JOURNAL OF MATHEMATICAL ANALYSIS AND APPLICATIONS 8, 503-514 (1964)

Reversal of the Lyapunov, Hölder, and Minkowski Inequalities and other Extensions of the Kantorovich Inequality*

ALBERT W. MARSHALL

Boeing Scientific Research Laboratories

AND

INGRAM OLKIN

Stanford University, Stanford, California

Submitted by Richard Bellman

I. INTRODUCTION

If $x_1, \cdots, x_n$ and $p_1, \cdots, p_n$ are nonnegative real numbers with $\Sigma p_i = 1$, and we define $\mu_r = \Sigma p_i x_i^r$, then according to Lyapunov's inequality [1, p. 27],

$$\mu_v^{w-u} \leqslant \mu_u^{w-v} \mu_w^{v-u}, \qquad 0 \leqslant u \leqslant v \leqslant w. \tag{1.1}$$

Since the p_i are probabilities, we may consider a random variable X with the distribution $P\{X = x_i\} = p_i$, $i = 1, \cdots, n$, in which case $\mu_r = EX^r$. But (1.1) holds for arbitrary nonnegative random variables, so that in the following we need only assume that $P\{X \geqslant 0\} = 1$ and $\mu_r = EX^r$.

Generally speaking, there is no positive constant γ for which

$$\mu_v^{w-u} \geqslant \gamma \mu_u^{w-v} \mu_w^{v-u}, \qquad u \leqslant v \leqslant w. \tag{1.2}$$

(A related result is given by Karlin, Proschan, and Barlow [3].) However, such a constant may exist if further restrictions are placed on the distribution F of X. For example, if $\log [1 - F(x)]$ is concave and u is a positive integer, then Barlow, Marshall, and Proschan [2] obtain

$$\gamma = [\Gamma(v + 1)]^{w-u} [\Gamma(u + 1)]^{-(w-v)} [\Gamma(w + 1)]^{-(v-u)}.$$

In this paper, we consider the restriction that X is positive and bounded, i.e., $P\{m \leqslant X \leqslant M\} = 1$, with $m > 0$, and do not require that u be nonnegative. Under this condition, the three special cases of (1.2) obtained by taking $u = 0$, $v = 0$, or $w = 0$ have each been recently obtained for discrete

* This work was supported in part by the National Science Foundation at Stanford University.

random variables by Cargo and Shisha [4]. The special case $u = v - r$, $w = v + r$, $r > 0$ yields

$$\mu_{v-r}\mu_{v+r} \leqslant \gamma\mu_v^2,$$

and follows from results of Greub and Rheinboldt [5]. When $v = 0$, $r = 1$, this reduces to the well-known inequality of L. V. Kantorovich (for a list of references and general discussion see [6]).

As is well known, Lyapunov's inequality can be obtained directly from Hölder's inequality. It is not surprising then, that in the course of deriving (1.2), we obtain some general results which also yield reversals of Hölder's and Minkowski's inequalities.

II. A Fundamental Inequality

By determining conditions when a linear combination of x^r and x^s is non-negative in the interval $[m, M]$, we obtain a fundamental inequality from which we are able to derive all of the succeeding results of this paper.

For $0 < m < M$ and $r < s$, $rs \neq 0$, we introduce the notation

$$a \equiv a(m, M) = \frac{M^r - m^r}{M^s - m^s}, \qquad b \equiv b(m, M) = \frac{M^s m^r - M^r m^s}{M^s - m^s}.$$

Note that $a > 0$ if and only if $rs > 0$, and $b > 0$ if and only if $s > 0$.

Lemma 2.1. *If X is a random variable satisfying $P\{m \leqslant X \leqslant M\} = 1$, with $m > 0$, and Z is a positive random variable, then*

$$r[EZX^r - aEZX^s - bEZ] \geqslant 0, \qquad \textit{for} \qquad r < s. \tag{2.1}$$

Equality holds if and only if $P\{X = m\} + P\{X = M\} = 1$.

Proof. To prove (2.1) it is sufficient to show that for $m \leqslant x \leqslant M$,

$$f(x) = r[x^r - ax^s - b] \geqslant 0, \qquad r < s, \tag{2.2}$$

for then

$$r[ZX^r - aZX^s - bZ] \geqslant 0, \qquad r < s, \tag{2.3}$$

with probability one, and (2.1) follows by integrating (2.3).

