The Annals of Statistics
1982, Vol. 10, No. 1. 297–301

AN INEQUALITY COMPARING SUMS AND MAXIMA WITH APPLICATION TO BEHRENS-FISHER TYPE PROBLEM

BY SIDDHARTHA R. DALAL[1] AND PETER FORTINI

Bell Laboratories and American Cyanamid Company

A sharp inequality comparing the probability content of the ℓ_1 ball and that of ℓ_∞ ball of the same volume is proved. The result is generalized to bound the probability content of the ℓ_p ball for arbitrary $p \geq 1$. Examples of the type of bound include

$$P\{(|X_1|^p + |X_2|^p)^{1/p} \leq c\} \geq F^2(c/2^{1/2p}), \qquad p \geq 1,$$

where X_1, X_2 are independent each with distribution function F. Applications to multiple comparisons in Behrens-Fisher setting are discussed. Multivariate generalizations and generalizations to non-independent and non-exchangeable distributions are also discussed. In the process a majorization result giving the stochastic ordering between $\Sigma a_i X_i$ and $\Sigma b_i X_i$, when $(a_1^2, a_2^2, \cdots, a_n^2)$ majorizes $(b_1^2, b_2^2, \cdots, b_n^2)$, is also proved.

1. Introduction and summary. For independent random variables X and Y, it is often difficult to evaluate the distribution of $|X| + |Y|$, while the distribution of $\max(|X|, |Y|)$ is easily obtained. In Section 2 of this paper, conditions for a stochastic ordering between the two are found using majorization results. Several illustrative examples are included in Section 3. An inequality obtained in this section is applied in Section 4 to a Behrens-Fisher type multiple comparisons problem (Dalal, 1978). In the last section, a multivariate analogue to Theorem 1 of Section 2 is studied. As an example, it is shown that, for independent standard normals, $|X_1| + \cdots + |X_k|$ is stochastically larger than $\sqrt{k}\max(|X_1|, \cdots, |X_k|)$.

2. Basic inequality. Before stating the basic inequality, notions associated with majorization are defined. An n-vector $\mathbf{x} = (x_1, x_2, \cdots, x_n)$ *majorizes* an n-vector $\mathbf{y}$ if $\Sigma y_i = \Sigma x_i$ and for a decreasing arrangement of coordinates of $\mathbf{x}$ and $\mathbf{y}$, $\sum_1^k x_i \geq \sum_1^k y_i$ $(k = 1, 2, \cdots, n)$. A permutation symmetric function g of n variables is *Schur-convex* if $g(\mathbf{x}) \geq g(\mathbf{y})$ whenever $\mathbf{x}$ majorizes $\mathbf{y}$; g is *Schur-concave* if $g(\mathbf{x}) \leq g(\mathbf{y})$. General discussion of majorization theory and some probabilistic implications can be found in Marshall and Olkin (1979).

THEOREM 1. *Let (X, Y) be a nonnegative random vector with symmetric density $f(x, y)$. If $f(\sqrt{x}, \sqrt{y})$ is Schur-convex, then for any real c,*

$$P(X + Y \leq c) \geq P\{\sqrt{2}\max(X, Y) \leq c\}. \tag{1}$$

For Schur-concave $f(\sqrt{x}, \sqrt{y})$, the following reverse inequality holds:

$$P\{X + Y \leq c\} \leq P\{\sqrt{2}\max(X, Y) \leq c\}. \tag{2}$$

PROOF. We prove the theorem for Schur-convex $f(\sqrt{x}, \sqrt{y})$. The proof for Schur-concave $f(\sqrt{x}, \sqrt{y})$ is similar. From the definition it follows that for any nonnegative (x, y)

Received March 1979; revised June 1981.

[1] This author's research partially supported by N.S.F. Grant MCS 78-02160 at Rutgers University.

AMS 1980 *subject classifications.* Primary, 62F25, 60E15.

