MONOTONICITY PROPERTIES OF DIRICHLET INTEGRALS
WITH APPLICATIONS TO THE MULTINOMIAL DISTRIBUTION
AND THE ANOVA TEST

BY
INGRAM OLKIN

TECHNICAL REPORT NO. 60
AUGUST 23, 1971

PREPARED UNDER THE AUSPICES
OF
NATIONAL SCIENCE FOUNDATION GRANT GP-27210
Ingram Olkin, Project Director

DEPARTMENT OF STATISTICS
STANFORD UNIVERSITY
STANFORD, CALIFORNIA
MONOTONICITY PROPERTIES OF DIRICHLET INTEGRALS WITH APPLICATIONS TO THE MULTINOMIAL DISTRIBUTION AND THE ANOVA TEST

by

Ingram Olkin

TECHNICAL REPORT NO. 60
August 23, 1971

PREPARED UNDER THE AUSPICES OF
NATIONAL SCIENCE FOUNDATION GRANT 27210

Ingram Olkin, Project Director

This report is also appearing as Technical Report No. 139 under Contract N00014-67-A-0112-0052

DEPARTMENT OF STATISTICS
STANFORD UNIVERSITY
STANFORD, CALIFORNIA
MONOTONICITY PROPERTIES OF DIRICHLET INTEGRALS WITH APPLICATIONS
TO THE MULTINOMIAL DISTRIBUTION AND THE ANOVA TEST

Ingram Olkin
Stanford University

1. Introduction. The present paper is concerned with establishing bounds for the Dirichlet integrals:

\begin{align}
L_1(a) &= \int_{0}^{a_k} \ldots \int_{0}^{a_k} \frac{k!}{\Pi_{i=1}^{k} t_i^{l_i}} (1-t_i)^{d-1} d \Pi_{i=1}^{k} t_i,
\end{align}

\begin{align}
L_2(a) &= \int_{0}^{a_k} \ldots \int_{0}^{a_k} \frac{k!}{\Pi_{i=1}^{k} t_i^{l_i}} (1-t_i)^{d-1} d \Pi_{i=1}^{k} t_i,
\end{align}

\begin{align}
U(a) &= \int_{a_1}^{a_k} \ldots \int_{a_1}^{a_k} \frac{k!}{\Pi_{i=1}^{k} t_i^{l_i}} (1+t_i)^{d-1} d \Pi_{i=1}^{k} t_i,
\end{align}

where \( a = (a_1, \ldots, a_k) \), and \( d \) is such that the integrals exist.

The main result is that under certain conditions, \(-L_1(a), -L_2(a),\) and \(-U(a)\) are Schur functions in \((a_1, \ldots, a_k)\), (see e.g., Marshall, Olkin and Proschan (1967)). As a consequence of this fact it follows that if

\begin{align}
(a_1, \ldots, a_k) > (b_1, \ldots, b_k)
\end{align}
in the sense that, after possible reordering, with \( a_1 \geq \ldots \geq a_k \),
\( b_1 \geq \ldots \geq b_k \), that

\[
(1.4) \quad \sum_{l=1}^{m} a_{l} \geq \sum_{l=1}^{m} b_{l}, \quad m=1, \ldots, k-1, \quad \sum_{l=1}^{k} a_{l} = \sum_{l=1}^{k} b_{l}
\]

holds, then

\[
(1.5) \quad L_1(a) \leq L_1(b), \quad L_2(a) \leq L_2(b), \quad U(a) \leq U(b)
\]

In this way we can generate many inequalities. Both the multinomial
distribution and the simultaneous ANOVA test are related to \( L_2(a) \)
and \( U(a) \), respectively, and we provide several new results as a
consequence of (1.4).

2. The main results. We next prove the main results concerning
\( L_1(a), L_2(a) \) and \( U(a) \). Theorem 1 is stated more generally and
encompasses both \( L_1(a) \) and \( L_2(a) \).

