ANDERSON'S THEOREM ON THE INTEGRAL OF A SYMMETRIC UNIMODAL
FUNCTION OVER A SYMMETRIC CONVEX SET,
AND ITS APPLICATIONS IN PROBABILITY AND STATISTICS

by

Michael D. Perlman

TECHNICAL REPORT No. 126
March 1988

Department of Statistics, GN-22
University of Washington
Seattle, Washington 98195 USA
Anderson's Theorem on the Integral of a Symmetric Unimodal Function over a Symmetric Convex Set, and Its Applications in Probability and Statistics

by

Michael D. Perlman *

1. Anderson's theorem and Sherman's extension.

In 1955, T. W. Anderson [20] established the following elegant moving-set inequality, which probably is the most widely cited result in multivariate statistical analysis.

**Theorem.** Let \( f \) be a nonnegative symmetric unimodal function on \( \mathbb{R}^n \) and let \( K \) be a symmetric convex subset of \( \mathbb{R}^n \). Then for every fixed \( y \in \mathbb{R}^n \), the integral

\[
\int_{K+y} f(x) \, dx = h_1(t)
\]

is a symmetric unimodal function of the real variable \( t \). In particular, \( h_1 \) achieves its maximum value at \( t = 0 \).

As will be noted below, this result has found a remarkable number of applications in multivariate theory, including the unbiasedness and power monotonicity of hypothesis tests, lower bounds for the probability of simultaneous confidence regions, and probability inequalities for elliptically contoured

*Department of Statistics, University of Washington, Seattle, Washington 98195. Preparation of this review was supported in part by National Science Foundation Grant No. DMS 86-03489.
According to Anderson [20], the convex set $K$ is symmetric (also called balanced) if $K = -K$, while the function $f$ is symmetric if $f(x) = f(-x)$ for every $x \in \mathbb{R}^n$ and is unimodal if $\{x \in \mathbb{R}^n | f(x) \geq c \}$ is convex for every real number $c$. If $f$ and $K$ are symmetric then the symmetry of $h_1$ is immediate: simply replace $x$ by $-x$ in (1) to see that $h_1(t) = h_1(-t)$.

Anderson established the unimodality of $h_1$ by first considering the special case where $f = I_C$ with $C$ another symmetric convex set. In this case the integral in (1) reduces to

$$\int_{K + ty} I_C(x) \, dx = \mu[C \cap (K + ty)] = h_2(t)$$

where $\mu$ denotes Lebesgue measure on $\mathbb{R}^n$. Anderson's elegant demonstration of the unimodality of $h_2$ is based on the following set-inclusion: for $0 \leq t_0 \leq t$ and $\alpha = (t_0 + t) / 2t$,

$$C \cap (K + t_0y) \supseteq \alpha[C \cap (K + ty)] + (1 - \alpha)[C \cap (K - ty)].$$

Since $\frac{1}{2} \leq \alpha \leq 1$, the right-hand side of (4) is a convex combination of the sets $C \cap (K + ty)$ and $C \cap (K - ty)$, which have identical Lebesgue measures by the symmetry of $C$ and $K$. The classical Brunn-Minkowski inequality then yields

$$\mu^{1/n} [C \cap (K + t_0y)] \geq \alpha \mu^{1/n} [C \cap (K + ty)] + (1 - \alpha)\mu^{1/n} [C \cap (K - ty)]$$

which is equivalent to the unimodality of $h_2$.

Because the integrals in (1) and (2) are linear in $f$ and $I_C$, respectively, the unimodality of $h_2$ implies the unimodality of $h_1$ whenever $f$ belongs to the class $C$ defined to be the closure (in an appropriate topology) of the convex cone consisting of all functions on $\mathbb{R}^n$ of the form $\Sigma a_i I_{C_i}$, where $\{C_i\}$ is a finite collection of symmetric convex subsets of $\mathbb{R}^n$ and each $a_i > 0$. It is easily verified that $f \in C$ whenever $f$ is symmetric and unimodal in Anderson's sense, hence the theorem is proved.

Sherman (1955) noted that the class $C$ is strictly larger than the class of all nonnegative symmetric unimodal functions, thereby extending the scope of Anderson's result to its natural domain. Additionally, Sherman treated the symmetry and unimodality aspects of Anderson's theorem separately, thereby further broadening its scope and also strengthening its conclusion. Following the outline of Anderson's argument, he applied the Brunn-Minkowski inequality to show that for any two convex (but not
necessarily symmetric) subsets $C, K$ of $\mathbb{R}^n$,

$$\mu[C \cap (K + y)] \equiv h_3(y)$$

is unimodal on $\mathbb{R}^n$ in Anderson’s sense. Separately, $h_3$ is clearly symmetric if $C$ and $K$ are symmetric (not necessarily convex). Thus, if $C$ and $K$ are both symmetric and convex, then $h_3$ is symmetric and unimodal, in particular $h_3 \in \mathcal{C}$. By writing

