The Annals of Mathematical Statistics
1972, Vol. 43, No. 6, 1852-1860

RANDOM QUOTIENTS AND THE BEHRENS-FISHER PROBLEM

BY MORRIS L. EATON[1] AND RICHARD A. OLSHEN[2]

University of Chicago and Columbia University

Let $\mathscr{P}_n$ be the space of $n \times n$ positive definite symmetric matrices. If S_1 and S_2 are random matrices in $\mathscr{P}_n$, S_1 is a *better* α *denominator* than S_2 (written $S_1 \prec_{(\alpha)} S_2$) iff $U(x'S_1^{-1}x)^{\alpha/2} \ll_{\text{st}} U(x'S_2^{-1}x)^{\alpha/2}$ for all $x \in R^n$ where U is uniform on [0, 1], independent of S_1 and S_2, $\alpha > 0$, and "$\ll_{\text{st}}$" means stochastically smaller than. A principal result is this.

THEOREM. *Let* $S_1, \cdots, S_m$ *be exchangeable random matrices in* $\mathscr{P}_n$. *If* $0 < \alpha \leqq 2$, *then* $\sum_{i=1}^m \eta_i S_i \prec_{(\alpha)} \sum_{i=1}^m \psi_i S_i$ *provided* $(\psi_1, \cdots, \psi_m)$ *majorizes* $(\eta_1, \cdots, \eta_m)$.

This has applications in establishing probability inequalities for certain common test statistics. The results in this paper extend those of Lawton. (Some inequalities for central and non-central distributions. *Ann. Math. Statist.* (1965) **36** 1521-1525; Concentration of random quotients. *Ann. Math. Statist.* (1968) **39** 466-480.)

1. Introduction. Assume that X_{ji}, $j = 1, \cdots, n_i$, $i = 1, 2$ are random samples from two normal populations with (respective) means μ_i and variances σ_i^2. Throughout, "$\sim$" means "is distributed as." Since $\bar{X}_1 - \bar{X}_2 \sim N(\mu_1 - \mu_2, n_1^{-1}\sigma_1^2 + n_2^{-1}\sigma_2^2)$, the statistic

$$T_1^2 \equiv \frac{(\bar{X}_1 - \bar{X}_2)^2}{(n_1^{-1}s_1^2 + n_2^{-1}s_2^2)}$$

is of interest in deciding whether or not $\mu_1 = \mu_2$ (Scheffé (1970)). As always, $\bar{X}_i = n_i^{-1} \sum_j X_{ji}$ and $s_i^2 = (n_i - 1)^{-1} \sum_j (X_{ji} - \bar{X}_i)^2$. When $\mu_1 = \mu_2$, T_1^2 has the same distribution as

$$T_2^2 = \frac{Z^2}{\lambda(n_1 - 1)^{-1}\chi^2_{n_1-1} + (1 - \lambda)(n_2 - 1)^{-1}\chi^2_{n_2-1}} \tag{1.1}$$

where Z, $\chi^2_{n_1-1}$ and $\chi^2_{n_2-1}$ are independent; Z is standard normal and χ_m^2 is a chi-square random variable with m degrees of freedom. Here,

$$\lambda = n_1^{-1}\sigma_1^2/(n_1^{-1}\sigma_1^2 + n_2^{-1}\sigma_2^2)\,, \tag{1.2}$$

so $0 < \lambda < 1$. In studying T_2^2, Hsu (1938) showed that T_2^2 is stochastically smaller than $F_{1,\min\{n_1-1,n_2-1\}}$ and stochastically larger than F_{1,n_1+n_2-2}, where $F_{p,q}$ denotes an F-random variable with (p, q) degrees of freedom. Hájek (1962) generalized Hsu's results as follows. Let $Z, U_1, \cdots, U_k$ be independent where $Z \sim N(0, 1)$ and $U_i \sim \chi_1^2$. Suppose that $\lambda_1, \cdots, \lambda_k$ are positive numbers satisfying $\sum \lambda_i = 1$. Then

$$\frac{Z^2}{k^{-1}\sum_{i=1}^k U_i} \ll_{\text{st}} \frac{Z^2}{\sum_{i=1}^k \lambda_i U_i} \ll_{\text{st}} \frac{Z^2}{\nu^{-1}\sum_{i=1}^\nu U_i} \tag{1.3}$$

Received May 17, 1971; revised April 13, 1972.