To show that a function \( F(a) = F(a_1, \ldots, a_k) \) is a Schur function,
we must show that

\[
(2.1) \quad \left( \frac{\partial F(a)}{\partial a_i} - \frac{\partial F(a)}{\partial a_j} \right) (a_i - a_j) \geq 0,
\]

for all \( i \) and \( j \).
Theorem 1. If \( f(x) \) is a non-negative monotone decreasing function,

\[
L(a;w) = \int_0^{a_1} \cdots \int_0^{a_k} f(\Sigma t_i) \prod_{i=1}^k w_i^{1-1} dt_i ,
\]

where \( a_1 \geq \cdots \geq a_k \geq 0, \ 0 \leq w_1 \leq \cdots \leq w_k \), then \(-L(a;w)\) is a Schur function in \((a_1, \ldots, a_k)\), whenever the integral is finite.

Proof. Because of symmetry, we need only consider \((i,j) = (1,2)\) in showing that \((2.1)\) holds for \(-L(a;w)\). It is immediate from \((2.2)\) -- noting the condition on \( f(x) \) -- that

\[
\frac{\partial L(a;w)}{\partial a_1} = \int_0^{a_2} \cdots \int_0^{a_k} f(a_1 + \Sigma t_i) \prod_{i=1}^k w_i^{1-1} dt_i
\]

\[
= \prod_{i=1}^k \int_0^{a_2} \cdots \int_0^{a_k} a_1^{-1} f(a_1 + a_2 z + \Sigma a_i y_i) \prod_{i=1}^k w_i^{1-1} dz \prod_{i=1}^k dy_i .
\]

Condition \((2.1)\) for \(-L(a;w)\) will be satisfied if

\[
\int_0^{a_2} \cdots \int_0^{a_k} [a_1^{-1} f(a_1 + a_2 z + \Sigma a_i y_i) z^{1-1} - a_2^{-1} f(a_1 + a_2 x + \Sigma a_i y_i) w_i^{1-1}] dz \prod_{i=1}^k dy_i \leq 0 .
\]

A sufficient condition for \((2.4)\) to hold is that, pointwise,

\[
a_2 f(a_1 + a_2 z + q) z^{1-1} \leq a_1 f(a_1 + a_2 z + q) w_i^{1-1} .
\]
For $0 \leq z \leq 1$, $(a_1 + a_2 z + Q) \geq (a_1 z + a_2 + Q)$, so that $f(a_1 + a_2 z + Q) \leq f(a_1 z + a_2 + Q)$. Since $0 \leq w_1 \leq w_2$, $z w_2 \leq z w_1$, and (2.4) holds, the result for $U(a)$ is more delicate and a pointwise argument does not carry through.

Theorem 2. If $a_1 \geq \ldots \geq a_k \geq 0$, $0 \leq w_1 \leq \ldots \leq w_k$, and

$$U(a; w) = \int_{a_1}^{\infty} \ldots \int_{a_k}^{\infty} \frac{k w_{i-1}^l}{(1 + \sum t_i)^d} \, dt_1 \ldots dt_k,$$

(2.6)

where $d \geq d w_1$, then $-U(a; w)$ is a Schur function.

Proof. Because of symmetry, we need only consider $(i, j) = (1, 2)$ in showing that (2.1) holds for $-U(a; w)$. It is immediate that

$$\frac{\partial U(a; w)}{\partial a_1} = - \int_{a_2}^{\infty} \ldots \int_{a_k}^{\infty} \frac{k w_{i-1}^l a_1}{2 (1 + a_1 + t_2 + \ldots + t_k)^d} \, dt_1 \ldots dt_k,$$

(2.7)

Let $t_2 - a_2 = z$ and interchange order of integration; then (2.7) becomes

$$\frac{\partial U(a; w)}{\partial a_1} = - \int_{a_2}^{\infty} \ldots \int_{a_k}^{\infty} \frac{k w_{i-1}^l a_1}{2 (1 + a_1 + a_2 + z + t_3 + \ldots + t_k)^d} \, dz \int_{0}^{\infty} \frac{w_1 w_{i-1}}{1 + a_1 + a_2 + z + t_3 + \ldots + t_k} \, dz,$$