Key words and phrases. Majorization, inequality, sums of powers, multiple comparisons, Behrens-Fisher problem.

and (w, z) such that $x^2 + y^2 = w^2 + z^2$ and $\max(x, y) \le \max(w, z)$, (x^2, y^2) is majorized by (w^2, z^2). Consequently $f(x, y) \le f(w, z)$. Thus, along any circle $x^2 + y^2 = r^2$ the density increases as one moves away from the line $x = y$. The opposite conclusion holds for Schur-concave $f(\sqrt{x}, \sqrt{y})$.

Let S_1 be the square region for which $\sqrt{2}\max(x, y) \le c$ and S_2 be the triangular region for which $x + y \le c$. Denoting the locus of the circle $x^2 + y^2 = r^2$ on the positive quadrant by C_r, it follows that the arc-length of $C_r \cap S_1$ coincides with that of $C_r \cap S_2$, while $C_r \cap S_1$ is closer to the line $x = y$ than $C_r \cap S_2$. Consequently by the monotonicity of f along the circle described above,

$$P\{\sqrt{2}\max(X, Y) \le c \,|\, X^2 + Y^2 = r^2\} \le P(X + Y \le c \,|\, X^2 + Y^2 = r^2).$$

The theorem follows by averaging over the distribution of $X^2 + Y^2$.

In applying this result the needed Schur-convexity, or concavity, of a permutation symmetric $f(\sqrt{x}, \sqrt{y})$ may be verified by Ostrowski's criteria, namely,

$$(x - y)\left(\frac{\delta}{\delta x} - \frac{\delta}{\delta y}\right) f(\sqrt{x}, \sqrt{y}) \ge 0,$$

or ≤ 0, respectively, for all (x, y). In the important special case of i.i.d. random variables with density f, $f(x, y) = f(x)f(y)$; thus by Remark 2.2 of Marshall and Olkin (1974), Schur-convexity or concavity of $f(\sqrt{x}, \sqrt{y})$ is equivalent to log-convexity or concavity of $f(\sqrt{x})$. We state this important result as follows.

COROLLARY 1. *For* i.i.d. *positive random variables X, Y with density f,* (1) *holds whenever* $\log f(\sqrt{x})$ *is convex. The reverse inequality* (2) *holds if* $\log f(\sqrt{x})$ *is concave.*

A generalization to nonexchangeable random variables X and Y is provided by

COROLLARY 2. *Let X and Y be positive random variables with joint density $f(x, y)$. Then $P(X + Y \le c) \ge P\{\sqrt{2}\max(X, Y) \le c\}$ iff $f(\sqrt{x}, \sqrt{y}) + f(\sqrt{y}, \sqrt{x})$ is Schur-convex. The inequality reverses for Schur-concavity.*

PROOF. $\frac{1}{2}f(\sqrt{x}, \sqrt{y}) + \frac{1}{2}f(\sqrt{y}, \sqrt{x})$ is the density function of exchangeable variables (X', Y') generated by taking X and Y in random order.

Finally we state without proof the following theorem comparing the minimum and the range of two i.i.d. nonnegative random variables. The proof, although involving different regions, follows from arguments similar to those employed in the proof of Theorem 1.

THEOREM 2. *Under the conditions of Theorem* 1, *for Schur-convex* $f(\sqrt{x}, \sqrt{y})$,

$$P(|X - Y| \le c) \le P\{\min(X, Y) \le c/\sqrt{2}\},$$

and

$$P((\tfrac{1}{2}|X^2 - Y^2|)^{1/2} \le c) \le P((|XY|)^{1/2} \le c).$$

The inequality between the probabilities reverses if $f(\sqrt{x}, \sqrt{y})$ *is Schur-concave.*