To prove (2.2), note that $f'(x) = rx^{r-1}(r - asx^{s-r}) = 0$ has a unique zero in $(0, \infty)$, so that $f(x)$ has at most two zeros in $(0, \infty)$. But by the choice of a and b, $f(m) = f(M) = 0$, so that $f(x)$ is of one sign for $m \leqslant x \leqslant M$. If $0 < r < s$, then $f(0) = -rb < 0$ implies that $f(x) \geqslant 0$, $m \leqslant x \leqslant M$. If $r < 0$ and $r < s$, then $\lim_{x \downarrow 0} f(x) \leqslant 0$ implies that $f(x) \geqslant 0$, $m \leqslant x \leqslant M$.

Equality holds in (2.1) if and only if equality holds in (2.3) with probability one, the condition for which is immediate. ∎

We note that the positivity of Z can be weakened to nonnegativity, but with some resulting minor changes in the conditions for equality.

LEMMA 2.2. *If* $P\{m \leqslant X \leqslant M\} = 1$ *with* $m > 0$, *and* $P\{Z > 0\} = 1$, *then for* $r < s$,

$$\frac{(EZX^s)^{1/s}}{(EZX^r)^{1/r}} \leqslant \kappa(EZ)^{(1/s)-(1/r)}, \tag{2.4}$$

where

$$\kappa = \left[\frac{r(\delta^s - \delta^r)}{(s-r)(\delta^r - 1)}\right]^{1/s} \left[\frac{s(\delta^s - \delta^r)}{(s-r)(\delta^s - 1)}\right]^{-1/r},$$

and $\delta = M/m$. *Equality holds if and only if* $P\{X = m \text{ or } X = M\} = 1$ *and* $EZX^s = rb[a(s-r)]^{-1} EZ$.

PROOF. From (2.1), it follows that

$$\frac{(EZX^s)^{1/s}}{(EZX^r)^{1/r}} \leqslant \frac{(EZX^s)^{1/s}}{(aEZX^s + bEZ)^{1/r}} \equiv \varphi(EZX^s).$$

It is easily seen that the unrestricted maximum of $\varphi(y)$ occurs at

$$y_0 = rb[a(s-r)]^{-1} EZ,$$

and $\varphi(y_0)$ is the bound of (2.4).

Since $m \leqslant X \leqslant M$ with probability one, it might appear that (2.4) could be improved if φ is maximized subject to the restriction

$$sm^s EZ \leqslant sEZX^s \equiv sy \leqslant sM^s EZ.$$

However, we find by a straightforward check that y_0 does satisfy this restriction. We conclude that equality occurs if and only if equality holds in (2.1), and in addition, $\varphi(EZX^s) = \varphi(y_0)$. ∎

III. REVERSAL OF THE LYAPUNOV, HÖLDER, AND MINKOWSKI INEQUALITIES

To apply the inequalities of Section II, we need only specify the choice of X and Z. In particular, we consider the cases

$$(X, Z) = (X, X^t) \qquad \text{and} \qquad (X, Z) = (VU^{-1}, V^{-r}U^s).$$

When $(X, Z) = (X, X^t)$, we obtain from (2.1) and (2.4) with $t = u$, $r + t = v$, $s + t = w$, that if $u < v < w$, then

$$(M^{w-u} - m^{w-u})\,\mu_v - (M^{v-u} - m^{v-u})\,\mu_w - (M^{w-u}m^{v-u} - M^{v-u}m^{w-u})\,\mu_u \geqslant 0, \tag{3.1}$$

$$\mu_v^{w-u} \geqslant \gamma \mu_u^{w-v} \mu_w^{v-u}, \tag{3.2}$$

where

$$\gamma = \left[\frac{(\delta^{w-u} - \delta^{v-u})\,(w-u)}{(\delta^{w-u} - 1)\,(w-v)}\right]^{w-u} \left[\frac{(\delta^{w-u} - \delta^{v-u})\,(v-u)}{(\delta^{v-u} - 1)\,(w-v)}\right]^{-(v-u)}.$$

Inequality (3.2) is the reversal of Lyapunov's inequality mentioned in the introduction.