$$h_3(y) = \int_I C(x)I_K(x-y)dx$$

$$= \int_I C(x)I_K(y-x)dx$$

$$\equiv (I_C * I_K)(y)$$

and noting that the convolution $f * g$ is bilinear in $f$ and $g$, Sherman observed that

$$\left(\sum a_i I_{C_i}\right) * \left(\sum b_i I_{K_i}\right) \in \mathcal{C}$$

whenever $C_i$ and $K_i$ are symmetric convex sets and $a_i, b_i > 0$. By taking limits, Sherman concluded that $\mathcal{C}$ is closed under convolution, i.e.,

$$f, g \in \mathcal{C} \implies f * g \in \mathcal{C}.$$ 

Since $h(ty)$ is symmetric and unimodal in $t$ whenever $h \in \mathcal{C}$, (8) is a strengthening of Anderson’s theorem.

2. Extensions involving group invariance and log concavity.

These results of Anderson and Sherman subsequently have been extended in a number of interesting ways. Mudholkar (1966) generalized the notion of symmetry by considering a finite (or compact) subgroup $G$ of orthogonal (hence Lebesgue measure-preserving) transformations acting on $\mathbb{R}^n$. He replaced the symmetry assumptions on $f, C, K$ by the assumption that they are $G$-invariant, i.e.,

$f(gx) = f(x), C = gC, K = gK$ for every $g \in G$. When $G = \{ \pm I \}$ with $I$ the identity transformation, $G$-invariance reduces to the Anderson-Sherman symmetry assumption. It is readily seen that if $f$ and $g$ are $G$-invariant functions on $\mathbb{R}^n$, then their convolution $f * g$ also is $G$-invariant. Thus, if $\mathcal{C}_G$ is defined to be the closure of the convex cone consisting of all finite linear combinations $\sum a_i I_{C_i}$ with $C_i$ convex and $G$-invariant and $a_i > 0$, then the Anderson-Sherman argument shows that $\mathcal{C}_G$ is closed under convolution, i.e.,
Mudholkar (1966) defined the function \( f \) on \( \mathbb{R}^n \) to be \( G \)-monotone if \( f(x) \geq f(y) \) whenever \( x \) lies in the convex hull of the \( G \)-orbit of \( y \), i.e., whenever
\[
\alpha_g \sum_{g \in G} \alpha_g g y,
\]
where each \( \alpha_g \geq 0 \) and \( \sum \alpha_g = 1 \). (In particular, a \( G \)-monotone function is \( G \)-invariant.) If \( C \) is a convex \( G \)-invariant subset of \( \mathbb{R}^n \), then \( f \equiv l_C \) is \( G \)-monotone, hence so is every \( f \in C_G \). Mudholkar (1966) concluded from (8') that
\[
f, g \in C_G \quad \Rightarrow \quad f \ast g \quad \text{is \( G \)-monotone}.
\]
When \( G = \{ \pm 1 \} \), \( G \)-monotonicity of \( f \) is equivalent to Anderson's condition that \( f(ty) \) is symmetric and unimodal in \( t \) for every fixed \( y \). The import of Mudholkar's result is that by imposing a stronger symmetry assumption on \( f \) and \( g \) while maintaining the convexity (≡ unimodality) assumption, a stronger monotonicity property can be deduced for \( f \ast g \).

It is readily seen that the class of all \( G \)-monotone functions is strictly larger than \( C_G \): for any two convex \( G \)-invariant subsets \( C_1, C_2 \) of \( \mathbb{R}^n \), \( I_{C_1 \cup C_2} \) is \( G \)-monotone but not necessarily in \( C_G \). It is natural to ask whether the result (10) can be strengthened as follows:
\[
f, g \quad \text{\( G \)-monotone} \quad \Rightarrow \quad f \ast g \quad \text{is \( G \)-monotone}?
\]

The case where \( G = S_n \), the symmetric group represented as the group of all \( n \times n \) permutation matrices acting on \( \mathbb{R}^n \), has deserved and received special attention. In this case the condition (9) is equivalent to the condition that \( x = (x_1, \ldots, x_n) \) is majorized by \( y = (y_1, \ldots, y_n) \), i.e., that
\[
\sum_{j=1}^k x(j) \leq \sum_{j=1}^k y(j), \quad k = 1, \ldots, n-1,
\]
\[
\sum_{j=1}^n x(j) = \sum_{j=1}^n y(j),
\]
where \( x(1) \geq \cdots \geq x(n) \) and \( y(1) \geq \cdots \geq y(n) \) denote the ordered values of the components of \( x \) and \( y \), respectively. Here the class of \( S_n \)-monotone functions \( f \) is precisely the class of Schur-concave functions, i.e., \( f(x) \geq f(y) \) whenever \( x \) is majorized by \( y \) (cf. Marshall and Olkin (1979)). Marshall and Olkin (1974) showed that (11) is indeed true when \( G = S_n \), while Eaton and Perlman (1977) extended this
to the case where \( G \) is any reflection group, i.e., a group of orthogonal transformations generated by simple reflections in \( \mathbb{R}^n \). (Since any permutation is a product of simple transpositions, \( S_n \) is a reflection group.) Later, Eaton (1984) produced a counterexample to show that (11) fails when \( G \) is a group of rotations acting on \( \mathbb{R}^2 \). It remains an open question to characterize those subgroups \( G \) of orthogonal transformations for which (11) is valid.