[1] Now at the University of Minnesota.

[2] On leave from Stanford University. Now at the University of Michigan.

where ν is an integer no larger than $\min\{1/\lambda_i\}$. Here "$\ll_{\text{st}}$" means "stochastically smaller than." Mickey and Brown (1966) independently obtained Hsu's conclusion by an argument similar to Hájek's. Obviously, (1.3) is sharp.

It is interesting and pertinent to our work that Hájek actually proved more than (1.3). More precisely, he showed that if the vector $(\psi_1, \cdots, \psi_k)$ majorizes the vector $(\eta_1, \cdots, \eta_k)$ then

$$\frac{Z^2}{\sum \eta_i U_i} \ll_{\text{st}} \frac{Z^2}{\sum \psi_i U_i} \tag{1.4}$$

provided only that the U_i are exchangeable χ_1^2 variables (see Section 2 for definitions). He really deduced (1.3) from (1.4).

More recently, Lawton (1965; 1968) established (1.3) (actually, implicitly (1.4)) for a wider class of random variables than those considered by Hájek. It is assumed that the reader is familiar with Lawton's papers. In his paper, Lawton defines an ordering "$\prec$" for positive random variables. If $V_1 \prec V_2$, then V_1 is called a "uniformly better denominator" than V_2. Unfortunately, Lawton's definition of "$\prec$" depends on a rather complicated condition (called Condition (A)). In this paper, we present an alternative definition of $\prec$ in which Lawton's condition is replaced by others. We feel that our definition is perhaps the "right" definition—not only because of its simplicity but because it affords a natural generalization of the ordering to random positive definite symmetric matrices. Moreover, Corollary 3 of the paper of Olshen and Savage (1970) renders our conditions easy to check in those practical applications known to us. For a more detailed discussion and a conjecture concerning the relationship between our conditions (in the one-dimensional case) and Condition (A) see Section 3.

In Section 2, we give our definition of $\prec$ for random positive definite symmetric matrices. In fact, our definition of $\prec$ depends on a real parameter $\alpha > 0$ which is intimately connected with recent results in the cited paper by Olshen and Savage on α-unimodal random vectors. Matrix versions of (1.4) and of many of Lawton's results are presented in Section 2. At one point we employ a multivariate version of an inequality for convex functions due to Marshall and Proschan (1965).

In Section 3, we give an extension of Hájek's result to two dimensions which has applications to the Behrens–Fisher problem. Also, we discuss two conjectures, one of which has been mentioned; the other concerns Hotelling's T^2 statistic.

Conversations and correspondence, respectively, with Samuel Karlin and Ingram Olkin helped us to see the relationships of exchangeability and majorization to our work.

2. Main results. We begin this section with a discussion of stochastic ordering and α-unimodal random vectors. Let $\mathscr{G}$ be the set of all functions $g: [0, \infty) \to [0, 1]$ such that g is left continuous, non-increasing, continuous at 0 and satisfies $g(0) = 1$ and $g(\infty) = 0$. Note that if G is a distribution function for which

$G(0)=0$, then $g(x)=1-G(x-)$ is in $\mathscr{G}$. Conversely, for $g \in \mathscr{G}$, $G(x)=1-g(x+)$ is a distribution function with $G(0)=0$. Recall that, for random variables, X_1, X_2 with distribution functions F_1, F_2, $X_1 \ll_{\text{st}} X_2$ iff $F_1(x) \geqq F_2(x)$ for all real x. Clearly, for positive random variables, $X_1 \ll_{\text{st}} X_2$ iff

$$\mathscr{E}g(X_1) \geqq \mathscr{E}g(X_2) \qquad \text{for all} \quad g \in \mathscr{G}. \tag{2.1}$$

