum_1^n u_i$, and consequently $u_1/\sum_1^n u_i \le \frac{1}{2}$. Of course, $(u_1+u_2)/\sum_1^n u_i \le 1$.

From (2.2), it follows that for $0 \le r \le 1$,

$$\left(\frac{u_1^r}{\Sigma u^r}, \ldots, \frac{u_n^r}{\Sigma u^r}\right) \prec \left(\frac{u_1}{\Sigma u}, \ldots, \frac{u_n}{\Sigma u}\right) . \quad (7.3)$$

Since $\varphi(x_1, \ldots, x_n) = \Sigma x^t$, $t \ge 1$, is a Schur function, we have from Theorem 2.3 and (7.2),

$$\left(\frac{u^r}{\Sigma u^r}\right)^t \le 2\,\frac{1}{2^t}, \qquad t \ge 1 .$$

The choice $t = s/r$, $s \geq r$, yields

$$(7.4) \qquad (\tfrac{1}{2}\Sigma u^r)^{1/r} \geq (\tfrac{1}{2}\Sigma u^s)^{1/s}, \qquad 0 \leq r \leq 1,\ r \leq s \ .$$

The result (7.1) corresponds to the case $r = 1$. The restriction $0 \leq r \leq 1$ cannot be relaxed to $0 \leq r$, as may be seen from the choice $c_1 = 1$, $c_2 = \cdots = c_n = -\frac{1}{n-1}$ and $r > 1$.

Remark. If in Lemma 7.1 we define $u_i = h(|c_i|)$ where h is non-negative and subadditive, i.e., $h(x+y) \leq h(x) + h(y)$, then (7.2) remains valid. The function $h(x) = x^r$, $0 \leq r \leq 1$, leads to (7.4). If $h(x)$ is non-negative and concave for $x \geq 0$, with $h(0) = 0$, then $h(x)$ is subadditive in $x \geq 0$.

As a consequence of (7.1), De Novellis (1958) proves that

$$(7.5) \qquad \Sigma u^{s+1} \leq \tfrac{1}{2}\, \Sigma u\, \Sigma u^s \qquad \text{for} \quad s \geq 0 \ .$$

Using the generalized version (7.4) of (7.1), a generalized version of (7.5) can be obtained, namely,

$$(7.6) \qquad \Sigma u^{r+s} \leq \tfrac{1}{2}\, \Sigma u^r\, \Sigma u^s \ , \qquad 0 \leq r \leq 1 \ .$$

From (7.2) and (7.3), we see that $\frac{1}{2} \geq u_1^r/\Sigma u^r$, so that

$$\Sigma u^{r+s}/\Sigma u^s = \Sigma u^r\left(\frac{u^s}{\Sigma u^s}\right) \leq u_1^r \leq \tfrac{1}{2}\Sigma u^r \ ,$$

which yields (7.6).

The role of $\frac{1}{2}$ in (7.2) stems from the fact that $c_1, \ldots, c_n$ can be divided into two sets with equal sums (except for sign). Suppose we have three sets of positive numbers $\{x\}$, $\{y\}$, $\{z\}$ such that $\sum_1^{n_1} x_i = \sum_1^{n_2} y_j = \sum_1^{n_3} z_k$. Let $u_1 \geq \cdots \geq u_n \, (n = n_1 + n_2 + n_3)$ be the ordered x's, y's, and z's. Then

$$\left(\frac{u_1}{\Sigma u}, \ldots, \frac{u_n}{\Sigma u}\right) \prec \left(\frac{1}{3}, \frac{1}{3}, \frac{1}{3}, 0, \ldots, 0\right) .$$

By using arguments analogous to those for the case of two sets, we obtain

$$\left(\frac{1}{3}\, \Sigma u^r\right)^{1/r} \geq \left(\frac{1}{3}\, \Sigma u^s\right)^{1/s}, \qquad 0 \leq r \leq 1,\ r \leq s ,$$

which parallels (7.4). The extension to the case of an arbitrary number of sets is immediate.

APPENDIX

Theorem. There exists a decreasing failure rate (DFR) distribution with moments $\mu_r = r^r$, $r > 0$.