Many other modifications and/or extensions of Anderson's theorem also have been obtained, of which we mention only one. Returning to (5) and (6), it may be established by means of the Brunn-Minkowski inequality that for any two convex subsets \( C, K \) of \( \mathbb{R}^n \), the convolution \( I_C * I_K \) is not only Anderson-unimodal but is in fact logarithmically concave, a stronger property. If we notice that \( I_C \) and \( I_K \) are both log concave, this raises the following question: for any two nonnegative functions \( f, g \) on \( \mathbb{R}^n \), is it true that

\[
(12) \quad f, g \text{ log concave } \implies f * g \text{ is log concave?}
\]

This question was answered affirmatively by Davidović, Korenbljum, and Hacet (1969), who again invoked Anderson's theorem in their proof. Prekopa (1971, 1973) and Rinott (1976) established the following generalization of (12): if \( f(x, y) \) is log concave on \( \mathbb{R}^{m+n} \), then

\[
(13) \quad h_4(y) = \int f(x, y) \, dx
\]

is log concave on \( \mathbb{R}^n \). Brascamp and Lieb (1974) presented a simple and elegant proof of this result based directly on the Brunn-Minkowski inequality.

The monographs by Tong (1980) and Dharmadhikari and Joag-dev (1986) contain excellent and comprehensive surveys of these topics, as well as complete bibliographies. Other valuable reviews appear in the papers by Das Gupta (1980) and Eaton (1982) and in the recent monograph by Eaton (1987).

3. Applications to the concentration of multivariate distributions.

Now we turn to a discussion of several of the many applications that have been found for Anderson's theorem and its extensions. In his original paper [20], Anderson immediately noted the following application. Suppose that
\[ X \sim N_n(0, \Sigma) \]
\[ Z \sim N_n(0, \Psi) \]

with \( \Delta \equiv \Psi - \Sigma \) positive semidefinite (\( \equiv \) psd), where \( N_n(0, \Sigma) \) denotes the \( n \)-dimensional normal distribution with mean vector 0 and covariance matrix \( \Sigma \). Then Anderson showed that the distribution of \( X \) is more concentrated (\( \equiv \) more peaked) about 0 than that of \( Z \), in the sense that

\[ P[Z \in K] \leq P[X \in K] \]

for every symmetric convex set \( K \) in \( \mathbb{R}^n \). To establish this fact, Anderson noted that \( Z \) may be represented as

\[ Z = X + Y, \]

where \( Y \sim N_n(0, \Delta) \) is independent of \( X \). Then

\[ P[Z \in K] = E\{P[X \in K - Y | Y]\} \]
\[ \leq E\{P[X \in K | Y]\} \]
\[ = P[X \in K], \]

where the inequality follows from Anderson’s theorem.

Fefferman, Jodeit, and Perlman (1972) showed that (15) remains valid for any family of elliptically contoured distributions. That is, if \( X: 1 \times n \) has probability density on \( \mathbb{R}^n \) of the form

\[ f_\Sigma(x) = |\Sigma|^{-\frac{1}{2}} \phi(x\Sigma^{-1}x'), \]

then for any symmetric convex set \( K \),

\[ P_\Psi[X \in K] \leq P_\Sigma[X \in K] \]

whenever \( \Psi - \Sigma \) is psd.

Remarkably, the proof of this result again is based on Anderson’s fundamental theorem. The inequality (19) may be restated in the form

\[ P_l[X \in \bar{K}] \leq P_l[X \in D\bar{K}], \]

where \( \bar{K} \) is the image of \( K \) under an appropriate linear transformation (hence \( \bar{K} \) is also symmetric and convex) and where \( D \) is the diagonal matrix \( \text{Diag}(d_1, \ldots, d_n) \) where \( d_1^2, \ldots, d_n^2 \) are the characteristic roots of \( \Psi \Sigma^{-1} \), hence each \( d_i \geq 1 \) (thus \( D \) is a dilation). Because the distribution of \( X \) is spherically
symmetric when $\Sigma = I$, the conditional distribution of $X$ given $\|X\| = r$ is uniform on the sphere $S_r$ of radius $r$, so (20) will follow from the stronger inequality

$$(21) \quad \nu(\tilde{K} \cap S_r) \leq \nu(D\tilde{K} \cap S_r),$$

where $\nu$ denotes the uniform surface measure (similar to Lebesgue measure) on $S_r$. (Note that $\tilde{K} \cap S_r$ is not necessarily contained in $D\tilde{K} \cap S_r$.) Since