A positve random variable Y with distribution functon F is *unimodal about* 0 iff F is a concave function on $(0, \infty)$. A fundamental result due to Khintchine (1938) implies that Y is unimodal about 0 iff Y has the same distribution as U/X where V and X are independent, U has a uniform distribution on $[0, 1]$, and X is some positive random variable. The reason for writing U/X rather than UX is to make our results correspond more easily to Lawtọn's. Now, let $\mathscr{F}$ be the class of functions $F\colon [0, \infty) \to [0, 1]$ which are continuous, concave, nondecreasing and satisfy $F(0)=0$ and $F(\infty)=1$. The remarks above imply that the correspondence between $\mathscr{F}$ and $\mathscr{G}$ defined by

$$F(x) = \int_0^1 g\left(\frac{u}{x}\right) du \tag{2.2}$$

is one-to-one and onto. For if G is the distribution function of X, then

$$\begin{aligned} F(x) &= P\{Y \leqq x\} = P\{U/X \leqq x\} \\ &= \int_0^1 P\{X \geqq u/x\}\, du = \int_0^1 \left(1 - G\left(\frac{u}{x}-\right)\right) du \\ &= \int_0^1 g(u/x)\, du \equiv \mathscr{E}g(U/x)\,. \end{aligned} \tag{2.3}$$

In order to motivate our definition of better denominators, we first consider a result of Lawton (1968). His Theorem 4 and the monotone convergence theorem show that "Z_1 is a better denominator than Z_2" is equivalent to

$$\mathscr{E}F(Z_1) \geqq \mathscr{E}F(Z_2) \qquad \text{for all} \quad F \in \mathscr{F}. \tag{2.4}$$

However, (2.3) implies that (2.4) is equivalent to

$$\mathscr{E}g(U/Z_1) \geqq \mathscr{E}g(U/Z_2) \qquad \text{for all} \quad g \in \mathscr{G}. \tag{2.5}$$

Then, from (2.1), (2.5) is equivalent to

$$U/Z_1 \ll_{\text{st}} U/Z_2\,. \tag{2.6}$$

Of course, U, Z_1 and Z_2 are independent. From Khintchine's (1938) Theorem, it follows that (2.6) is equivalent to

$$Y/Z_1 \ll_{\text{st}} Y/Z_2 \tag{2.7}$$

for all positive unimodal random variables Y where Y, Z_1 and Z_2 are independent.

In this paper, we will use (2.7) as a definition of Z_1 being a better denominator than Z_2. In fact, it is (2.7) which leads us to a definition of S_1 being a better denominator than S_2, where S_1 and S_2 are random positive definite symmetric matrices. Before proceeding to the definition, we need to introduce a notion of unimodality for random vectors.

Following Olshen and Savage (1970), we have

DEFINITION 2.1. A random vector Y taking values in an n-dimensional real vector space is *α-unimodal about* 0 iff the function $t \to t^{\alpha}\mathscr{E}f(tY)$ is non-decreasing in t for $t > 0$ for every bounded, nonnegative, measurable function f.

A result concerning α-unimodal random vectors, which generalizes Khintchine's Theorem, is

THEOREM 2.1 (Olshen and Savage (1970)). *$Y \in R^n$ is α-unimodal* iff $Y \sim U^{1/\alpha}X$ *where U and X are independent, U is uniform on* [0, 1] *and X is some random vector in R^n.*

Now, let $\mathscr{P}_n$ be the set of $n \times n$ positive definite symmetric matrices.

DEFINITION 2.2. Let S_1, S_2 taking values in $\mathscr{P}_n$ be random matrices. S_1 is a better α-denominator than S_2 (written $S_1 \prec_{(\alpha)} S_2$) iff

$$Y'S_1^{-1}Y \ll_{\mathrm{st}} Y'S_2^{-1}Y \tag{2.8}$$

for all α-unimodal random vectors Y which are independent of S_1 and S_2.

It follows from Definition 2.1 that if the random vector Y is α-unimodal then Y is β-unimodal for all $\beta \geqq \alpha$, so $S_1 \prec_{(\alpha)} S_2$ implies $S_1 \prec_{(\beta)} S_2$ for all $\beta \leqq \alpha$. Examples abound in which $\alpha_2 > \alpha_1$ and $S_1 \prec_{(\alpha_1)} S_2$ but not $S_1 \prec_{(\alpha_2)} S_2$. In the special case $n = 1$ Theorem 2.1 dictates that $S_1 \prec_{(\alpha)} S_2$ iff $S_1^{\alpha/\beta} \prec_{(\beta)} S_2^{\alpha/\beta}$; the cited definition and theorem also guarantee, when $n = 1$, that $S_1 \prec_{(\alpha)} S_2$ and $\beta > 0$ together imply $S_1^{\beta} \prec_{(\delta)} S_2^{\beta}$ for all $\delta \leqq \alpha/\beta$.