(2.8)
We now use a pointwise argument on the inner integral with $\sum_{i=1}^{k} \frac{a_i}{2} \frac{x_i}{\frac{a_i}{2}}$ fixed. Let $z = (1 + a_1 + a_2 + \sum_{i=1}^{k} x_i) \nu = sv$ then the inner integral becomes

$$
(2.9) \quad \int_{0}^{s} \frac{w_{1}^{l-1}(sv+a_{2})^{w_{2}^{l-1}}} {s^{d-1}(1+v)^{d}} \, dv.
$$

Consequently, if $a_1 \geq a_2$, $\left(\frac{\partial U(a;w)} {\partial a_1} - \frac{\partial U(a;w)} {\partial a_2}\right) \leq 0$ provided

$$
(2.10) \quad \int_{0}^{w_{1}} \frac{w_{1}^{l-1}(sv+a_{2})^{w_{2}^{l-1}} - a_1^{w_{1}^{l-1}(sv+a_{1})^{w_{1}^{l-1}}}} {(1+v)^{d}} \, dv \geq 0.
$$

The ordering $0 \leq w_1 \leq w_2$ guarantees that the integrand be non-negative, so that (2.10) holds.\|

3. An application to the multinomial distribution. Let $X = (X_1, \ldots, X_k)$ have the multinomial distribution

$$
(3.1) \quad P\{X=x\} = \binom{n}{x_1, \ldots, x_k} \prod_{i=1}^{k} \frac{x_i}{\theta_i},
$$

where $x = (x_1, \ldots, x_k)$, $\sum x_i = n$, $\theta_1 \geq \ldots \geq \theta_k \geq 0$, $\sum \theta_i = 1$, and consider the tail probability $P\{X_1 \geq r, \ldots, X_k \geq r | \theta_1, \ldots, \theta_k\}$, with $r \leq n/k$.

Alam (1970) obtains lower and upper bounds for $P\{X_1 \geq r, \ldots, X_k \geq r | \theta_1, \ldots, \theta_k\}$ by averaging some of the $\theta$'s, namely
(3.2) \( P(X_1 \geq r, \ldots, X_k \geq r \mid \theta^*, \theta_k, \ldots, \theta_k) \leq P(X_1 \geq r, \ldots, X_k \geq r \mid \theta_1, \ldots, \theta_k) \leq P(X_1 \geq r, \ldots, X_k \geq r \mid \bar{\theta}, \ldots, \bar{\theta}) \),

where \( \theta^* = 1-(k-1)\theta_k \) and \( \bar{\theta} = \Sigma \theta_i / k \). It has been shown by Olkin and Sobel (1965) that \( P(X_1 \geq r, \ldots, X_k \geq r \mid \theta_1, \ldots, \theta_k) \) has a representation in terms of the Dirichlet integral

\[
(3.3) \quad P(X_1 \geq r, \ldots, X_k \geq r \mid \theta_1, \ldots, \theta_k) = \kappa(n,r,k) \int_0^{\theta_1} \cdots \int_0^{\theta_k} \frac{k^k}{\Gamma(1) \cdots \Gamma(k)} r^r \frac{1}{1} \frac{1}{1} \cdots \frac{1}{1} d\theta_1 \cdots d\theta_k,
\]

where \( d = n - kr \geq 0 \), and \( \kappa(n,r,k) = \Gamma(n+1)/[[\Gamma(r)]^k \Gamma(n-kr+1)] \). We may now make use of Theorem 1 with \( w_1 = \ldots = w_k = r \). Thus, if \( 0 \leq p_i \leq 1 \), \( 0 \leq q_i \leq 1 \), and \( (p_1, \ldots, p_k) \succ (q_1, \ldots, q_k) \), then

\[
(3.4) \quad P(X_1 \geq r, \ldots, X_k \geq r \mid p_1, \ldots, p_k) \leq P(X_1 \geq r, \ldots, X_k \geq r \mid q_1, \ldots, q_k) \).
\]