$$D = \prod_{i=1}^{n} D_i,$$

where $D_i = \text{Diag}(1, \ldots, 1, d_i, 1, \ldots, 1)$, it suffices to establish (21) when $D = D_1 = \text{Diag}(d_1, 1, \ldots, 1)$. By means of the Divergence Theorem, however, it can be shown that for every suitably smooth $K$,

$$(22) \quad \frac{\partial}{\partial d_1} [\nu(D_1\tilde{K} \cap S_r)] = -c^* \frac{d^2}{dt^2} \left[ (I_{B_r} \ast I_{D_1\tilde{K}})(ty) \right]_{t=0}$$

where $c^*$ is a positive constant, $B_r$ is the ball of radius $r$ (so $S_r = \partial B_r$), and $y = (1, 0, \ldots, 0)$. Because both $B_r$ and $D_1\tilde{K}$ are symmetric convex sets, Anderson's theorem implies that $(I_{B_r} \ast I_{D_1\tilde{K}})(ty)$ achieves its maximum at $t = 0$, hence the second derivative on the right-hand side of (22) is negative. Therefore $\nu(D_1\tilde{K})$ is nondecreasing in $d_1$, which, as already noted, establishes (21), (20), and hence (19). (An alternate derivation of (19), also based on Anderson's theorem, was given by Das Gupta, Eaton, Olkin, Perlman, Savage, and Sobel (1972).)

4. Applications to simultaneous confidence intervals.

A second application of Anderson's theorem yields an important lower bound for the coverage probability of simultaneous confidence intervals for the components of the mean vector of a multivariate normal population. Suppose that

$$(23) \quad X \equiv (X_1, \ldots, X_n) \sim N_n(\mu, \Sigma),$$

where $\mu = (\mu_1, \ldots, \mu_n)$ and $\Sigma = (\sigma_{ij} | i, j = 1, \ldots, n)$. Khatri (1967), Sidak (1967, 1968), and Jogdeo (1970) applied Anderson's theorem to show that for any positive constants $a_1, \ldots, a_n$,

$$(24) \quad P_{\Sigma} \left[ |X_1 - \mu_1| \leq a_1, \ldots, |X_n - \mu_n| \leq a_n \right] \geq P_D(\sigma) \left[ |X_1 - \mu_1| \leq a_1, \ldots, |X_n - \mu_n| \leq a_n \right]$$

$$\equiv \prod_{i=1}^{n} P_{\sigma_i} \left[ |X_i - \mu_i| \leq a_i \right],$$
where $D(\sigma) = \text{Diag}(\sigma_{11}, \ldots, \sigma_{nn})$, thus providing conservative lower bounds for the probability content of any set of simultaneous confidence intervals for $\mu_1, \ldots, \mu_n$ when $\sigma_{11}, \ldots, \sigma_{nn}$ are known.

Here we sketch Khatri’s proof of a slightly more general result. Suppose that

$$Y = (Y_1, Y_2) \sim N_{n_1+n_2} \left( (0, 0), \Sigma = \begin{bmatrix} \Sigma_{11} & \Sigma_{12} \\ \Sigma_{21} & \Sigma_{22} \end{bmatrix} \right),$$

with $Y_i : 1 \times n_i$, $i = 1, 2$. Let $C_1, C_2$ be two symmetric convex subsets in $\mathbb{R}^{n_1}$ and $\mathbb{R}^{n_2}$, respectively. Khatri proved that

$$P_{\Sigma} [Y_1 \in C_1, Y_2 \in C_2] \geq P_{\Sigma_1} [Y_1 \in C_1] P_{\Sigma_2} [Y_2 \in C_2]$$

when $\text{rank} (\Sigma_{12}) = 1$, i.e., when $\Sigma_{12} = \alpha_1' \alpha_2$ with $\alpha_i : 1 \times n_i$. The result (24) follows from (26) by taking $Y = X - \mu$, $n_1 = 1$, $n_2 = n - 1$, $C_1 = \{|y_1| \leq a_1\}$, and $C_2 = \{|y_2| \leq a_2, \ldots, |y_n| \leq a_n\}$, and then applying induction on $n$.