PROPOSITION 2.1. *These are equivalent:*

(i) $S_1 \prec_{(\alpha)} S_2$;

(ii) $U(x'S_1^{-1}x)^{\alpha/2} \ll_{\mathrm{st}} U(x'S_2^{-1}x)^{\alpha/2}$ *for all $x \in R^n$, where U is uniform on* [0, 1] *and independent of S_1 and S_2.*

PROOF. The equivalence of (i) and (ii) follows from Theorem 2.1 because $Y'S_i^{-1}Y \sim U^{2/\alpha}X'S_i^{-1}X$ for each α-unimodal Y. □

What follows is the multivariate analogue of parts (i) and (ii) of Lawton's (1968) Theorem 4.

THEOREM 2.2. *The two conditions of Propositon* 2.1 *are equivalent to*

(iii) $\mathscr{E}F((x'S_1^{-1}x)^{-\alpha/2}) \geqq \mathscr{E}F((x'S_2^{-1}x)^{-\alpha/2})$ *for all $F \in \mathscr{F}$, and $x \in R^n$, $x \neq 0$.*

PROOF. $S_1 \prec_{(\alpha)} S_2$ iff for each $g \in \mathscr{G}$,

$$\mathscr{E}g(U(x'S_1^{-1}x)^{\alpha/2}) \geqq \mathscr{E}g(U(x'S_2^{-1}x)^{\alpha/2})\,. \tag{2.9}$$

From (2.3), each $F \in \mathscr{F}$ has the form

$$F(t) \equiv \mathscr{E}g(U/t)\,, \qquad t > 0\,, \tag{2.10}$$

so (2.9) is equivalent to (iii). □

Theorem 4 of Lawton (1968) includes a criterion equivalent to (i), (ii), and (iii) of Proposition 2.1 and Theorem 2.2 in case $n = 1$ and $\alpha = 2$. Namely, $S_1 \prec_{(2)} S_2$ iff there is a probability space on which are defined two random variables Y_1, Y_2, with $Y_i \sim S_i$, and $E(Y_2 \mid Y_1) \leqq Y_1$. The equivalence of this latter condition and (ii) of Lawton's Theorem 4 was proved first by Strassen ((1965), page 435). Strassen proved also a multivariate version of this equivalence. We have not found such results helpful in our work, though in view of what is known at least one such is implicit in Theorem 2.2. For example, $S_1 \prec_{(2)} S_2$ iff for each $x \in R^n$ there is a probability space on which are defined (univariate) random variables Z_1, Z_2, with $Z_i \sim (x'S_i^{-1}x)^{-1}$, and $E(Z_2 \mid Z_1) \leqq Z_1$.

For each $x \in R^n$, $x \neq 0$, define f_x on $\mathscr{P}_n$ to $(0, \infty)$ by

$$f_x(M) = (x'M^{-1}x)^{-1}\,, \qquad M \in \mathscr{P}_n\,. \tag{2.11}$$

The next theorem allows us to establish Lawton's Lemma 1 for random $S \in \mathscr{P}_n$. Note that (iii) of Theorem 2.2 can be written in the form

$$\mathscr{E}F(f_x^{\alpha/2}(S_1)) \geqq \mathscr{E}F(f_x^{\alpha/2}(S_2))\,. \tag{2.12}$$

THEOREM 2.3. *For each $F \in \mathscr{F}$ and $x \neq 0$, the function $F \circ f_x : \mathscr{P}_n \to [0, 1]$ is a concave function on the convex set $\mathscr{P}_n$.*

PROOF. It suffices to establish the theorem for F having a continuous second derivative. For M_1, $M_2 \in \mathscr{P}_n$ and $\lambda \in [0,1]$, let

$$h(\lambda) = (F \circ f_x)(\lambda M_1 + (1-\lambda)M_2)\,. \tag{2.13}$$

It is sufficient to show $h(\lambda)$ is concave in λ. Write $M_1 = WW'$ and $M_2 = WD_\theta W'$ where W is a nonsingular $n \times n$ matrix and D_θ is a diagonal matrix with ith diagonal element $d_{ii} = \theta_i > 0$ (see Rao (1965), page 37). Put $u = W^{-1}x$; so

$$h(\lambda) = F\left[\frac{1}{\sum_i u_i^2(\lambda + (1-\lambda)\theta_i)^{-1}}\right] \equiv F(k^{-1}(\lambda)) \tag{2.14}$$

where $k(\lambda) = \sum_i u_i^2(\lambda + (1-\lambda)\theta_i)^{-1}$. Straightforward differentiation show that

$$h''(\lambda) = F''(k^{-1}(\lambda))\left[-\frac{k'(\lambda)}{k^2(\lambda)}\right]^2 - F'(k^{-1}(\lambda))k^{-3}(\lambda)[k''(\lambda) - 2(k'(\lambda))^2]\,. \tag{2.15}$$