The results of Alam are special cases of (3.4) since

\[
(\theta^*, \theta_k, \ldots, \theta_k) \succ (\theta_1, \ldots, \theta_k) \succ (\bar{\theta}, \ldots, \bar{\theta}) .
\]

Clearly, many other intermediate bounds can now be obtained.

For the lower tail of the multinomial distribution (3.1), we have the representation
\[(3.5) \quad c(\theta; r) = P(X_1 < r, \ldots, X_m < r \mid \theta_1, \ldots, \theta_k) / \kappa(n, r, m) \]

\[
= \int_{\theta_1}^{1-\Sigma \theta_1} \int_{\theta_2}^{1-\Sigma \theta_2} \cdots \int_{\theta_m}^{1-\Sigma \theta_m} (1 - \Sigma t_i) \prod_{i=1}^{m} t_i^{r-1} \, dt_1 \cdots dt_m,
\]

where \( m \leq \min(k-1, n) \) and \( d = n - kr \geq 0 \). As in Theorem 1, a direct differentiation of (3.5) with respect to \( \theta_j \), followed by the change of variables \( t_2 - \theta_2 = v, \quad t_j - \theta_j = z_j, \quad j = 3, \ldots, m \), yields

\[
\frac{\partial c(\theta; r)}{\partial \theta_1} = -\int_0^{\theta_0} \int_0^{\theta_0 - v} \cdots \int_0^{\theta_0 - v - \Sigma z_j} z_j^{m-1} [\theta_1(v + \theta_2) \prod_{j=3}^{m}(z_j + \theta_j)]^{r-1}(\theta_0 - v - \Sigma z_j)^d \, dv \prod_{j=3}^{m} dz_j,
\]

where \( \theta_0 = 1 - \Sigma \theta_i \). By symmetry,

\[
\frac{\partial c(\theta; r)}{\partial \theta_2} = -\int_0^{\theta_0} \int_0^{\theta_0 - v} \cdots \int_0^{\theta_0 - v - \Sigma z_j} z_j^{m-1} [\theta_2(v + \theta_1) \prod_{j=3}^{m}(z_j + \theta_j)]^{r-1}(\theta_0 - v - \Sigma z_j)^d \, dv \prod_{j=3}^{m} dz_j.
\]

That \(-c(\theta; r)\) is a Schur function follows from the fact that

\[(\theta_1 - \theta_2)[(\theta_1 + \theta_2)^{r-1} - (\theta_2 + \theta_1)^{r-1}] \geq 0 \quad \text{for all} \quad v.\]

As a consequence, we have the

**Corollary.** If \((p_1, \ldots, p_k) \succ (q_1, \ldots, q_k)\), then

\[P(X_1 < r, \ldots, X_m < r \mid q_1, \ldots, q_k) \geq P(X_1 < r, \ldots, X_m < r \mid p_1, \ldots, p_k) .\]
4. **An application to the simultaneous ANOVA model.** Suppose two hypotheses are tested using the same error variance for each test, so that we have

\[ F_1 = \frac{q_1/n_1}{q_0/n_0} , \quad F_2 = \frac{q_2/n_1}{q_0/n_0} , \]

where \( q_0, q_1, \) and \( q_2 \) are independently distributed as \( \chi^2 \) variates with \( n_0, n_1, \) and \( n_2 \) d.f. respectively. Kimball (1951) obtained the inequality

\[(4.1) \quad P\{F_1 \leq F_{1\alpha}, F_2 \leq F_{2\alpha}\} \geq P\{F_1 \leq F_{1\alpha}\} P\{F_2 \leq F_{2\alpha}\} ,\]

where \( F_{1\alpha} \) and \( F_{2\alpha} \) are the 100 \( \alpha \) percent points of the distributions of \( F_1 \) and \( F_2 \). This inequality is of interest in that it provides a bound for the probability of making no errors of the first kind. We may use Theorem 2 to obtain a bound for \( P\{F_1 \geq F_{1\alpha}, F_2 \geq F_{2\alpha}\} \).