To verify (26), Khatri (1967) introduced the extended normal random vector

$$(Y_1, Y_2, Z) \sim N_{n_1+n_2+1} \left( (0, 0, 0), \begin{bmatrix} \Sigma_{11} & \Sigma_{12} & \alpha_1' \\ \Sigma_{21} & \Sigma_{22} & \alpha_2' \\ \alpha_1 & \alpha_2 & 1 \end{bmatrix} \right).$$

It is readily verified that the covariance matrix appearing in (27) is psd and that, conditionally,

$$(Y_1, Y_2 | Z) \sim N_{n_1, n_2} \left( (Z\alpha_1, Z\alpha_2), \begin{bmatrix} \Sigma_{11} - \alpha_1' \alpha_1 & 0 \\ 0 & \Sigma_{22} - \alpha_2' \alpha_2 \end{bmatrix} \right).$$

Therefore $Y_1$ and $Y_2$ are conditionally independent given $Z$, so

$$P_{\Sigma} [Y_1 \in C_1, Y_2 \in C_2] = E \left\{ P [Y_1 \in C_1 | Z] P [Y_2 \in C_2 | Z] \right\}$$

$$\geq E \left\{ P [Y_1 \in C_1 | Z] \right\} E \left\{ P [Y_2 \in C_2 | Z] \right\}$$

$$= P_{\Sigma_1} [Y_1 \in C_1] P_{\Sigma_2} [Y_2 \in C_2].$$

The inequality follows from the fact that for $i = 1, 2,$

$$P [Y_i \in C_i | Z] = P [N_{n_i}(0, \Sigma_{ii} - \alpha_i' \alpha_i) \in C_i - Z \alpha_i | Z]$$

are each nonincreasing functions of $|Z|$ by Anderson’s theorem, hence are nonnegatively correlated.
It has long been conjectured that the inequality (26) remains valid without the assumption that rank(\(\Sigma_{12}\)) = 1, but the proof of this fact remains elusive (although several incorrect proofs have been published). Pitt (1977) has established the validity of (26) when rank(\(\Sigma_{12}\)) = 2. Recently, Pitt informed this reviewer of the existence of a proof of (26) in the general case, by W. Beckner, but that proof remains unpublished.

Anderson’s theorem may be used to derive the following reversal of (26) with no restriction on \(\Sigma\) other than that it be positive definite:

\[
(27) \quad P_{\Sigma} [Y_1 \in C_1, Y_2 \in C_2] \leq P_{\Sigma_{11}^{-1}} [Y_1 \in C_1] P_{\Sigma_2} [Y_2 \in C_2],
\]

where \(\Sigma_{11}^{-1} = \Sigma_{11} - \Sigma_{12} \Sigma_{22}^{-1} \Sigma_{21}\). To obtain this, note that

\[
(28) \quad V = Y_1 - Y_2 \Sigma_{22}^{-1} \Sigma_{21} - N_n(0, \Sigma_{11}^{-1})
\]

\[
Y_2 \sim N_n(0, \Sigma_{22}),
\]

and \(V, Y_2\) are independent. Therefore

\[
(29) \quad P_{\Sigma} [Y_1 \in C_1, Y_2 \in C_2] = E \left\{ P \left[ V \in C_1 - Y_2 \Sigma_{22}^{-1} \Sigma_{21}, Y_2 \in C_2 | Y_2 \right] \right\}
\]

\[
= E \left\{ P \left[ V \in C_1 - Y_2 \Sigma_{22}^{-1} \Sigma_{21} | Y_2 \right] I_{C_1}(Y_2) \right\}
\]

\[
\leq E \left\{ P \left[ V \in C_1 | Y_2 \right] I_{C_1}(Y_2) \right\}
\]

\[
= P \left[ V \in C_1 \right] P \left[ Y_2 \in C_2 \right],
\]

where the inequality follows from Anderson’s theorem. Clearly, (29) is equivalent to (27).

An extension of the Khatri-Sidak-Jogdeo result from the normal distribution to general elliptically contoured distributions was obtained by Das Gupta, Eaton, Olkin, Perlman, Savage, and Sobel (1972). Their result takes the following form. Let \(X = (X_1, X_2)\) with \(X : 1 \times n\) and \(X_2 : 1 \times (n-1)\), let \(a > 0\), and let \(C\) be a symmetric convex subset of \(\mathbb{R}^{n-1}\). If \(X\) has probability density of the form (18) with \(\Sigma\) replaced by

\[
\Sigma_{\lambda} = \begin{bmatrix} \Sigma_{11} & \lambda \Sigma_{12} \\ \lambda \Sigma_{21} & \Sigma_{22} \end{bmatrix},
\]

where \(\Sigma_{12} \sim 1 \times (n-1)\) and \(0 \leq \lambda \leq 1\), then

\[
(30) \quad P_{\Sigma_{\lambda}} [ | X_1 | \leq a, X_2 \in C ] \text{ is increasing in } \lambda;
\]
in particular

\begin{equation}
P_{\Sigma}[|X_1| \leq a, X_2 \in C] \geq P_{\Sigma}[|X_1| \leq a, X_2 \in C].
\end{equation}

Note, however, that the right-hand side of (31) cannot be factored as in (26) unless \( X \) is normally distributed. Again, it is remarkable that the proof of (30) (too lengthy to present here) rests on Anderson's fundamental theorem. (Sidak (1968) and Jogdeo (1970) established (30) when \( X \) is normally distributed.)