3. Examples. We briefly illustrate Theorem 1 for several pairs of random variables. a) *i.i.d. random variables and their powers.* For i.i.d. (S, T) bilateral exponential, Student's t, logistic, or Weibull with exponent ≤ 1, $\log f(\sqrt{s})$ is convex for $s \ge 0$, hence (1) holds by Corollary 1 with $(X, Y) = (|S|, |T|)$. For (S, T) i.i.d. normal, $N(0, \sigma^2)$, $\log f(\sqrt{s})$ is in fact linear so that equality holds in (2). In all these cases convexity of $\log f(\sqrt{x})$ implies that the density of X^q, $q \ge 1$, has the same property and thus

$$\Pr\{(|S|^q + |T|^q)^{1/q} \le c\} \ge \Pr\{\max(|S|, |T|) \le c/2^{1/2q}\}.$$

For the Weibull distribution with exponent ≥ 2 and Pearson type II distributions, $\log f(\sqrt{x})$ is concave and in these cases (2) holds.

b) *Non-i.i.d. random variables.*

i) *Bivariate normal variables.* For (S, T) bivariate normal with means 0, variances σ^2 and correlation ρ, $f(\sqrt{x}, \sqrt{y})$ is Schur-concave by Ostrowski's criterion and hence (2) holds with $(X, Y) = (|S|, |T|)$.

ii) *Scale mixtures of independent normals and powers of t.* Let $X^* = |aZ_1|^q$, $Y^* = |bZ_2|^q$, where Z_1, Z_2 are i.i.d. $N(0, 1)$, and $q \geq 1$. By Corollary 2 it can be shown that (1) holds for (X^*, Y^*). Now let $X = |t_\nu|^q$, $|Y| = |t_\mu|^q$ where t_ν and t_μ are independent Student's t with ν and μ degrees of freedom and $q \geq 1$. Then, since Student's t is a scale mixture of normals, the preceding argument about (X^*, Y^*) implies that

$$P\{(|t_\nu|^q + |t_\mu|^q)^{1/q} \leq c\} \geq P\{\max(|t_\nu|, |t_\mu|) \geq c/2^{1/2q}\}. \tag{3}$$

This result will be applied in the next section. Equality holds for $\mu = \nu = \infty$ and $q = 1$. To gauge the precision of (3) for the application we compare the two sides of (3) in Table 1 below for various c, q and $\nu = \mu$.

TABLE 1

Comparison of $P\{(|t_\nu|^q + |t_\mu|^q)^{1/q} \leq c\}$ *with the lower bound* (3)

$\nu = \mu =$		5		10		∞ (Normal)	
q	c	Value	Lower bound	Value	Lower bound	Value	Lower bound
1	1	.239	.239	.255	.254	—	—
	2	.617	.614	.663	.660	—	—
	3	.845	.833	.891	.884	—	—
	4	.941	.928	.971	.965	—	—
2	1	.344	.315	.368	.336	.393	.360
	2	.759	.717	.811	.768	.864	.823
	3	.925	.897	.962	.940	.989	.977
3	1	.374	.344	.400	.367	.428	.393
	2	.787	.748	.839	.801	.892	.856
	3	.935	.914	.970	.954	.993	.985

4. An application to multiple comparisons in Behrens-Fisher type problems. This investigation was originally motivated by the following Behrens-Fisher type situation. From each of k $N(\mu_i, \sigma_i^2)$ populations a random sample $(X_{i1}, \cdots, X_{in_i})$ is obtained. The purpose is to obtain simultaneous confidence intervals for all linear combinations of the μ_i's with exact confidence $1 - \alpha$.

For this problem, if $t(q, \alpha)$ denotes the upper α-point of the distribution of $\| t \|_q = (\Sigma |t_{n_i-1}|^q)^{1/q}$, then Dalal (1978) showed that

$$\Pr\{\Sigma a_i\mu_i \in \Sigma a_i\bar{X}_i \pm t(q, \alpha)(\Sigma |a_i|^p S_i^p/n_i^{p/2})^{1/p} \text{ for all } \mathbf{a} \text{ in } R^k\} = 1 - \alpha \tag{4}$$

where nonnegative (p, q) are such that $1/p + 1/q = 1$.

