The Annals of Mathematical Statistics
1972, Vol. 43, No. 1, 123-141

LARGE SAMPLE TESTS FOR THE EQUALITY OF TWO COVARIANCE MATRICES[1]

BY M. W. J. LAYARD

University of California, Davis

0. Introduction. It is well known that the standard normal-theory techniques for testing hypotheses about variances are extremely non-robust, even asymptotically, against departures from the assumed normality of the underlying distributions. As Box [4] points out, the reason for this frailty is that the relevant statistics, though asymptotically normal under general conditions, do not incorporate a corrective component to ensure the stability of the asymptotic variance under departures from normality. (By way of contrast, the t-statistic is "self-normalizing.")

Not surprisingly, the same difficulty arises in the multivariate case when we test hypotheses about covariance matrices. There is a well-developed body of normal-theory procedures for testing such hypotheses as the equality of two covariance matrices, independence of sets of variates, sphericity, etc., as well as for testing hypotheses about certain functions of covariance matrices, such as correlation coefficients and regression coefficients. It is the purpose of this paper to point out that the standard tests for equality of two covariance matrices are non-robust against departures from normality, and to discuss several procedures which are, at least asymptotically, robust. The non-robustness of normal-theory tests about correlation coefficients and the structure of covariance matrices will be discussed elsewhere.

In Section 1 we describe some notation and state some large-sample theory results which are needed in the sequel.

Section 2 demonstrates the non-robustness of two normal-theory tests for the equality of two covariance matrices. If S and T are the sample covariance matrices, then in the normal case the characteristic roots of ST^{-1} constitute a maximal invariant (under the full linear group), and the normal-theory tests which have been proposed employ functions of those roots. The two tests examined are the likelihood-ratio test, which uses a function of all the roots, and Roy and Gnanadesikan's test, which is based on the smallest and largest roots.

Section 3 considers four asymptotically robust procedures for testing the equality of two covariance matrices. The first of these, which is based on the elementary symmetric functions of the roots of ST^{-1} (where S and T are the

Received March 19, 1970; revised June 10, 1971.

[1] Adapted from the author's doctoral dissertation at Stanford University. The research was supported in part by National Science Foundation Grant 6681.

sample covariance matrices), does not appear to have much practical utility. The remaining three tests exploit the asymptotic normality of the second-order moments by treating the problem as a test for equality of mean vectors. They are

(i) a "standard error" test, in which the vector of differences of the transformed second-order sample moments is standardized by means of an estimate of its asymptotic covariance matrix;

(ii) a test based on Box's [4] idea of dividing the data into groups, computing certain functions of the second-order moments for each group, and treating the resulting statistics as approximately normal;

(iii) a test based on the jackknife procedure (Quenouille [15], Tukey [17], Miller [11]).

In Section 4 we make some asymptotic comparisons of the above three tests for some simple classes of alternatives, using the concepts of Pitman efficiency and approximate Bahadur efficiency. In terms of these comparisons, it is seen that the Box test with small group sizes is somewhat less powerful than the standard error and jackknife tests, and that the latter tests are not inferior to the normal-theory likelihood-ratio test when the distributions are in fact normal.

1. Notation and preliminary results. In this section we set up some notation and state some results from large sample theory which will be needed in subsequent sections. Since the discussion throughout the paper is mainly confined to bivariate distributions, the notation is established for that case. The notation $\operatorname{diag}(\nu_1, \nu_2, \cdots, \nu_k)$ is used for the diagonal matrix whose diagonal elements are $(\nu_1, \nu_2, \cdots, \nu_k)$.

Suppose we have a sample of n i.i.d. observations (X_i, Y_i) from a bivariate distribution with distribution function $F(x, y)$ and finite fourth moments. Denote the central moments by

$$\mu_{ij} = E(X - EX)^i(Y - EY)^j ,$$

the covariance matrix by

$$\Sigma = \begin{bmatrix} \mu_{20} & \mu_{11} \\ \mu_{11} & \mu_{02} \end{bmatrix},$$