When $(X, Z) = (VU^{-1}, V^{-r}U^s)$, we obtain from (2.1) and (2.4) that if $P\{m \leqslant VU^{-1} \leqslant M\} = 1$, with $m > 0$, then for $r < s$,

$$r[EU^{s-r} - aEV^{s-r} - bEU^sV^{-r}] \geqslant 0, \tag{3.3}$$

$$(EU^{s-r})^{-1/r}\,(EV^{s-r})^{1/s} \leqslant \kappa(EU^sV^{-r})^{(1/s)-(1/r)}. \tag{3.4}$$

An equivalent formulation may be obtained by writing $EU^t = \int U^t\,d\lambda$, $EV^t = \int V^t d\lambda$. When λ is uniform on $[\alpha, \beta]$ or $\{1, 2, \cdots, n\}$, and $r = -1$, $s = 1$, both (3.3) and (3.4) have been obtained by Diaz and Metcalf [6].

To obtain a reversal of Hölder's inequality, let $f^p = U^{s-r}$, $g^q = V^{s-r}$, $p = (s-r)/s$, $r < 0$, and $\theta = \delta^{s-r}$. A direct substitution in (3.4) and (3.3) yields

THEOREM 3.1. *Let f and g be nonnegative functions such that $l \leqslant f^{-p}g^q \leqslant l\theta$ and $\int fg\,d\lambda$ exists. If $p \geqslant 1$ and $(1/p) + (1/q) = 1$, then*

$$\int fg d\lambda \geqslant c \left(\int f^p d\lambda\right)^{1/p} \left(\int g^q d\lambda\right)^{1/q}, \tag{3.5}$$

and

$$l^{1/q}\theta^{1/q}(\theta^{1/p} - 1) \int f^p d\lambda + l^{-1/p}(\theta^{1/q} - 1) \int g^q d\lambda \leqslant (\theta - 1) \int fg d\lambda, \tag{3.6}$$

where

$$c \equiv c(p, q, \theta) = \frac{q^{1/q}p^{1/p}\theta^{1/pq}(\theta^{1/q} - 1)^{1/q}\,(\theta^{1/p} - 1)^{1/p}}{(\theta - 1)}.$$

Just as Minkowski's inequality is derived from Hölder's inequality, we obtain from (3.5)

$$\left(\int (f+g)^p\,d\lambda\right)^{1/p} \geqslant c\left[\left(\int f^p d\lambda\right)^{1/p} + \left(\int g^p d\lambda\right)^{1/p}\right]. \tag{3.7}$$

REMARK. Inequalities (3.5) and (3.7) are reversed for $0 < p \leqslant 1$. It should be noted that the proofs did not require λ to be finite.

A particular case of interest is $p = q = 2$, which yields the reversal of the Cauchy inequality due to Schweitzer [7], namely,

$$\left(\int fg d\lambda\right)^2 \geqslant \frac{4\theta^{1/2}}{(\theta^{1/2}+1)^2}\left(\int f^2 d\lambda\right)\left(\int g^2 d\lambda\right). \tag{3.8}$$

Some interesting results involving the geometric mean can be obtained from the various inequalities. From (3.1) with $v = 0$, we obtain

$$\left[\frac{(M^w m^u - m^w M^u) - (m^u - M^u)\mu_w}{M^w - m^w}\right]^{1/u} \leqslant \mu_u^{1/u},$$

and a similar upper bound for $\mu_w^{1/w}$. After taking limits as $u \to 0$ (and $w \to 0$), we obtain

$$r\{(M^r - m^r)e^{E\log X} - (EX^r)\log(M/m) - M^r \log m + m^r \log M\} \geqslant 0. \tag{3.9}$$

By a similar argument, from (3.2), we obtain with $\rho = \delta^{r(\delta^r - 1)^{-1}}$,

$$rE\log X - \log EX^r + \log\rho - \log\log\rho - 1 \geqslant 0, \tag{3.10}$$

a result also obtained by Cargo and Shisha [4].

The two choices of (X, Z) in (2.1) and (2.4) which we have utilized in this section exemplify the methods. However, other choices, e.g.,

$$(X, Z) = (UV, U^t V^t) \qquad \text{or} \qquad (X, Z) = (X, e^{tX})$$

lead to other types of inequalities.