Suppose that \( n_1 = n_2 = n \), then

\[(4.2) \quad P\{F_1 \geq F_{\alpha}, F_2 \geq F_{\alpha}\} = k \int_c^\infty \int_c^\infty \frac{x^{n-1} y^{n-1}}{(1 + x + y)^n} \, dx \, dy ,\]

where \( c = \frac{n \alpha}{n_0} \) and \( k \) is a normalizing constant. Since \((2c,0) \succ (c,c)\), we obtain from Theorem 2 that

\[(4.3) \quad P\{F_1 \geq F_{\alpha}, F_2 \geq F_{\alpha}\} \geq P\{F_1 \geq 2F_{\alpha}, F_2 \geq 0\} = P\{F_1 \geq 2F_{\alpha}\} .\]
Since

\[(4.4) \quad P(F_1 \leq F_{\alpha}, F_2 \leq F_{\alpha}) = P(F_1 \leq F_{\alpha}) + P(F_2 \leq F_{\alpha}) + P(F_1 \geq F_{\alpha}, F_2 \geq F_{\alpha}) - 1, \]

we obtain an alternative inequality to that of (4.1), namely,

\[(4.5) \quad P(F_1 \leq F_{\alpha}, F_2 \leq F_{\alpha}) \geq 2P(F_1 \leq F_{\alpha}) - P(F_1 \leq 2F_{\alpha}). \]

This is to be compared with

\[(4.6) \quad P(F_1 \leq F_{\alpha}, F_2 \leq F_{\alpha}) \geq \left[P(F_1 \leq F_{\alpha})\right]^2. \]

We wish the larger bound, so that we need to determine the sign of

\[(4.7) \quad \left[P(F_1 \leq F_{\alpha})\right]^2 - 2P(F_1 \leq F_{\alpha}) + 1 - P(F_1 \geq 2F_{\alpha}) = \left[P(F_1 \geq F_{\alpha})\right]^2 - P(F_1 \geq 2F_{\alpha}). \]

It turns out that the difference is not always of one sign. When \( n=2, \)

\[\left[P(F \geq c)\right]^2 = \left(1 + \frac{2c}{n^0}\right) ^{-n_0} < \left(1 + \frac{4c}{n^0}\right) ^{-n_0} = P(F \geq 2c), \]

so that (4.5) yields a better bound than (4.6). With \( n_0 \to \infty, \) the
difference becomes

\[(4.8) \quad \left[P(X_n^2 \geq c)\right]^2 - P(X_n^2 \geq 2c). \]
When \( n = 2m \), \( P(X^2_n \geq c) = \sum_{j=0}^{m-1} \frac{c^j}{j!} \), and a straightforward analysis shows that (4.8) is nonnegative.

For small values of \( c \) (less than \( c_0(n, n_0) \)), \( [P(F \geq c)]^2 \) is larger than \( P(F \geq 2c) \), whereas for \( c > c_0(n, n_0) \), \( P(F \geq 2c) \) is larger than \( [P(F \geq c)]^2 \), where \( c_0(n, n_0) \) depends on \( n \) and \( n_0 \). As either \( n \) or \( n_0 \) increase, the constant \( c_0(n, n_0) \) tends to increase. Since \( c \) will, in general, be of moderate size, it appears that (4.5) is the better bound in practice.

Acknowledgement. The author is indebted to M. Sobel and V.R.R. Uppuluri for their comments, and for pointing out an application which led to Theorem 2; and to R. Miller for his suggestions concerning Section 4.
REFERENCES