Since F is concave and non-decreasing, to show $h''(\lambda) \leqq 0$, it suffices to show that

$$k''(\lambda)k(\lambda) - 2(k'(\lambda))^2 \geqq 0\,. \tag{2.16}$$

However,

$$k'(\lambda) = \textstyle\sum_i u_i^2(\lambda + (1-\lambda)\theta_i)^{-2}(1-\theta_i) \tag{2.17}$$

and

$$k''(\lambda) = 2\textstyle\sum_i u_i^2(\lambda + (1-\lambda)\theta_i)^{-3}(1-\theta_i)^2\,. \tag{2.18}$$

Now, if $a_i = u_i(\lambda + (1-\lambda)\theta_i)^{-\frac{3}{2}}(1-\theta_i)$ and $b_i = u_i(\lambda + (1-\lambda)\theta_i)^{-\frac{1}{2}}$, (2.16) follows from the Cauchy–Schwarz inequality: $(\sum a_i b_i)^2 \leqq (\sum a_i^2)(\sum b_i^2)$. □

COROLLARY 2.1. *If* $0 < \alpha \leqq 2$, *then* $F(f_x^{\alpha/2}(M))$ *is a concave functon on* $\mathscr{P}_n$.

PROOF. First note that for $0 < \alpha \leqq 2$ and $F \in \mathscr{F}$, the function $H(u) = F(u^{\alpha/2})$ is also in $\mathscr{F}$. Theorem 2.3 applied to H gives the conclusion. □

In what follows all random matrices which appear are defined on the same probability space. The next results involve two concepts whose definitions we recall. A sequence $(S_1, S_2, \cdots, S_m)$ of random matrices is *exchangeable* iff for each permutation π of the integers $1, \cdots, m$, $(S_1, S_2, \cdots, S_m) \sim (S_{\pi(1)}, S_{\pi(2)}, \cdots, S_{\pi(m)})$. Also, the vector $\boldsymbol{\phi} = (\phi_1, \cdots, \phi_m)$ *majorizes* the vector $\boldsymbol{\eta} = (\eta_1, \cdots, \eta_m)$ if, possibly after reordering their components, $\phi_1 \geqq \phi_2 \geqq \cdots \geqq \phi_m$, $\eta_1 \geqq \eta_2 \geqq \cdots \geqq \eta_m$; $\sum_{i=1}^k \phi_i \geqq \sum_{i=1}^k \eta_i$, $k = 1, 2, \cdots, m-1$; $\sum_{i=1}^m \phi_i = \sum_{i=1}^m \eta_i$. An important characterization is this: ϕ majorizes $\boldsymbol{\eta}$ iff $\boldsymbol{\eta} = \mathbf{P}\boldsymbol{\phi}$ for some doubly stochastic matrix $\mathbf{P}$ (see Hardy, Littlewood, and Pólya (1964), pages 45 and 49).

THEOREM 2.4. *Let* $(X_1, \cdots, X_m)$ *be an exchangeable sequece of real random vectors with values in* R^k. *Assume the function* $\phi: (R^k)^m \to R$ *satisfies* $\phi(x_1, \cdots, x_m) = \phi(x_{\pi(1)}, \cdots, x_{\pi(m)})$, *and further assume* ϕ *is continuous and convex. Then*

$$\mathscr{E}\phi(\phi_1 X_1, \phi_2 X_2, \cdots, \phi_n X_m) \geqq \mathscr{E}\phi(\eta_1 X_1, \eta_2 X_2, \cdots, \eta_m X_m)$$

whenever $\boldsymbol{\phi} = (\phi_1, \cdots, \phi_m)$ *majorizes* $\boldsymbol{\eta} = (\eta_1, \cdots, \eta_m)$.