IV. MATRIX THEORETIC INTERPRETATIONS

If A is an $n \times n$ Hermitian matrix with (real) eigenvalues $\theta_1, \cdots, \theta_n$, and if x is a unit row vector, then

$$xA^r x^* = (x\Gamma)D_\theta^r(\Gamma^* x^*) \equiv yD_\theta^r y^* = \sum y_i \bar{y}_i \theta_i^r,$$

where Γ is unitary, and $D_\theta = \operatorname{diag}(\theta_1, \cdots, \theta_n)$. Since $y_i \bar{y}_i$ are nonnegative and add to unity, $xA^r x^*$ can be regarded as the rth moment, μ_r, of a random variable taking values on the eigenvalues of A. (For another application see Marshall and Olkin [8].) With this interpretation we obtain from (3.1) and (3.2), that if A is positive definite and $0 < m \leqslant \theta_i \leqslant M$, $i = 1, \cdots, n$, then

$$\begin{aligned}(M^{w-u} - m^{w-u})\,xA^v x^* &- (M^{v-u} - m^{v-u})\,xA^w x^* \\ &\geqslant (M^{w-u}m^{v-u} - M^{v-u}m^{w-u})\,xA^u x^*,\end{aligned} \tag{4.1}$$

$$(xA^v x^*)^{w-u} \geqslant \gamma(xA^u x^*)^{w-v}(xA^w x^*)^{v-u}. \tag{4.2}$$

Of particular interest is the choice $u = -1$, $v = 0$, $w = 1$, which in (4.2) yields the more familiar form of Kantorovich's inequality. From (4.1) we get a strengthened version of Kantorovich's inequality,

$$xA^{-1}x^* \leqslant (M + m - xAx^*)/Mm. \tag{4.3}$$

If A and B are commutative Hermitian matrices, then they can be simultaneously diagonalized by a unitary matrix, so that

$$xA^{\alpha}x^* = \sum y_i\bar{y}_i\theta_i^{\alpha}, \qquad xB^{\beta}x^* = \sum y_i\bar{y}_i\varphi_i^{\beta},$$

and

$$xA^{\alpha}B^{\beta}x^* = \sum y_i\bar{y}_i\theta_i^{\alpha}\varphi_i^{\beta}.$$

Consequently, these quadratic forms may be regarded as moments $E\Theta^{\alpha}$, $E\Phi^{\beta}$, and $E\Theta^{\alpha}\Phi^{\beta}$, respectively, where

$$P\{\Theta = \theta_i\} = P\{\Phi = \varphi_i\} = P\{\Theta = \theta_i, \Phi = \varphi_i\} = y_i\bar{y}_i, \qquad i = 1, \cdots, n.$$

Using (3.3) and (3.4) we obtain, for $r < s$,

$$r[xA^{s-r}x^* - axB^{s-r}x^* - bxA^sB^{-r}x^*] \geqslant 0, \tag{4.4}$$

$$[xB^{s-r}x^*]^{1/s}[xA^{s-r}x^*]^{-1/r} \leqslant \kappa[xA^sB^{-r}x^*]^{(1/s)-(1/r)}. \tag{4.5}$$

Here the roots θ_i and φ_i of A and B satisfy $m \leqslant \varphi_i/\theta_i \leqslant M$. The case $r = -1$ and $s = 1$ in (4.5) was obtained in [5].

Some further results can be obtained by considering compound matrices. The pth compound $B_{(p)}$ of a $k \times l$ matrix B is defined for $p \leqslant \min(k, l)$ to be the $\binom{k}{p} \times \binom{l}{p}$ matrix of pth order minors of B arranged in lexicographic order (see [9] or [10]). We write $\operatorname{tr}_p B \equiv \operatorname{tr} B_{(p)}$, and make use of the Cauchy-Binet Theorem $(BC)_{(p)} = B_{(p)}C_{(p)}$.

As before, let $A : n \times n$ be Hermitian, with characteristic roots $\theta_1 \geqslant \cdots \geqslant \theta_n$; then the characteristic roots $\lambda_1 \geqslant \cdots \geqslant \lambda_{\binom{n}{p}}$ of $A_{(p)}$ are the products of the θ_i taken p at a time.

Let $X : l \times n$ satisfy $\operatorname{tr}_p XX^* = 1$ (so that rank $(X) \geqslant p$). Then

$$XA^kX^* = X\Gamma D^k\Gamma^*X^* \equiv YD^kY^*.$$

Thus

$$\operatorname{tr}_p YD^kY^* = \operatorname{tr} D^k_{(p)}(Y^*Y)_{(p)} \equiv \operatorname{tr} D^k_{(p)}Z = \sum \lambda_i^k z_{ii}.$$

Since

$$\operatorname{tr} Z = \operatorname{tr}_p Y^*Y = \operatorname{tr}_p YY^* = \operatorname{tr}_p XX^* = 1, \text{ and } z_{ii} > 0,$$

we may regard $\operatorname{tr}_p YD^kY^*$ as the kth moment of a random variable taking values on the $\binom{n}{p}$ points λ_i, i.e., on the products of the roots of A taken p at a time.