5. Unbiasedness and power monotonicity of invariant tests for MANOVA, independence, and related problems.

Many applications of Anderson's theorem concern the unbiasedness and power monotonicity of multivariate tests. Das Gupta, Anderson, and Mudholkar [41] applied this theorem to establish the unbiasedness and power monotonicity of a wide class of invariant tests arising in the multivariate analysis of variance (\( \equiv \) MANOVA).

The MANOVA testing problem may be stated in the following canonical form (cf. Anderson (1984), Chapter 8). Suppose that \( X \) and \( Y \) are independent random matrices such that

\begin{align*}
X & \sim N_{pr}(\mu, \Sigma \otimes I): p \times r \\
Y & \sim N_{pn}(0, \Sigma \otimes I): p \times n;
\end{align*}

that is, the columns of \( X \) and \( Y \) are mutually independent \( p \)-variate normal random vectors with common covariance matrix \( \Sigma \). The problem is to test

\begin{equation}
H_0: \mu = 0 \quad \text{vs.} \quad H_1: \mu \neq 0
\end{equation}

with \( \Sigma \) positive definite but unknown, where \( \mu: p \times r \) is a matrix of unknown means. This testing problem is invariant under the group of linear transformations

\begin{equation}
(X, Y) \rightarrow (BX\Gamma, BY\Psi),
\end{equation}

where \( B: p \times p \) is nonsingular and \( \Gamma: r \times r, \Psi: n \times n \) are orthogonal. A maximal invariant statistic under this group of transformations is the vector

\begin{equation}
c = (c_1, \ldots, c_t),
\end{equation}

where \( t = \min(p, r) \) and \( c_1 > \cdots > c_t > 0 \) are the nontrivial characteristic roots of the matrix
XX'(YY)^{-1}. (Assume that $n \geq p$ so that YY' is nonsingular with probability one.) A maximal invariant parameter is the vector of noncentrality parameters

\[(\lambda_1, \ldots, \lambda_r),\]

where $\lambda_1 \geq \cdots \geq \lambda_r \geq 0$ are the nontrivial characteristic roots of $\mu \Sigma^{-1}$. The power function of an invariant test for (33) depends on $(\mu, \Sigma)$ only through the value of $\lambda$ (see (38)). Note also that $\mu = 0$ if and only if $\lambda = 0$.

Das Gupta, Anderson, and Mudholkar [41] derived several sufficient conditions for the power function of an invariant test for (33) to be monotonically increasing in each noncentrality parameter $\lambda_1, \ldots, \lambda_r$, and therefore unbiased. Their basic result states that if the acceptance region $A$ of an invariant test (i.e., one depending on $(X, Y)$ only through the value of $c$) is convex in each column vector $X_i$ of $X$ when $Y$ and the remaining columns of $X$ are held fixed, then the power function

\[\pi(\mu, \Sigma) = P_{\mu, \Sigma}((X, Y) \notin A)\]

is monotonically increasing in $\lambda_1, \ldots, \lambda_r$. First, they observed that the invariance of $A$ implies that

\[\pi(\mu, \Sigma) = \pi(D_{\lambda}, I),\]

where $D_{\lambda}$ is the $p \times r$ matrix with $i$-th entry $\lambda_i^{j-1}$, $i = 1, \ldots, r$, and all other entries 0. Second, the invariance of $A$ under the transformations (34) implies that each $X_i$-section of $A$ is a symmetric subset of $\mathbb{R}^p$. Together with the convexity assumption, this allows one to apply Anderson's theorem to the conditional probability

\[P_{D_{\lambda}, I}[(X, Y) \in A \mid Y, X_j, j \neq i].\]

Since

\[X_i \sim N_p[\lambda_i e_i, I]\]

when $(\mu, \Sigma) = (D_{\lambda}, I)$, where $e'_i = (0, \ldots, 0, 1, 0, \ldots, 0)$ with the 1 in position $i$, Anderson's theorem implies that (39) is decreasing in $\lambda_i$. Furthermore, the distributions of $Y$ and $X_j, j \neq i$, do not depend on $\lambda_i$, hence the unconditional probability must also decrease with $\lambda_i$, which establishes the stated result.

The following four invariant test statistics frequently have been proposed for testing (33):

\[c_1 = \text{ch}_{max}[XX'(YY)^{-1}], \quad (S. N. Roy)\]

\[\sum_{i=1}^t c_i = \text{tr}[XX'(YY)^{-1}], \quad \text{(Lawley-Hotelling)}\]
Das Gupta, Anderson, and Mudholkar [41] showed that the three tests based on (41), (42) and (43) satisfy their convexity assumption, hence possess monotone power functions. Perlman (1974) noted, however, that for the usual significance levels, the test (44) violates the convexity assumption of [41]. Monotonicity of the power function of the test (44) remains an open question, although its unbiasedness was established by Perlman and Olkin (1980). They showed that any invariant test statistic \( g(c_1, \ldots, c_t) \) with \( g \) nondecreasing in each \( c_i \) determines an unbiased test for (33). Their proof is not based on Anderson’s theorem but instead upon monotone likelihood ratio properties of the noncentral probability density function of \( c \equiv (c_1, \ldots, c_t) \).