Marshall and Proshan ((1965), pages 87 and 88) have proved Theorem 2.4 in the case $k = 1$. Their proof applies here without change. In the terminology of Berge ((1963), page 219), the function $(a_1, \cdots, a_m) \to \mathscr{E}\phi(a_1 X_1, \cdots, a_n X_n)$ is S-convex (sometimes called Schur convex). The interested reader is encouraged to pursue the eleventh section of Chapter VIII of Berge's book. In the next theorem and its corollary, Theorem 2.4 is applied together with our previous results to extend the scope of Lawton's conclusions.

THEOREM 2.5. *Let* $S_1, \cdots, S_m$ *be random matrices (considered as vectors) in* $\mathscr{P}_n$ *which satisfy the assumptions of Theorem* 2.4, *where* $k = n(n+1)/2$. *If* $0 < \alpha \leqq 2$ *then*

$$\sum_{i=1}^m \eta_i S_i \prec_{(\alpha)} \sum_{i=1}^m \phi_i S_i \tag{2.19}$$

provided $\boldsymbol{\phi}$ *majorizes* $\boldsymbol{\eta}$.

PROOF. Let $H: \mathscr{P}_n \to (0, \infty)$ by $H(S) = F(f_x^{\alpha/2}(S))$, where $F(f_x^{\alpha/2}(\cdot))$ is as in Corollary 2.1. In view of that corollary H is convex. Define $\phi: (\mathscr{P}_n)^m \to (0, \infty)$ by $\phi(S_1, \cdots, S_m) = -H(\sum_{i=1}^m S_i)$; ϕ satisfies the assumptions of Theorem 2.4. Therefore, Theorems 2.4 and 2.2 imply (2.19). □

COROLLARY 2.2. *Let* $S_1, \cdots, S_m$ *be independent and identically distributed random matrices in* $\mathscr{P}_n$. *If* $0 < \alpha \leqq 2$, *then*

$$\frac{1}{m} \sum_{i=1}^m S_i \prec_{(\alpha)} \sum_{i=1}^m \lambda_i S_i \prec_{(\alpha)} \frac{1}{\nu} \sum_{i=1}^{\nu} S_i\,, \tag{2.20}$$

where $0 < \lambda_i \leqq 1$, $\sum \lambda_i = 1$, *and* ν *is the largest integer not more than* $\min\{1/\lambda_i\}$.

PROOF. Independent random vectors are automatically exchangeable. It is trivial that the probability vector whose every entry is $1/m$ is majorized by $\boldsymbol{\lambda} = (\lambda_1, \cdots, \lambda_m)$. The arguments of the second section of Hájek's paper (1962) show that the probability vector whose first ν entries are $1/\nu$ majorizes $\boldsymbol{\lambda}$ and ν is the largest j for which a probability vector with entries $1/j$ and 0 majorizes $\boldsymbol{\lambda}$. □

As has been mentioned, Hájek's results and Lawton's extensions of them involve majorization, but only implicitly. Also, their arguments apply verbatim to exchangeable (not merely independent and identically distributed) variables.

3. Applications and discussion. Statistical applications of Corollary 2.2 in one dimension for $\alpha = 1$ and $\alpha = 2$ are given in the papers of Lawton (1965; 1968), and are not repeated here. A multivariate version of Hsu's result would be relevant to the p-dimensional Behrens–Fisher problem in case the population covariance matrices are proportional. To see this, alter the assumptions of Section 1 as follows: let $X_{j,i}$ be p-dimensional normal with respective mean vectors μ_i, $i = 1, 2$ and covariance matrices Σ and $k\Sigma$ where $\Sigma \in \mathscr{P}_p$ and k is an unknown constant. In an obvious notation, $S_i^2 = (n_i - 1)^{-1} \sum_j (X_{j,i} - \bar{X}_i)(X_{j,i} - \bar{X}_i)'$ are the sample covariance matrices and the analogue of T_1^2 is

$$T_p^2 = (\bar{X}_1 - \bar{X}_2)'(n_1^{-1}S_1^2 + n_2^{-1}S_2^2)^{-1}(\bar{X}_1 - \bar{X}_2) . \tag{3.1}$$