With this interpretation, we obtain from (3.1) and (3.2) that if $A : n \times n$ is Hermitian with characteristic roots $\lambda_1, \cdots, \lambda_n$, $m \leqslant \lambda_i \leqslant M$, and if $X : l \times n$ satisfies $\mathrm{tr}_p\, XX^* = 1$, then

$$(M^{w-u} - m^{w-u})\,\mathrm{tr}_p\, XA^vX^* - (M^{v-u} - m^{v-u})\,\mathrm{tr}_p\, XA^wX^* \geqslant (M^{w-u}m^{v-u} - M^{v-u}m^{w-u})\,\mathrm{tr}_p\, XA^uX^*, \tag{4.6}$$

$$(\mathrm{tr}_p\, XA^vX^*)^{w-u} \geqslant \gamma(\mathrm{tr}_p\, XA^uX^*)^{w-v}\,(\mathrm{tr}_p\, XA^wX^*)^{v-u}. \tag{4.7}$$

In the case $l = p$, $\mathrm{tr}_p\, XA^kX^* = \det(XA^kX^*)$, and these inequalities take a particularly simple form. The special case of (4.7) with

$$(u, v, w) = (k - 1, k, k + 1)$$

was obtained by Schopf [11].

V. Related Inequalities

In (2.2) we defined $f(x) = r[x^r - ax^s - b]$ which for $r = -1$ and $s = 1$ becomes

$$f(x) = \frac{(x - m)(M - x)}{mMx} = \frac{M + m - x}{mM} - x^{-1},$$

so that $f(x) \geqslant 0$ is immediate. However, the nonnegativity also follows from the convexity of x^{-1}. The essential point is that the function $x^{-1} \leqslant ax + b$, $m \leqslant x \leqslant M$, with equality at the end points $x = m$ and $x = M$ by choice of a and b. This suggests that we consider functions $g(x)$ and $h(x)$ satisfying $g(x) \leqslant ah(x) + b$, $m \leqslant x \leqslant M$, $g(m) = ah(m) + b$, and $g(M) = ah(M) + b$. The latter two conditions determine

$$a = \frac{g(M) - g(m)}{h(M) - h(m)}, \qquad b = \frac{g(m)\,h(M) - g(M)\,h(m)}{h(M) - h(m)}. \tag{5.1}$$

If in the interval $[m, M]$, either h is monotone and gh^{-1} is convex, or g is monotone and hg^{-1} is concave, then $g(x) \leqslant ah(x) + b$, and hence

$$Eg(X) \leqslant aEh(X) + b, \tag{5.2}$$

where a and b are defined in (5.1).

If $Eh(X) > 0$ and $a < 0$, then from (5.2),

$$Eh(X)\,Eg(X) \leqslant a[Eh(X)]^2 + bEh(X) \leqslant \frac{b^2}{4a} = \frac{[g(m)\,h(M) - h(m)\,g(M)]^2}{4[g(M) - g(m)]\,[h(m) - h(M)]}. \tag{5.3}$$

Equality in (5.3) can be achieved by a distribution concentrating on m and M.

Of course, if $g(x) = x^{-1}$ and $h(x) = x$, then (5.3) is just Kantorovich's inequality.

Inequality (5.2) is a simple extension of the case that $h(x) = x$ and g is convex; in this special form it is given by Edmundson [12], andmay be regarded as a reversal of Jensen's inequality. Multivariate extensions of Edmundson's results have been obtained by Madansky [13]. In both instances, the results were obtained using convexity properties of moment spaces.

We now consider the multivariate case from the point of view used to derive (5.2). Let $X = (X_1, \cdots, X_n)$ be a random vector with $EX_i = \mu_i$, and $P\{0 \leqslant X_i \leqslant 1\} = 1$, $i = 1, \cdots, n$. Let g be a function defined on the unit hypercube $0 \leqslant x_i \leqslant 1$, $i = 1, \cdots, n$, with the property that $g(x_1, \cdots, x_n)$ is convex in each x_i for fixed x_α, $\alpha \neq i$.