It is also interesting to note that Anderson and Perlman (1988) have shown that, unlike (41), (42), (43), the test based on (44) fails to be parameter consistent whenever it fails to satisfy the convexity assumptions of [41]. That is, in such cases the power of the test based on (44) remains bounded below 1 as \( \lambda_1 \to \infty \) with \( \lambda_2, \ldots, \lambda_t \) bounded. Although test (44) is the locally most powerful invariant test for (33) (cf. Schwartz (1967)), Anderson and Perlman (1988) recommend against its use on this basis.

Eaton and Perlman (1974) applied Mudholkar’s extension of Anderson’s theorem to exhibit more detailed monotonicity properties for the power functions of a subclass of invariant tests for (33); this subclass includes the tests (41) and (42) but not (43) or (44).

This discussion of the paper [41] shows that Anderson’s theorem may be applied directly to establish the unbiasedness and power monotonicity of a broad class of tests for a multivariate location parameter. Remarkably, for hypothesis-testing problems involving the covariance matrix of the multivariate normal distribution, Anderson’s result is also fundamental for determining the behavior of the power functions of invariant tests, as will now be indicated.

In a companion paper to [41], Anderson and Das Gupta [42] obtained conditions for the unbiasedness and power monotonicity of invariant tests for the independence of two sets of normal variates. Suppose that \( S \) is the (unnormalized) sample covariance matrix based on a random sample of size \( m \) from the \( p \)-dimensional normal population \( N_p(\mu, \Sigma) \). Partition \( S \) and \( \Sigma \) as

\[
\prod_{i=1}^{t} (c_i + 1) \equiv |XX' + YY'| / |YY'| \tag{Likelihood Ratio Criterion}
\]

\[
\sum_{i=1}^{n} c_i (c_i + 1)^{-1} = \text{tr} [XX'(XX' + YY')^{-1}] \tag{Bartlett-Nanda-Pillai}
\]
\( S = \begin{bmatrix} S_{11} & S_{12} \\ S_{21} & S_{22} \end{bmatrix}, \quad \Sigma = \begin{bmatrix} \Sigma_{11} & \Sigma_{12} \\ \Sigma_{21} & \Sigma_{22} \end{bmatrix}, \)

with \( S_{ij}, \Sigma_{ij} : p_i \times p_j, i, j = 1, 2, \) where \( p = p_1 + p_2. \) The problem of testing independence is that of testing

\( H_0 : \Sigma_{12} = 0 \quad \text{vs.} \quad H_1 : \Sigma_{12} \neq 0. \)

This problem is invariant under linear transformations of the form

\[
\begin{bmatrix} S_{11} & S_{12} \\ S_{21} & S_{22} \end{bmatrix} \to \begin{bmatrix} B_1 & 0 \\ 0 & B_2 \end{bmatrix} \begin{bmatrix} S_{11} & S_{12} \\ S_{21} & S_{22} \end{bmatrix} \begin{bmatrix} B_1' & 0 \\ 0 & B_2' \end{bmatrix},
\]

where \( B_i : p_i \times p_i \) is nonsingular, \( i = 1, 2. \) The maximal invariant statistic and parameter can be represented by

\[
r = (r_1, \ldots, r_t) \quad \text{and} \quad \rho = (\rho_1, \ldots, \rho_t),
\]

respectively, where \( t = \min(p_1, p_2) \) and \( 1 > r_1 > \cdots > r_t > 0 \) and \( 1 \geq \rho_1 \geq \cdots \geq \rho_t \geq 0 \) are, respectively, the squared sample and population canonical correlation coefficients, i.e., the nontrivial characteristic roots of the matrices

\[
S_{12}S_{22}^{-1}S_{21}^{-1}, \quad \Sigma_{12}S_{22}^{-1}\Sigma_{21}^{-1},
\]

respectively. Invariant tests for (46) depend on \( S \) only through \( r. \)

The \( p \times p \) random matrix \( S \) has the Wishart distribution with \( n = m - 1 \) degrees of freedom, i.e.,

\[
S \sim W_p(\Sigma, n).
\]

Anderson and Das Gupta [42] noted that if \( n \geq p, \) then \( S_{12}S_{22}^{-1/2} \) and \( S_{11:2} = S_{11} - S_{12}S_{22}^{-1}S_{21} \) are conditionally independent given \( S_{22}, \) and that

\[
S_{12}S_{22}^{-1/2}|S_{22} \sim N_{p_1p_2}(\Sigma_{12}S_{22}^{-1/2}, \Sigma_{11:2} \otimes I): p_1 \times p_2
\]

\[
S_{11:2}|S_{22} \sim W_p(\Sigma_{11:2}, n-p_2): p_1 \times p_1.
\]