and the sample covariance matrix by

$$S = \begin{bmatrix} s_{20} & s_{11} \\ s_{11} & s_{02} \end{bmatrix}.$$

Then it is well known (see, e.g., Cramér [5] page 365), that

$$n^{\frac{1}{2}}(s - \mu) \to_{\mathcal{L}} N(0, \Gamma) ,$$

where

$$s' = (s_{20}, s_{02}, s_{11}) , \qquad \mu' = (\mu_{20}, \mu_{02}, \mu_{11})$$

and

$$\Gamma = \begin{bmatrix} \mu_{40} - \mu_{20}^2 & \mu_{22} - \mu_{20}\mu_{02} & \mu_{31} - \mu_{20}\mu_{11} \\ & \mu_{04} - \mu_{02}^2 & \mu_{13} - \mu_{02}\mu_{11} \\ & & \mu_{22} - \mu_{11}^2 \end{bmatrix}. \tag{1.1}$$

(The symbol $\rightarrow_{\mathscr{L}}$ denotes convergence in distribution, and $N(0, \Gamma)$ denotes the multivariate normal distribution with mean vector 0 and covariance matrix Γ.)

Suppose we are considering the asymptotic distribution of the sample correlation coefficient, $\hat{\rho} = s_{11}(s_{20}s_{02})^{-\frac{1}{2}}$, or of a function of the roots of $|S - \theta T| = 0$ (where T is the sample covariance matrix of a second sample from a distribution with the same covariance matrix and fourth moments). It is easily seen that a transformation of the observations with matrix $\operatorname{diag}(\mu_{20}^{-\frac{1}{2}}, \mu_{02}^{-\frac{1}{2}})$ leaves such a statistic unchanged, so that we can without loss of generality assume $\mu_{20} = \mu_{02} = 1$. This transformation has the effect of multiplying s by the matrix $\operatorname{diag}(\mu_{20}^{-1}, \mu_{02}^{-1}, (\mu_{20}\mu_{02})^{-\frac{1}{2}})$. Noting that

$$\begin{aligned} \mu_{40} &= \kappa_{40} + 3\mu_{20}^2 & \mu_{31} &= \kappa_{31} + 3\mu_{20}\mu_{11} \\ \mu_{04} &= \kappa_{04} + 3\mu_{02}^2 & \mu_{13} &= \kappa_{13} + 3\mu_{02}\mu_{11} \\ & & \mu_{22} &= \kappa_{22} + \mu_{20}\mu_{02} + 2\mu_{11}^2 \end{aligned}$$

(where the κ_{ij} are the fourth-order cumulants of the bivariate distribution), and defining the standardized cumulants

$$\begin{aligned} \lambda_{40} &= \kappa_{40}\mu_{20}^{-2} & \lambda_{31} &= \kappa_{31}\mu_{20}^{-\frac{3}{2}}\mu_{02}^{-\frac{1}{2}} \\ \lambda_{04} &= \kappa_{04}\mu_{02}^{-2} & \lambda_{13} &= \kappa_{13}\mu_{20}^{-\frac{1}{2}}\mu_{02}^{-\frac{3}{2}} \\ & & \lambda_{22} &= \kappa_{22}\mu_{20}^{-1}\mu_{02}^{-1} \end{aligned}$$

we see that we can assume

$$\Gamma = \begin{bmatrix} \lambda_{40} + 2 & \lambda_{22} + 2\rho^2 & \lambda_{31} + 2\rho \\ & \lambda_{04} + 2 & \lambda_{13} + 2\rho \\ & & \lambda_{22} + (1 + \rho^2) \end{bmatrix}. \tag{1.2}$$

(Note that λ_{40} and λ_{04} are the kurtoses of the marginal distributions.) In the bivariate normal case, in which all the λ_{ij} vanish, we have

$$\Gamma = \begin{bmatrix} 2 & 2\rho^2 & 2\rho \\ & 2 & 2\rho \\ & & 1 + \rho^2 \end{bmatrix}. \tag{1.3}$$

(Cf. Anderson [1] page 77). In the case of a bivariate distribution with i.i.d. components, we have $\Gamma = \operatorname{diag}(2 + \gamma, 2 + \gamma, 1)$, where γ is the common kurtosis of the marginal distributions.