An upper bound for $Eg(X)$ in terms of the μ_i may be obtained as follows. Let $\mathscr{H}$ be the class of functions h on $0 \leqslant x_i \leqslant 1$, $i = 1, \cdots, n$ such that $g(x) \leqslant h(x)$ and $Eh(X)$ is a function of the μ_i. For any $h \in \mathscr{H}$, we have the inequality $Eg(X) \leqslant Eh(X)$, and hence

$$Eg(X) \leqslant \inf_{h \in \mathscr{H}} Eh(X).$$

If $Eg(X) = Eh_0(X)$, $h_0 \in \mathscr{H}$, for some random vector satisfying the moment conditions, then

$$\inf_{\mathscr{H}} Eh(X) = Eh_0(X).$$

In the case g is invariant under permutations of its arguments, we obtain inequalities illustrating the method, when the x_i may be dependent, and when the x_i are independent.

The following Lemma is useful in proving that $g(x) \leqslant h(x)$.

LEMMA 5.1. *Let Z_m be the set of m-dimensional vectors with components 0 or 1. Suppose $g(x_1, \cdots, x_n)$ is convex in each x_i for fixed x_α, $\alpha \neq i$, and $h(x_1, \cdots, x_n)$ is linear in each x_i for fixed x_α, $\alpha \neq i$. If $g(z) \leqslant h(z)$, $z \in Z_n$, then $g(x) \leqslant h(x)$ for all x in the unit hypercube.*

PROOF. Let $x = (x_1, \cdots, x_n)$ be an arbitrary but fixed vector in the unit hypercube, and $x_{(k)} = (x_1, \cdots, x_k)$.

If $z_{(n-1)} \in Z_{n-1}$, then by the linearity of h and convexity of g,

$$h(x_{(1)}, z_{(n-1)}) \geqslant g(x_{(1)}, z_{(n-1)}).$$

Assume that for all $z_{(n-k)} \in Z_{n-k}$,

$$h(x_{(k)}, z_{(n-k)}) \geqslant g(x_{(k)}, z_{(n-k)}).$$

Then

$$\begin{aligned} h(x_{(k+1)}, z_{(n-k-1)}) &= h(x_{(k)}, x_{k+1}, z_{(n-k-1)}) \\ &= x_{k+1} h(x_{(k)}, 1, z_{(n-k-1)}) + (1 - x_{k+1}) h(x_{(k)}, 0, z_{(n-k-1)}) \\ &\geqslant x_{k+1} g(x_{(k)}, 1, z_{(n-k-1)}) + (1 - x_{k+1}) g(x_{(k)}, 0, z_{(n-k-1)}) \\ &\geqslant g(x_{(k+1)}, z_{(n-k-1)}), \end{aligned}$$

and by induction, $h(x) \geqslant g(x)$. ∎

When $X_1, \cdots, X_n$ are independent, we consider the function

$$h(x) = b_0 + b_1 \sum x_i + b_2 \sum x_i x_j + \cdots + b_n \prod x_i ,$$

where the b_i are determined so that $g_k \equiv g(v_k) = h(v_k)$, $k = 0, 1, \cdots, n$, where v_k is the vector with first k components equal to 1, and remaining components equal to zero. These equations may be written in the form

$$g \equiv (g_0, g_1, \cdots, g_n) = (b_0, \cdots, b_n) \begin{bmatrix} \binom{0}{0} & \binom{1}{0} & \binom{2}{0} & \cdots & \binom{n}{0} \\ 0 & \binom{1}{1} & \binom{2}{1} & \cdots & \binom{n}{1} \\ 0 & 0 & \binom{2}{2} & \cdots & \binom{n}{2} \\ \vdots & \vdots & \vdots & & \vdots \\ 0 & 0 & 0 & \cdots & \binom{n}{n} \end{bmatrix} \equiv bU.$$

It is easily seen that $U^{-1} \equiv (u^{ij})$ has elements $u^{ij} = (-1)^{i+j} u_{ij}$, so that

$$b_k = \binom{k}{k} g_k - \binom{k}{k-1} g_{k-1} + \binom{k}{k-2} g_{k-2} - \cdots + (-1)^k \binom{k}{0} g_0 ,$$

$$k = 0, 1, \cdots, n.$$

By the Lemma, it remains to show that $g(z) \leqslant h(z)$ for $z \in Z_n$. Note that each $z \in Z_n$ is equal to $v_k P$ for some k and some permutation matrix P. Since both g and h are invariant under permutations,

$$g(z) = g(v_k) = h(v_k) = h(z).$$

To summarize, we have obtained

THEOREM 5.2. *Let $X = (X_1, \cdots, X_n)$ be a random vector with independent components such that $P\{0 \leqslant X_i \leqslant 1\} = 1$ and $EX_i = \mu_i$, $i = 1, \cdots, n$. Let*