Under the hypothesis that $\mu_1 = \mu_2$, $\bar{X}_1 - \bar{X}_2 \sim N_p(0, (n_1^{-1} + kn_2^{-1})\Sigma)$, $S_1^2 \sim W(\Sigma, p, n_1 - 1)$ and $S^2 \sim W(k\Sigma, p, n_1 - 1)$ where $W(\Sigma, p, m)$ denotes a Wishart distribution on $\mathscr{P}_p$ with m degrees of freedom and expectation $m\Sigma$. In this notation, we allow $m < p$. The distribution of T_p^2 does not depend on Σ when $\mu_1 = \mu_2$. If we put $\Sigma = I$ and let $V = (n_1^{-1} + kn_2^{-1})^{-\frac{1}{2}}(\bar{X}_1 - \bar{X}_2)$, then when $\mu_1 = \mu_2$,

$$T_p^2 \sim V'[\lambda(n_1 - 1)^{-1}S_{n_1-1} + (1 - \lambda)(n_2 - 1)^{-1}S_{n_2-1}]^{-1}V \tag{3.2}$$

where $V \sim N_p(0, I)$ and $S_m \sim W(I, p, m)$ for $m = n_1 - 1$ and $m = n_2 - 1$. Of course, $\lambda = n_1^{-1}(n_1^{-1} + kn_2^{-1})^{-1}$. Note that $\lambda(n_1 - 1)^{-1}S_{n_1-1} + (1 - \lambda)(n_2 - 1)^{-1}S_{n_2-1}$ has the same distribution as a convex combination of $(n_1 + n_2 - 2)$ independent $W(I, p, 1)$ random matrices. Our interest in the present problem arose from attempting to generalize (1.3) to the multivariate case. The following theorem accomplishes this for two dimensions ($p = 2$).

THEOREM 3.1. *Suppose $Z \sim N_2(0, I)$, $W_1, \cdots, W_k$ are independent and independent of z and that $W_i \sim W(I, 2, 1)$. Let $\lambda_1, \cdots, \lambda_k$ be nonnegative, $\sum_i \lambda_i = 1$ and assume at least two λ_i are positive. Then for any integer, ν, $2 \leqq \nu \leqq \min\{1/\lambda_i\}$,*

$$Z'\left[\frac{S_k}{k}\right]^{-1} Z \ll_{\mathrm{st}} Z'(\textstyle\sum_1^k \lambda_i W_i)^{-1}Z \ll_{\mathrm{st}} Z'\left[\frac{S_\nu}{\nu}\right]^{-1} Z \tag{3.3}$$

where $S_m = \sum_{i=1}^m W_i$, $m = 1, 2, \cdots, k$.

PROOF. Z is 2-unimodal according to Corollary 3 of Olshen and Savage (1970). But, Corollary 2.2 does not apply directly as each W_i is a.s. singular. Thus, for each $\varepsilon > 0$, let $W_i^{(\varepsilon)} = W_i + \varepsilon I$ where I is the 2×2 identity matrix. From Corollary 2.2, (3.3) holds for each $\varepsilon > 0$ with W_i replaced by $W_i^{(\varepsilon)}$, and $S_m^{(\varepsilon)} = \sum_{i=1}^m W_i^{(\varepsilon)}$. Noting that the map $P \to P^{-1}$ is continuous on $\mathscr{P}_2$ and letting $\varepsilon \to 0$, we have $(S_m^{(\varepsilon)})^{-1} \to_{\text{a.s.}} S_m^{-1}$, $m \geqq 2$, and $(\sum_{i=1}^k \lambda_i W_i^{(\varepsilon)})^{-1} \to_{\text{a.s.}} (\sum_{i=1}^k \lambda_i W_i)^{-1}$. The last assertion follows from the assumption that at least two λ_i are nonnegative. Since a.s. convergence implies convergence in distribution, (3.3) follows. □

The restriction in Theorem 2.4 that $\alpha \leqq 2$ is essentially a restriction on dimension because any probability density on R^n which is bounded at 0 cannot be the probability density of an α-unimodal random vector for $\alpha < n$ (see Corollary 3 of Olshen and Savage (1970)). Of course, we had hoped to decide whether or not Theorem 3.1 is true for $p > 2$. We once thought Theorem 3.1 was false for $p > 2$, but are now unsure. A special instance of this question concerns Hotelling's T^2 with non-standard normalizing constants. Namely, is $k\chi_p^2/\chi_{k-p+1}^2$ stochastically decreasing in k? (See Rao (1962) page 458.) An affirmative answer for $p = 1$ follows from Hsu (1939) and for $p = 2$ from Theorem 3.1. From Pearson's (1934) *Tables of the Incomplete Beta Function*, the result appears to be true for all p. One might conjecture (what is almost the same) that $F_{p,q}$ is stochastically decreasing in q for each p. The case $p = 1$ has been discussed and the case $p = 2$ follows from Lawton (1968) or our Corollary 2.2. However, this result is false for $p > 2$ as can be seen by a cursory examination of standard F-tables.