In the characteristic roots problem we can multiply the observations by $\Sigma^{-\frac{1}{2}}$ and leave the roots unchanged, in which case Γ can again be assumed to be a

function of ρ and the λ_{ij}, though not so simple an expression as (1.2). However, in the normal case Γ takes the particularly simple form diag(2, 2, 1).

The remarks above extend in an obvious way to the case of p-variate distributions, the matrix Γ being a function of correlation coefficients and fourth-order standardized cumulants.

For ease of reference we state without proof some results from large-sample theory which are used in the sequel.

THEOREM 1.1. *If $g: E^p \to E^k$ is a continuous function and $\{X_n\}$ is a sequence of p-dimensional random vectors such that*

$$X_n \to_{\mathscr{L}} X, \quad \text{then} \quad g(X_n) \to_{\mathscr{L}} g(X).$$

THEOREM 1.2. *Let $\{X_n\}$ be a sequence of random vectors such that $X_n \to_{\mathscr{L}} X$, and let $\{B_n\}$ be a sequence of symmetric random matrices such that $B_n \to_P B$, where B is a positive definite symmetric matrix of real elements. (The symbol $\to_P$ denotes convergence in probability.) Then*

$$X_n' B_n^{-1} X_n \to_{\mathscr{L}} X' B^{-1} X$$

THEOREM 1.3. *Suppose $\{X_n\}$ is a sequence of p-dimensional random vectors such that*

$$n^{\frac{1}{2}}(X_n - \mu) \to_{\mathscr{L}} N(0, \Sigma).$$

(i) *Let $f: E^p \to E^k$ be a function admitting continuous first partial derivatives at μ, at least one of which does not vanish at μ. Then $n^{\frac{1}{2}}(f(X_n) - f(\mu)) \to_{\mathscr{L}} N(0, A'\Sigma A)$, where A is the $p \times k$ matrix of first partials of f evaluated at μ.*

(ii) *Let $f: E^p \to E^1$ be a function admitting continuous second partials at μ, at least one of which does not vanish at μ, and such that all the first partials vanish at μ. Then $n(f(X_n) - f(\mu)) \to_{\mathscr{L}} \frac{1}{2} X'BX$, where $X \sim N(0, \Sigma)$ and B is the $p \times p$ symmetric matrix of second partials of f evaluated at μ.*

2. Non-robustness of normal-theory procedures for testing $\Sigma_1 = \Sigma_2$. Suppose we have independent samples of size n and m respectively from two p-variate distributions with cdf's F and G, covariance matrices Σ_1 and Σ_2 and finite fourth moments. The problem is to test

$$H_0: F(x_1, \cdots, x_p) = G(x_1 + \xi_1, \cdots, x_p + \xi_p),$$

where $\xi_1, \cdots, \xi_p$ are unspecified constants,

$$\text{vs. } H_A: \Sigma_1 \neq \Sigma_2.$$

(H_0 implies $\Sigma_1 = \Sigma_2$ but in general the reverse implication does not hold, though it does in the normal case. The choice of H_0 rather than the more general hypothesis $\Sigma_1 = \Sigma_2$ ensures that the fourth moments of the two distributions are equal.)

If F and G are assumed to be normal cdf's, then the sample covariance matrices S and T, together with the sample mean vectors, form a set of sufficient statistics. A maximal invariant (under the full linear group) for testing H_0 can be shown to be $\hat{\theta}_1 < \hat{\theta}_2 < \cdots < \hat{\theta}_p$, the ordered roots of the equation $|S - \hat{\theta}T| = 0$. The normal-theory procedures which have been proposed for testing H_0 are based on these roots. The likelihood-ratio test and Roy and Gnanadesikan's test, which we examine here, are consistent. Hotelling's test [10], which is based on the sum of the roots ($\operatorname{tr} ST^{-1}$), is not consistent against all alternatives in H_A; nor is a test based on a product of the roots ($|ST^{-1}|$) ($\operatorname{tr} \Sigma_1\Sigma_2^{-1} = p$ does not imply $\Sigma_1 = \Sigma_2$; nor does $|\Sigma_1\Sigma_2^{-1}| = 1$). That these functions of the roots are not asymptotically robust is demonstrated in Sub-section 3.1.