Thus the joint conditional distribution of \( S_{12}S_{22}^{-1/2} \) and \( S_{11:2} \) is of the same form as that of \((X, YY')\) with \((X, Y)\) as in (32). Since \( \Sigma_{12} = 0 \) if and only if \( \Sigma_{12}S_{22}^{-1}S_{22}^{-1/2} = 0, \) the testing problem (46) is of the same type as (33), hence the results of [41] can be applied to establish unbiasedness and power monotonicity (as \( \rho_1, \ldots, \rho_t \) increase) of invariant tests for (46), first conditionally given \( S_{22} \) and then unconditionally.
Because the characteristic roots of
\[ S_{12} S_{22}^{-1} S_{21} S_{11}^{-1} \quad [\leftrightarrow XX'(YY)^{-1}] \]
are
\[ r_1(1-r_1)^{-1}, \ldots, r_i(1-r_i)^{-1}, \]
their conditional distribution given \( S_{22} \) has the same form as the distribution of \((c_1, \ldots, c_t)\) in (35). Therefore, Anderson and Das Gupta [42] conclude that an invariant test for \((46)\) based on a statistic of the form
\[ g\left(r_1(1-r_1)^{-1}, \ldots, r_i(1-r_i)^{-1}\right) \]
is unbiased and has a monotone power function whenever the corresponding test for \((33)\) based on \(g(c_1, \ldots, c_i)\) enjoys these properties. In particular, the three tests based on the statistics
\[ r_1 \equiv \text{ch}_{\text{max}} \left[ S_{12} S_{22}^{-1} S_{21} S_{11}^{-1} \right] \quad \text{(S. N. Roy)} \]
\[ \sum_{i=1}^{t} \frac{r_i}{1-r_i} \equiv \text{tr} \left[ S_{12} S_{22}^{-1} S_{21} S_{11}^{-1} \right] \quad \text{(Lawley-Hotelling)} \]
\[ \prod_{i=1}^{t} \frac{1}{1-r_i} \equiv \frac{|S|}{|S_{11}| |S_{22}|} \quad \text{(Likelihood Ratio Criterion)} \]
are unbiased and have monotone power functions (compare to (41), (42), (43)).

Fujikoshi (1973) applied a similar conditioning argument to show that the results in [41] for the MANOVA problem extend directly to the generalized MANOVA (GMANOVA) testing problem, also referred to as the growth curves model testing problem.

Perlman (1980) applied Anderson’s theorem to establish the unbiasedness of Bartlett’s test (the modified likelihood ratio test) for the equality of the covariance matrices of several \(p\)-dimensional normal populations, i.e., for testing
\[ H_0: \Sigma_1 = \cdots = \Sigma_k \quad \text{vs.} \quad H_1: \text{not} \; H_0. \]
He also established the unbiasedness of the (unmodified) likelihood ratio test for the simultaneous equality of mean vectors and covariance matrices of several normal populations. The proofs use an extension of the conditioning argument appearing in (27), (28), and (29).
Finally, also in the 1964 volume of the *Annals of Mathematical Statistics*, Anderson and Das Gupta [44] presented sufficient conditions for the unbiasedness and power monotonicity of invariant tests of

\[(59) \quad H_0: \Sigma_1 = \Sigma_2 \quad \text{vs.} \quad H_1: \Sigma_1 - \Sigma_2 \text{ psd},\]

a one-sided version of problem (58) when \(k = 2\). If \(S_1\) and \(S_2\) are independent Wishart matrices with

\[(60) \quad S_1 \sim W_p(\Sigma_1, n_1), \quad S_2 \sim W_p(\Sigma_2, n_2),\]

invariant tests for (59) must depend on \(S_1, S_2\) through the values of \(d_1 > \cdots > d_p > 0\), the ordered characteristic roots of \(S_1 S_2^{-1}\), and their power functions must depend on \(\Sigma_1, \Sigma_2\) through \(\delta_1 \geq \cdots \geq \delta_p > 0\), the ordered characteristic roots of \(\Sigma_1 \Sigma_2^{-1}\). Anderson and Das Gupta [44] showed that any invariant test based on \(g(d_1, \ldots, d_p)\), with \(g\) nondecreasing in each \(d_i\), must have a power function that is nondecreasing in each \(\delta_i\). Their proof is not based on Anderson’s theorem (note that no convexity assumptions are imposed on \(g\) or on the acceptance region expressed in terms of \(S_1\) and \(S_2\)) but instead upon the more elementary fact that if \(F, F_1, F_2\) are \(p \times p\) positive definite matrices such that \(F_1 - F_2\) is psd, then

\[(61) \quad ch_i(FF_1) \geq ch_i(FF_2), \quad i = 1, \ldots, p,\]

where \(ch_1(T) \geq \cdots \geq ch_p(T)\) denote the ordered characteristic roots of \(T\). The result (61) was established by Anderson and Das Gupta in [40].
References to papers by T. W. Anderson


Additional references