Among this infinite class of procedures, only the procedure corresponding to $p = 1$ $(q = \infty)$ was examined earlier as the percentage point $t(\infty, \alpha)$ (tabulated in Dalal, 1978) is easily obtained by solving $\Pi \Pr(|t_{n_i-1}| \leq c) = 1 - \alpha$. Now we examine the case $k = 2$, and compare various procedures supposing that comparisons of the form $\mu_1 \pm \mu_2$ are of primary interest.

From (3) it follows that for $k = 2$, $t(q, \alpha) \leq 2^{1/2q} t(\infty, \alpha)$. Using this bound as a critical constant conservative simultaneous intervals can be constructed for all linear combinations for any given p. Table 1 as well as more detailed computations indicate that this bound yields a good approximation to the true value.

The above bound may also be used for comparing various p's. If $W_{p,\alpha,n}$ is the width of the confidence intervals for $\mu_1 \pm \mu_2$ for a given p, then the bound on $t(q, \alpha)$ implies that

$W_{1,\alpha,n}/W_{p,\alpha,n} \geq (1 + \rho)/2^{1/2q}(1 + \rho^p)^{1/p}$ where $\rho^2 = n_1S_2^2/n_2S_1^2$ is the ratio of sample variances. Values of the lower bound can be computed for comparing various procedures. Numerical computations using this bound indicate that $p \geq 2$ is superior to $p = 1$ for $\mu_1 \pm \mu_2$ unless the ratio of standard errors exceeds about 7.

5. A multivariate analogue. Two multivariate results are proved here. Theorem 4 is of independent interest as it compares linear combination of i.i.d. positive random variables in a majorization sense. We use this result to prove Theorem 5, a multivariate extension of Theorem 1 for Schur-concavity. The hypotheses of both Theorems 4 and 5 are satisfied in the important case of absolute normal random variables.

THEOREM 4. *Let $X_1, \cdots, X_n$ be nonnegative* i.i.d. *random variables with density $f(x)$, where* $\log f(\sqrt{x})$ *is concave. Further, let nonnegative vectors* **a** *and* **b** *be such that $(a_1^2, a_2^2, \cdots, a_n^2)$ majorizes $(b_1^2, b_2^2, \cdots, b_n^2)$. Then for all $c \geq 0$*

$$P(\Sigma a_iX_i \leq c) \geq P(\Sigma b_iX_i \leq c).$$

PROOF. It suffices to prove the theorem for $n = 2$. The $n > 2$ case follows from $n = 2$ by applying the technique of T-transforms described in Marshall and Olkin (1974). Without loss of generality, assume $a_1^2 + a_2^2 = b_1^2 + b_2^2 = 1$, $0 \leq a_2 \leq b_2 \leq b_1 \leq a_1 \leq 1$.

As in the proof of Theorem 1, we argue conditionally on the value of $R = (X^2 + Y^2)^{1/2}$. The loci of points for which $a_1x_1 + a_2x_2 \geq c$ and $b_1x_1 + b_2x_2 \geq c$ are half-spaces both tangent to a circle of radius c. Each of $P(a_1X_1 + a_2X_2 \geq c \mid X_1^2 + X_2^2 = r^2)$ and $P(b_1X_1 + b_2X_2 \geq c \mid X_1^2 + X_2^2 = r)$ is the conditional probability content of an arc obtained as the intersection of the circle $X_1^2 + X_2^2 = r^2$ with the appropriate half-space. The second arc, however, is more centrally placed relative to $x_1 = x_2$ and has equal or greater length in the positive quadrant than the first. Hence, when $f(\sqrt{x_1}, \sqrt{x_1})$ is Schur-concave,

$$P(a_1X_1 + a_2X_2 \geq c \mid X_1^2 + X_2^2 = r) \leq P(b_1X_1 + b_2X_2 \geq c \mid X_1^2 + X_2^2 = r^2).$$

The theorem follows by averaging over $X_1^2 + X_2^2$ and reversing all inequalities.