2.1. *The likelihood-ratio test for H_0.* The normal-theory likelihood-ratio criterion for testing H_0 is ([1] page 248)

$$\begin{aligned}\lambda &= |nS|^{n/2}|mT|^{m/2}|ns + mT|^{-(n+m)/2} \cdot (n+m)^{(n+m)p/2} n^{-np/2} m^{-mp/2} \\ &= |S|^{n/2}|T|^{m/2}|\alpha S + \beta T|^{-(n+m)/2} \\ &\qquad\qquad \text{(where } \alpha = n(n+m)^{-1},\ \beta = m(n+m)^{-1}\text{)} \\ &= \textstyle\prod_{i=1}^{p} \hat{\theta}_i^{n/2}(\alpha\hat{\theta}_i + \beta)^{-(n+m)/2}\,.\end{aligned}$$

THEOREM 2.1. *The asymptotic distribution of $-2\log\lambda$ under H_0, as n and m tend to infinity in constant ratio, is that of $\sum_{i=1}^{p(p+1)/2} c_i U_i$, where the c_i are functions of correlation coefficients and standardized fourth-order cumulants and the U_i are independent χ_1^2 variates.*

PROOF. (For notational simplicity, we prove the result for $p = 2$; the extension to general p is straightforward). As we saw in Section 1, we can take $\Sigma = I$ without loss of generality. Now

$$\begin{aligned}-2\log\lambda &= (n+m)(\log|\alpha S + \beta T| - \alpha\log|S| - \beta\log|T|) \\ &= (n+m)\textstyle\sum_{i=1}^{2}\log(\alpha\hat{\theta}_i + \beta) - n\sum_{i=1}^{2}\log\hat{\theta}_i\,.\end{aligned}$$

The function

$$(n+m)\log(n(n+m)^{-1}y + m(n+m)^{-1}) - n\log y\ (y > 0)$$

has a minimum of 0 at $y = 1$, so $-2\log\lambda$ has a minimum of 0 at $\hat{\theta}_1 = \hat{\theta}_2 = 1$. Since

$$\hat{\theta}_1 = \hat{\theta}_2 = 1 \Leftrightarrow S = T\,,$$

if follows that the first partials of $-2\log\lambda$ with respect to the sample moments vanish at $S = T = \Sigma$. Since

$$n^{\frac{1}{2}}((s, (\beta/\alpha)^{\frac{1}{2}}t) - (\mu, (\beta/\alpha)^{\frac{1}{2}}\mu)) \to_{\mathscr{L}} N\left(0, \begin{bmatrix}\Gamma & 0\\ 0 & \Gamma\end{bmatrix}\right),$$

we have by Theorem 1.3 (i) that

$$n^{\frac{1}{2}}((s, t) - (\mu, \mu)) \to_{\mathscr{L}} N\left(0, \begin{bmatrix} \Gamma & 0 \\ 0 & (\alpha/\beta)\Gamma \end{bmatrix}\right).$$

Now

$$\begin{aligned} -2 \log \lambda &= nf(s, t) \\ &= n((\log |\alpha S + \beta T|)/\alpha - \log |S| - (\beta/\alpha) \log |T|) . \end{aligned}$$

Since $f(\mu, \mu) = 0$, we have by Theorem 1.3 (ii) that $nf(s, t) \to_{\mathscr{L}} \frac{1}{2} X'BX$, where

$$X \sim N\left(0, \begin{bmatrix} \Gamma & 0 \\ 0 & (\alpha/\beta)\Gamma \end{bmatrix}\right)$$

and B is the 6×6 matrix of second partials of f evaluated at (μ, μ). We find that

$$B = \begin{bmatrix} D & -D \\ -D & D \end{bmatrix},$$

where $D = \beta E = \beta \operatorname{diag}(1, 1, 2)$. We thus have $\frac{1}{2} X'BX = Z'CZ$, where $Z \sim N(0, I)$ and

$$C = \frac{1}{2} \begin{bmatrix} \beta^{\frac{1}{2}}\Gamma^{\frac{1}{2}} & 0 \\ 0 & \alpha^{\frac{1}{2}}\Gamma^{\frac{1}{2}} \end{bmatrix} \begin{bmatrix} E & -E \\ -E & E \end{bmatrix} \begin{bmatrix} \beta^{\frac{1}{2}}\Gamma^{\frac{1}{2}} & 0 \\ 0 & \alpha^{\frac{1}{2}}\Gamma^{\frac{1}{2}} \end{bmatrix},$$