Proof. (H. Rubin). Consider

$$r^r = \int_0^\infty x^r f(x)dx = \int_{-\infty}^\infty e^{-yr} f(e^{-y})\, e^{-y}dy$$

$$\equiv \int_{-\infty}^\infty e^{-yr} g(y)dy .$$

From the inversion formula, Widder (1941) p. 241,

$$g(y) = \frac{1}{2\pi i} \int_{-\infty}^\infty r^r e^{yr}dr = \frac{1}{2\pi i} \int_{-\infty}^\infty \exp[r \log r + y^r]dr .$$

Transform from r to θ by $r = \rho(y,\theta)\, e^{i\theta}$ choosing ρ in such a manner that $r \log r + y^r$ is real. Since

$$r \log r + yr = \rho[\cos \theta(\log \rho + y) - \theta \sin \theta] + i\rho[\sin \theta(\log \rho + y) + \theta \cos \theta]$$

$$\equiv a + bi ,$$

we choose $b = 0$, i.e.,

$$\text{(A.1)} \qquad \log \rho = - \theta \cot \theta - y \; .$$

Then

$$\begin{aligned} r \log r + yr &= \rho[\cos \theta(-\theta \cot \theta) - \theta \cos \theta) = - \rho\theta \csc \theta \\ &= - e^{-y}(\theta \csc \theta \, e^{-\theta \cot \theta}) \\ &\equiv - e^{-y} A(\theta) \; . \end{aligned}$$

To determine $dr/d\theta$, note that from (A.1) $\frac{1}{\rho}\frac{d\rho}{d\theta} = \theta \csc^2\theta - \cot \theta$, so that

$$\begin{aligned} \frac{dr}{d\theta} &= \rho \frac{\theta}{d\theta} (\cos \theta + i \sin \theta) + e^{i\theta} \frac{d\rho}{d\theta} \\ &= \rho \csc \theta[\theta \cot \theta - 1] + i\rho\theta \csc \theta \equiv c + di \; . \end{aligned}$$

Hence

$$g(y) = \frac{1}{2\pi i} \int \exp[- e^{-y}A(\theta)] \; (c+di) \; d\theta \equiv L + Mi \; .$$

But $g(y)$ is real, so that the right-hand side must be real. This implies that the complex term vanishes, and that

$$\begin{aligned} g(y) &= \frac{1}{2\pi} \int_{-\pi}^{\pi} \exp[- e^{-y}A(\theta)] \; e^{-y}A(\theta)d\theta \\ &= \frac{1}{\pi} \int_0^{\pi} \exp[- e^{-y}A(\theta)] \; e^{-y}A(\theta)d\theta \; . \end{aligned}$$

Let $x = e^{-y}$, then

$$f(x) = \frac{g(-\log x)}{x} = \frac{1}{\pi} \int_0^{\pi} e^{-xA(\theta)}A(\theta)d\theta \; ,$$

which is a mixture of exponential distributions, and is automatically DFR.

REFERENCES

[1] Barlow, R. E., Marshall, A. W., and Proschan, F. (1963). Properties of probability distributions with monotone hazard rate. Ann. Math. Statist. 34, 374-389.

[2] Beckenbach, E. F., and Bellman, R. (1961). Inequalities, Springer-Verlag, Berlin.

[3] Birnbaum, Z. W. (1958). On an inequality due to S. Gatti. Metron 19, 3-4.

[4] Birnbaum, Z. W., Esary, J. D., and Marshall, A. W. (1965). Stochastic characterization of wearout for components and systems. Boeing Scientific Research Laboratories Document D1-82-0460.

[5] Bruckner, A. M. and Ostrow, E. (1962). Some function classes related to the class of convex functions. Pacific J. Math. 12, 1203-1215.

[6] De Novellis, M. (1958). Some applications and developments of Gatti-Birnbaum inequality. Metron, 245-247.

[7] Fan, K. (1951). Maximum properties and inequalities for the eigenvalues of completely continuous operators. Proc. Nat. Acad. Sci. U.S.A. 37, 760-766.

[8] Gatti, S. (1956). Sul massimo di un indice di anormalità, Metron 18, 181-188.

[9] Hardy, G. H., Littlewood, J. E., and Pólya, G. (1952). Inequalities, 2nd ed., Cambridge University Press, Cambridge.

[10] Marshall, A. W. and Olkin, I. (1965). Norms and inequalities for condition numbers. Pacific J. Math. 15, 241-247.

[11] Minc, H. and Sathre, L. (1964). Some inequalities involving $(r!)^{1/r}$. Proc. Edinburgh Math. Soc. 14 (Series II), 41-46.

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

[13] Pólya, G., and Szegö, G. (1925). Aufgaben and Lehrsatze aus der Analysis, Vol. I, Julius Springer, Berlin.

[14] Proschan, F. and Pyke, R. (1965). Tests for monotone failure rate, Boeing Scientific Research Laboratories, Math. Note No. 400, to appear in the Proceedings of the Fifth Berkeley Symposium on Mathematical Statistics and Probability.

[15] Schur, I. (1923). Über eine Klasse von Mittelbildungen mit Anwendungen auf die Determinantentheorie. Sitzber. Berl. Math. Ges. 22, 9-20.

[16] von Neumann, J. (1937). Some matrix-inequalities and metrization of matric-space. Tomsk. Univ. Rev. 1, 286-300 (in Collected Works, Vol. IV, Pergamon Press, 1962).

[17] Widder, David Vernon (1941). The Laplace transform. Princeton University Press.