Let S_n denote the sum and M_n the maximum of positive i.i.d. random variables $X_1, \cdots, X_n$ having density $f(x)$ and cdf $F(x)$. To derive a stochastic ordering between S_n and M_n we require the following.

LEMMA 1. *If* $\log f(\sqrt{x})$ *is concave and $f(x)/x$ is decreasing, then*
(i) $P(M_n + M_n' \leq c) \leq P(M_{2n} \leq c/\sqrt{2})$ *for* M_n, M_n' i.i.d.
(ii) $P(M_n + M_{n+1}' \leq c) \leq P(M_{2n+1} \leq c/\sqrt{2})$, *for M_n independent of M_{n+1}'.*

PROOF. Let $f_n(x) = nf(x)F^{n-1}(x)$ be the density function of M_n. Since $f(\sqrt{z})/\sqrt{z}$ is decreasing, $F(\sqrt{x}) = \int_0^x f(\sqrt{z})/2\sqrt{z}\ dz$ and hence $\log F(\sqrt{x})$ is concave. Therefore $\log f_n(\sqrt{x})$ is concave. (i) follows from Corollary 1, while (ii) follows from Corollary 2 and a short calculation.

Using the above results we finally prove a multivariate analogue of Theorem 1.

THEOREM 5. *Let $X_1, X_2, \cdots, X_n$ be* i.i.d. *positive random variables with density function $f(x)$. If* $\log f(\sqrt{x})$ *is concave and $f(x)/x$ is nonincreasing, then for all $c \geq 0$*

$$P(\Sigma X_i \leq c) \leq P\{\sqrt{n}\max(X_1, X_2, \cdots, X_n) \leq c\}.$$

PROOF. We proceed by induction on n. Abbreviate the relation $P(Z \leq c) \leq P(W \leq c)$ for all c by $Z \geq st W$. The conclusion, that $S_n \geq st \sqrt{n}\ M_n$, is trivial if $n = 1$, and is true by Corollary 1 if $n = 2$.

If n is even, set $n = 2p$. Write $S_n = \Sigma_1^p X_i + \Sigma_{p+1}^n X_i = S_p + S_p'$. Then

$$S_n = S_p + S_p' \geq st\sqrt{p}\, M_p + \sqrt{p}\, M_p' \geq st\sqrt{n}\, M_n$$

by the induction hypothesis and Lemma 1(i). It is easily shown that if $Z_1 \geq stZ_2$ and $W_1 \geq stW_2$ with (Z_1, Z_2) independent of (W_1, W_2), then $Z_1 + W_1 \geq stZ_2 + W_2$.

If n is odd, set $n = 2p + 1$. By Theorem 4 with $\mathbf{a}^2 = (n/2p, \cdots, n/2p, n/2p + 2, \cdots, n/2p + 2)$ and $\mathbf{b}^2 = (1, \cdots, 1)$, $S_n = S_p + S_{p+1}' \geq st\sqrt{n/2p}\; S_p + \sqrt{n/2p + 2}\;\; S_{p+1}'$. By induction and Lemma 1(ii), this is stochastically larger than $\sqrt{n}\;\; M_n$.

REFERENCES

DALAL, S. R. (1978). Simultaneous confidence procedures for univariate and multivariate Behrens-Fisher type problems. *Biometrika* **65** 221–225.

MARSHALL, A. W. and OLKIN, I. (1974). Majorization in multivariate distributions. *Ann. Statist.* **2** 1189–1200.

MARSHALL, A. W. and OLKIN, I. (1979). *Inequalities: Theory of Majorization and Its Applications.* Academic, New York.

BELL LABORATORIES
600 MOUNTAIN AVENUE
MURRAY HILL, NJ 07974

AMERICAN CYANAMID COMPANY
EASTON TURNPIKE
BOUND BROOK, NJ 08805