so that $\frac{1}{2} X'BX = \Sigma c_i U_i$, where the c_i are the eigenvalues of C and the U_i are independent χ_1^2 variates. The eigenvalues of C are those of

$$\frac{1}{2} \begin{bmatrix} \beta\Gamma & 0 \\ 0 & \alpha\Gamma \end{bmatrix} \begin{bmatrix} E & -E \\ -E & E \end{bmatrix} = \begin{bmatrix} \frac{1}{2}\beta\Gamma E & -\frac{1}{2}\beta\Gamma E \\ -\frac{1}{2}\alpha\Gamma E & \frac{1}{2}\alpha\Gamma E \end{bmatrix}.$$

Since

$$\begin{vmatrix} \frac{1}{2}\beta\Gamma E - cI & -\frac{1}{2}\beta\Gamma E \\ -\frac{1}{2}\alpha\Gamma E & \frac{1}{2}\alpha\Gamma E - cI \end{vmatrix} = \begin{vmatrix} -cI & 0 \\ -\frac{1}{2}\alpha\Gamma E & \frac{1}{2}\Gamma E - cI \end{vmatrix} = -c^3 \left| \frac{1}{2}\Gamma E - cI \right|,$$

we see that $c_i = 0$, $i = 1, 2, 3$, while c_4, c_5, c_6 are the eigenvalues of $\frac{1}{2}\Gamma E$, and are thus functions of ρ and the λ_{ij}. □

The non-robustness of the likelihood-ratio criterion is clearly illustrated by the following corollary.

COROLLARY 2.1. *Suppose F is the* cdf *of a bivariate distribution whose components are* i.i.d. *with common kurtosis γ. Then under H_0*

$$-2 \log \lambda \to_{\mathscr{L}} (1 + \tfrac{1}{2}\gamma)(U_1 + U_2) + U_3 ,$$

where the U_i are independent χ_1^2 variates.

PROOF. For this case we have $\Gamma = \operatorname{diag}(\gamma + 2, \gamma + 2, 1)$. Hence $\frac{1}{2}\Gamma E = \operatorname{diag}(1 + \frac{1}{2}\gamma, 1 + \frac{1}{2}\gamma, 1)$, so $c_4 = c_5 = 1 + \frac{1}{2}\gamma$, $c_6 = 1$. □

In the case of a bivariate normal distribution, the limiting distribution is χ_3^2, which we know independently from Wilks' theorem ([18] page 418).

It is at once evident that the normal-theory likelihood-ratio criterion is not asymptotically robust. If we had $\gamma > 0$, the critical constant obtained from the χ_3^2 distribution would be too small and the significance level greater than the nominal figure, while if $\gamma < 0$ the opposite situation would obtain.

2.2. *Roy and Gnanadesikan's test for H_0.* A consistent test for H_0 proposed by Roy and Gnanadesikan [16] for the normal case is: accept if $c_1 < \hat{\theta}_1 < \hat{\theta}_p < c_2$, where the c_i are suitably chosen critical constants. We demonstrate the non-robustness of this test by deriving the asymptotic distribution of $\hat{\theta}_1$ and $\hat{\theta}_2$ in the bivariate case.

THEOREM 2.2. *Suppose F is the* cdf *of a bivariate distribution. Then, under H_0,*

$$(nm/(n+m))^{\frac{1}{2}}((\hat{\theta}_1, \hat{\theta}_2) - (1, 1)) \to_{\mathscr{L}} (X_1 - (X_2^2 + X_3^2)^{\frac{1}{2}}, X_1 + (X_2^2 + X_3^2)^{\frac{1}{2}}) ,$$

where $(X_1, X_2, X_3) \sim N(0, C)$ and C is a function of the correlation coefficient and the standardized fourth-order cumulants.