$g(x_1, \cdots, x_n)$ be defined and convex in each x_i, $0 \leqslant x_i \leqslant 1$, with the further property that g is invariant under permutations of the x_i. Then

$$Eg(X_1, \cdots, X_n) \leqslant g_0 + \sum_{k=1}^{n} \left[\binom{k}{k} g_k - \binom{k}{k-1} g_{k-1} + \cdots + (-1)^k \binom{k}{0} g_0 \right] s_k, \tag{5.4}$$

where $g_j \equiv g(v_{(j)})$, $v_{(j)}$ is the vector with first j components one and remaining components zero, and where s_k is the kth *elementary symmetric function of the μ_i.*

Equality is obtained if $X_1, \cdots, X_n$ are independent with

$$P\{X_i = 1\} = \mu_i = 1 - P\{X_i = 0\}.$$

To see this note that this distribution of X is concentrated at points x such that $g(x) = h(x)$.

In the special case $g(x) = \Pi_1^n t(x_i)$ with t convex, b_k takes the simple form $b_k = t_0^{n-k}(t_1 - t_0)^k$, $t_o \equiv t(0)$, $t_1 \equiv t(1)$ and hence (5.4) becomes

$$E\prod_1^n t(X_i) \leqslant t_0^n + t_0^{n-1}(t_1 - t_0) \sum \mu_i + t_0^{n-2}(t_1 - t_0)^2 \sum_{i<j} \mu_i\mu_j + \cdots + (t_1 - t_0)^n (\mu_1 \cdots \mu_n). \tag{5.5}$$

This result for $g(x) = \exp(x_1 + x_2)$ was obtained by Madansky [13].

We now remove the condition of independence but retain the choice $g(x) = \Pi_1^n t(x_i)$. In this case

$$h(x) = a_0 + \sum_1^n a_i x_i,$$

and we determine the a_i by $g_k \equiv g(v_{(k)}) = h(v_{(k)})$, $k = 0, 1, \cdots, n$. These equations may be written as

$$g \equiv (g_0, \cdots, g_n) = (a_0, \cdots, a_n) \begin{pmatrix} 1 & 1 & 1 & \cdots & 1 \\ 0 & 1 & 1 & \cdots & 1 \\ 0 & 0 & 1 & \cdots & 1 \\ \vdots & & & & \\ 0 & 0 & 0 & \cdots & 1 \end{pmatrix} \equiv aT.$$

Since $T^{-1} \equiv (t^{ij})$ has elements $t^{ii} = 1$, $t^{i-1,i} = -1$, and all other elements zero, it is easily seen that

$$a = (g_0, g_1 - g_0, \cdots, g_n - g_{n-1}).$$

If $a_1 \leqslant \cdots \leqslant a_n$, or equivalently,

$$g_k - g_{k-1} \geqslant g_{k-1} - g_{k-2}, \qquad k = 1, \cdots, n, \tag{5.6}$$

then $h(v_{(k)}) \leqslant h(v_{(k)}P)$ for any permutation matrix P, and hence, by the permutation invariance of g,

$$g(v_{(k)}P) = g(v_{(k)}) = h(v_{(k)}) \leqslant h(v_{(k)}P).$$

By the Lemma, $g(x) \leqslant h(x)$ for all x in the unit hypercube.

The monotonicity condition (5.6) becomes

$$[t(1)]^k [t(0)]^{n-k} - 2[t(1)]^{k-1} [t(0)]^{n-k+1} + [t(1)]^{k-2} [t(0)]^{n-k+2} \geqslant 0,$$

which is clearly satisfied if $t(0) \geqslant 0$, $t(1) \geqslant 0$.