According to Lawton, the pair (X, h) satisfies Condition (A) if (i) the one-dimensional variable X has a continuous distribution, and (ii) for any $a < 0 < b$ the function $f_{a,b}(z) = P\{a < X/h(z) < b\}$ is concave for $z \geqq 0$. Here, h is a continuous real-valued function. In view of the similarity of Lawton's results and those of the present paper, and the arguments of page 27 of Olshen and Savage (1970), it seems plausible that if (X, h) satisfies Condition (A), then X is $\alpha = \alpha(h)$ unimodal for some $\alpha > 0$. If $h(x) = x^{1/\alpha}$, $\alpha > 0$, then a variant of Theorem 5 of Olshen and Savage (page 31) shows that X is α-unimodal. Theorem 3 of Olshen and Savage (page 30) shows that conversely, if X is α-unimodal, then $(X, x \to x^{1/\alpha})$ satisfies Condition (A). If $X \sim N(\delta, 1)$ and $h(x) = x^{\frac{1}{2}}$, then (X, h) satisfies Condition (A) iff $|\delta| \leqq 2$ (see Lawton (1965), Corollary 1). In this case, Corollary 3 of Olshen and Savage establishes that $X \sim N(\delta, 1)$ is 2-unimodal iff $|\delta| \leqq 2$. However, the general question of Condition (A) implying α-unimodality remains open.

REFERENCES

[1] Berge, C. (1963). *Topological Spaces*. Oliver and Boyd, Edinburgh.

[2] Hájek, J. (1962). Inequalities for the generalized Student's distributions. *Sel. Transl. Math. Statist. Prob.* **2** 63–74.

[3] Hardy, G. H., Littlewood, J. E., and Pólya, G. (1964) *Inequalities*, (2nd ed.). Reprinted. Cambridge Univ. Press.

[4] Hsu, P. L. (1938). Statistical Research Memoirs. Department of Statistics, University College, London.

[5] KARLIN, S. (1968). *Total Positivity*. Stanford Univ. Press.

[6] KHINTCHINE, A. Y. (1938). On unimodal distributions (in Russian). *Izv. Nauchno-Issled. Inst. Math. Mekh. Tomsk. Gos. Univ.* **2** 1–7.

[7] LAWTON, W. H. (1965). Some inequalities for central and non-central distributions. *Ann. Math. Statist.* **36** 1521–1525.

[8] LAWTON, W. H. (1968). Concentration of random quotients. *Ann. Math. Statist.* **39** 466–480.

[9] MARSHALL, A. W. and PROSCHAN, F. (1965). An inequality for convex functions involving majorization. *J. Math. Anal. Appl.* **12** 87–90.

[10] MICKEY, M. R. and BROWN, M. B. (1966). Bounds on the distribution functions of the Behrens-Fisher statistic. *Ann. Math. Statist.* **37** 639–642.

[11] OLSHEN, R. A. and SAVAGE, L. J. (1970). A generalized unimodality. *J. Appl. Prob.* **7** 21–34.

[12] PEARSON, K. (1934). *Tables of the Incomplete Beta Function*. Cambridge Univ. Press.

[13] RAO, C. R. (1965). *Linear Statistical Inference and Its Applications*. Wiley, New York.

[14] SCHEFFÉ, H. (1970). Practical solutions of the Behrens-Fisher problem. *J. Amer. Statist. Assn.* **65** 1501–1508.

[15] STRASSEN, V. (1965) The existence of probability measures with given marginals. *Ann. Math. Statist.* **36** 423–439.

DEPARTMENT OF STATISTICS
UNIVERSITY OF MICHIGAN
1447 MASON HALL
ANN ARBOR, MICHIGAN 48104