PROOF. We have

$$\hat{\theta}_1, \hat{\theta}_2 = (s_{20}t_{02} + s_{02}t_{20} - 2s_{11}t_{11} \pm ((s_{20}t_{02} + s_{02}t_{20} - 2s_{11}t_{11})^2 - 4|S|\,|T|)^{\frac{1}{2}})/2|T| ,$$

or, after a little rearrangement,

$$\hat{\theta}_1, \hat{\theta}_2 = f_1(s, t) \pm ((f_2(s, t))^2 + 4f_3(s, t)f_4(s, t))^{\frac{1}{2}} ,$$

where

$$\begin{aligned} f_1(s, t) &= (s_{20}t_{02} + s_{02}t_{20} - 2s_{11}t_{11})/2|T| \\ f_2(s, t) &= (s_{02}t_{20} - s_{20}t_{02})/2|T| \\ f_3(s, t) &= (s_{02}t_{11} - t_{02}s_{11})/2|T| \\ f_4(s, t) &= (s_{20}t_{11} - t_{20}s_{11})/2|T| . \end{aligned}$$

We first find the joint asymptotic distribution of the $f_i(s, t)$. As in Sub-section 2.1, we take $\Sigma = I$, we have samples of size n and m and we let n and m go to infinity in constant ratio. The matrix of first partials of $f(s, t) = (f_1(s, t), \cdots, f_4(s, t))$ evaluated at (μ, μ) is

$$A = \frac{1}{2}\begin{bmatrix} A_1 & A_2 \\ -A_1 & -A_2 \end{bmatrix}$$

where

$$A_1 = \begin{bmatrix} 1 & -1 \\ 1 & 1 \\ 0 & 0 \end{bmatrix} \quad \text{and} \quad A_2 = \begin{bmatrix} 0 & 0 \\ 0 & 0 \\ -1 & -1 \end{bmatrix} .$$

By Theorem 1.3 (i) under H_0

$$n^{\frac{1}{2}}\beta^{\frac{1}{2}}(f(s, t) - f(\mu, \mu)) = n^{\frac{1}{2}}\beta^{\frac{1}{2}}(f(s, t) - (1, 0, 0, 0)) \to_{\mathscr{L}} N(0, B) ,$$

where

$$B = \beta A' \begin{bmatrix} \Gamma & 0 \\ 0 & (\alpha/\beta)\Gamma \end{bmatrix} A$$
$$= \frac{1}{4} \begin{bmatrix} A_1'\Gamma A_1 & A_1'\Gamma A_2 \\ A_2'\Gamma A_1 & A_2'\Gamma A_2 \end{bmatrix},$$

which is a function of ρ and the λ_{ij}. We see that all the elements of $A_2'\Gamma A_2$ are equal, which means that $n^{\frac{1}{2}}\beta^{\frac{1}{2}}f_3$ and $n^{\frac{1}{2}}\beta^{\frac{1}{2}}f_4$ are asymptotically equal. Now $(nm/(n+m))^{\frac{1}{2}}((\hat{\theta}_1, \hat{\theta}_2) - (1, 1)) = n^{\frac{1}{2}}\beta^{\frac{1}{2}}(\hat{\theta}_1 - 1, \hat{\theta}_2 - 1)$ is a continuous function of $n^{\frac{1}{2}}\beta^{\frac{1}{2}}(f(s, t) - (\mu, \mu))$, so by Theorem 1.1

$$(nm/(n+m))^{\frac{1}{2}}((\hat{\theta}_1, \hat{\theta}_2) - (1, 1)) \to_{\mathscr{L}} (X_1 - (X_2^2 + X_3^2)^{\frac{1}{2}}, X_1 + (X_2^2 + X_3^2)^{\frac{1}{2}}),$$

where $(X_1, X_2, X_3) \sim N(0, C)$, and

$$C = \begin{bmatrix} b_{11} & b_{12} & 2b_{13} \\ b_{12} & b_{22} & 2b_{23} \\ 2b_{13} & 2b_{23} & 4b_{33} \end{bmatrix}. \qquad \square$$

COROLLARY 2.2. *Suppose F is the* cdf *of a bivariate distribution whose components are* i.i.d. *with common kurtosis γ. Then under H_0*

$$(nm/(n+m))^{\frac{1}{2}}((\hat{\theta}_1, \hat{\theta}_2) - (1, 1))$$
$$\to_{\mathscr{L}} ((1 + \tfrac{1}{2}\gamma)^{\frac{1}{2}}Z_1 - ((1 + \tfrac{1}{2}\gamma)Z_2^2 + Z_3^2)^{\frac{1}{2}}, (1 + \tfrac{1}{2}\gamma)^{\frac{1}{2}}Z_1 + ((1 + \tfrac{1}{2}\gamma)Z_2^2 + Z_3^2)^{\frac{1}{2}}),$$

where $(Z_1, Z_2, Z_3) \sim N(0, I)$.