To summarize, we have

THEOREM 5.3. *Let $(X_1, \cdots, X_n)$ be a random vector with $P\{0 \leqslant X_i \leqslant 1\} = 1$ and $EX_i = \mu_i$, $i = 1, \cdots, n$. Let $g(x) = \prod_1^n t(x_i)$, where t is convex on* $[0, 1]$, *$t(0) \geqslant 0$, $t(1) \geqslant 0$. Then*

$$Eg(X_1, \cdots, X_n) \leqslant g_0 + \sum_1^n (g_i - g_{i-1}) \mu_i . \tag{5.7}$$

The order of the X_i is arbitrary, and if we assume $1 \geqslant \mu_1 \geqslant \cdots \geqslant \mu_n \geqslant 0$, then equality may be attained for $P\{X = v_{(k)}\} = p_k$, $k = 0, \cdots, n$, where

$$(p_0, \cdots, p_n) = (1 - \mu_1, \mu_1 - \mu_2, \cdots, \mu_{n-1} - \mu_n, \mu_n)$$

is determined by $(p_0, \cdots, p_n) T' = (1, \mu_1, \cdots, \mu_n)$. Equality is attained because this distribution of X is concentrated at the points x for which $g(x) = h(x)$.

The result for $g(x) = \exp(x_1 + x_2)$ was obtained by Madansky [13].

REMARK. Note that (5.7) holds for any $g(x)$ which is invariant under permutations, convex in each element, and satisfying the monotonicity condition (5.6). This class of g's is convex, and in addition to $g(x) = \prod_1^n t(x_i)$, includes, e.g., $g(x) = \sum_1^n t(x_i)$.

The condition (5.6) was derived from the choice of T, which in turn is dependent upon the choice of n vertices from the 2^n possibilities for which we require $h(x) = g(x)$. The proper choice of vertices depends on the function g and also upon the μ_i. Thus, for example, if $0 \leqslant g_0 \leqslant \cdots \leqslant g_n$, $\sum_1^n \mu_i \leqslant 1$, then

$$T = \begin{pmatrix} 1 & v_{(n)} \\ 0 & I \end{pmatrix}$$

yields the sharp inequality

$$Eg(X_1, \cdots, X_n) \leqslant g_0 + \sum_1^n g_i \mu_i .$$

Two such functions g are $\sum_1^n x_i^k$ and $(\sum_1^n x_i)^k$.

Note added in proof. An alternative proof of (5.4) was suggested by W. Hoeffding, and is based on the fact that if $g(x_1, \cdots, x_n)$ is convex in each x_i for fixed x_α $(\alpha \neq i)$, then

$$g(x_1, \cdots, x_n) \leq \sum_{i_1=0}^{1} \cdots \sum_{i_n=0}^{1} g(i_1, \cdots, i_n) \prod_{j=1}^{n} x_j^{i_j}(1-x_j)^{1-i_j} .$$

References

1. Hardy, G. A., Littlewood, J. E., and Pólya, G. "Inequalities," 2nd ed. Cambridge Univ. Press, Cambridge, 1959.
2. Barlow, R. E., Marshall, A. W., and Proschan, F. Properties of probability distributions with monotone hazard rate. *Ann. Math. Statist.* **34**, 375-389 (1963).
3. Karlin, S., Proschan, F., and Barlow, R. E. Inequalities of Pólya frequency functions. *Pacific J. Math.* **11**, 1023-1033 (1961).
4. Cargo, G. T., and Shisha, O. Bounds on ratios of means. *J. Res. Natl. Bur. Std.* **66B**, 169-170 (1962).
5. Greub, W., and Rheinboldt, W. On a generalization of an inequality of L.V. Kantorovich. *Proc. Am. Math. Soc.* **10**, 407-415 (1959).
6. Diaz, J. B., and Metcalf, F. T. Stronger forms of a class of inequalities of G. Pólya-G. Szegö, and L. V. Kantorovich. *Bull. Am. Math. Soc.* **69**, 415-418 (1963).
7. Schweitzer, P. Egy Egyenlötlenség az Arithmetikai Középértékröl, (Hungarian) (An inequality concerning the arithmetic mean). *Mat. Fiz. Lapok* **23**, 257-261 (1914).
8. Marshall, A. W., and Olkin, I. Inclusion theorems for eigenvalues from probability inequalities. *Numer. Math.* **6** (1964).
9. Aitken, A. C. "Determinants and Matrices," 5th ed. Oliver and Boyd, Edinburgh, 1948.
10. Wedderburn, J. M. "Lectures on Matrices." *Am. Math. Soc. Colloq. Publ.* Vol. 17 (1934).
11. Schopf, A. H. On the Kantorovich inequality. *Numer. Math.* **2**, 344-346 (1960).
12. Edmundson, H. P. Bounds on the expectation of a convex function of a random variable, p. 982. The RAND Corporation, 1957.
13. Madansky, A. Bounds on the expectation of a convex function of a multivariate random variable. *Ann. Math. Statist.* **30**, 743-746 (1959).