PROOF. We have $\Gamma = \operatorname{diag}(\gamma + 2, \gamma + 2, 1)$, so that

$$B = \begin{bmatrix} B_1 & 0 \\ 0 & B_2 \end{bmatrix},$$

where

$$B_1 = \operatorname{diag}\left(1 + \frac{1}{2}\gamma, 1 + \frac{1}{2}\gamma\right) \quad \text{and} \quad B_2 = \begin{bmatrix} \frac{1}{4} & \frac{1}{4} \\ \frac{1}{4} & \frac{1}{4} \end{bmatrix}. \qquad \square$$

In the normal case, $\gamma = 0$, and $(nm/(n+m))^{\frac{1}{2}}(\hat{\theta}_i - 1) \to_{\mathscr{L}} U \pm V$, where $U \sim N(0, 1)$, $V \sim (\chi_2^2)^{\frac{1}{2}}$ and U and V are independent. The joint density of $Y_1 = U + V$ and $Y_2 = U - V$ is

$$f_{Y_1, Y_2}(y_1, y_2) = 2^{-5/2}\pi^{-\frac{1}{2}}(y_1 - y_2)\exp(-2^{-3}(y_1^2 + y_2^2)), \qquad y_1 > y_2 > 0$$
$$= 0, \text{ otherwise},$$

which agrees with the density derived by another method—see Gleser [8] page 162.

It is clear that the test is not robust. If $\gamma > 0$, the critical points derived from the normal theory are too small in absolute value, and the significance level is greater than the nominal value, while if $\gamma < 0$, the opposite situation obtains.

The method used above cannot be applied to the case of general p-dimensional distributions because of the impossibility of representing the $\hat{\theta}_i$ as explicit functions of the second-order sample moments. However, the calculations for the bivariate case suffice to demonstrate the non-robustness of the test.

3. Large-sample tests for equality of covariance matrices. In this section we describe several tests for H_0 which have the property of asymptotic robustness. In seeking a large sample test for $\Sigma_1 = \Sigma_2$ we might first look for a univariate function of the second-order sample moments which is asymptotically normal. We need, in order to have a test which is consistent against any alternative, a function which assumes a certain value in the null hypothesis (i.e. $S = T = \Sigma$) and which assumes that value nowhere else. But to have asymptotic normality, we want the function to have a linear approximation at the null hypothesis (i.e., to have at least one non-vanishing first partial there). The question is, can these requirements be simultaneously met? They can, for example, in the univariate case, since the function s^2/t^2 satisfies them. It is not clear that it is possible to do so in higher dimensions. In the bivariate case, for example, we want a differentiable real-valued function defined on E^6, which has a constant value on a 3-dimensional subspace (a hyperplane through the origin), does not assume that value off the hyperplane and does not have an extremum on the hyperplane. We have seen that the normal-theory likelihood-ratio criterion fails the third requirement; so does another obvious choice, $\Sigma(s_{ij} - t_{ij})^2$. On the other hand, the function $|S|/|T|$ fails the second requirement, since $|S| = |T| \nRightarrow S = T$.

If we are interested in testing only against a restricted class of alternatives —a subclass of H_A—there may be a function which fails the second requirement but nevertheless yields a test consistent against the subclass. In general, however, we want a test consistent against all alternatives, so we need to find some approach other than that of an asymptotically normal univariate test statistic. The four large-sample tests discussed below are consistent; the last three have a common basic idea, that of exploiting the asymptotic normality of the second-order moments by treating the problem as a test for equality of mean vectors.

3.1. *A test based on the elementary symmetric functions of the roots of* ST^{-1}. Suppose we wish to test H_0 (see Section 2), where we have two p-dimensional populations. Now

$$\Sigma_1 = \Sigma_2 \Leftrightarrow \text{ all roots of } \Sigma_1\Sigma_2^{-1} \text{ equal } 1 \Leftrightarrow e_k = \binom{p}{k}, \qquad k = 1, \cdots, p,$$

where e_k is the kth elementary symmetric function of the roots. Hence we might consider using the vector $(\hat{e}_1, \cdots, \hat{e}_p)$ as a test criterion for H_0, where $\hat{e}_k$ is the kth elementary symmetric function of the roots of ST^{-1}.

In the bivariate case, with both samples of size n, we find using Theorem 1.3 (i) that under H_0

$$n^{\frac{1}{2}}(\Sigma\hat{\theta}_i - 2, \prod \hat{\theta}_i - 1) \to_{\mathscr{L}} N\left(0, \begin{bmatrix} \sigma^2 & \sigma^2 \\ \sigma^2 & \sigma^2 \end{bmatrix}\right),$$

where

$$\sigma^2 = 2(\lambda_{40} + \lambda_{04} - 4\rho(\lambda_{31} + \lambda_{13}) + 2(1 - \rho^2)\lambda_{22})(1 - \rho^2)^{-2} + 8 .$$

(The asymptotic distribution has rank 1 under H_0, but this is not necessarily the case under H_A.)

Since if θ_1 and θ_2 are not equal to 1 we must have $\Sigma\theta_i \neq 2$ or $\prod \theta_i \neq 1$, for large samples an intuitively reasonable test procedure is: reject if

$$(n^{\frac{1}{2}}(\Sigma\hat{\theta}_i - 2)\hat{\sigma}^{-1}, n^{\frac{1}{2}}(\prod \hat{\theta}_i - 1)\hat{\sigma}^{-1})$$

lies outside the square with center at (0, 0) and side Δ, where Δ is obtained from the standard normal distribution function. Here $\hat{\sigma}$ is a consistent estimate of σ obtained from both samples.

It can be shown that the asymptotic distribution of the $\hat{e}_i$ is also of rank 1 in the p-variate case. Hence the same idea is theoretically applicable to a test of H_0 for any value of p. But an explicit representation of the asymptotic covariance matrix, such as we gave above in the bivariate case, would be very tedious to compute for a particular $p > 2$. As regards the practical worth of the test, the fact that the limiting distribution of the statistic has rank 1 leads one to suspect that the convergence may be rather slow. This matter could be investigated by Monte Carlo methods; a more immediate objection is the difficulty of deriving the asymptotic distribution for the case of more than two dimensions. The author suggests that one of the three tests next to be discussed would be a better choice for practical purposes.

3.2. *A "standard error" test.* Since the second-order sample moments are asymptotically multivariate normal, a possible approach to the problem of testing for equality of covariance matrices is to treat it as a test for equality of mean vectors. One way to do this is simply to normalize the difference vector by means of an estimate of its covariance matrix, and to use the fact that the inner product of the resulting vector is asymptotically χ^2 distributed. However, we may hope to hasten the convergence to normality by using transformations of the second-order sample moments. The log transformation of the sample variance is generally supposed to have beneficial effects in this direction. Further, the $\tanh^{-1}$ transformation of $\hat{\rho}$, the sample correlation coefficient, is well known to have useful properties in the normal case. Gayen [6] considered the distribution of $\tanh^{-1}\hat{\rho}$ in samples from non-normal populations specified by an Edgeworth expansion. He found that for samples of moderate size, the distribution of $\tanh^{-1}\hat{\rho}$ is approximately normal, though

the approach to normality is not in general as rapid as in the normal case.

Consequently, we propose using as a test statistic the vector $\varphi(s) - \varphi(t)$, where $\varphi(s)' = (\log s_{20}, \log s_{02}, \tanh^{-1}\hat{\rho}_1)$ and $\varphi(t)' = (\log t_{20}, \log t_{02}, \tanh^{-1}\hat{\rho}_2)$, normalized by an estimate of its covariance matrix (here $\hat{\rho}_1 = s_{11}(s_{20}s_{02})^{-\frac{1}{2}}$, $\hat{\rho}_2 = t_{11}(t_{20}t_{02})^{-\frac{1}{2}}$).

THEOREM 3.1. *Suppose we have samples of size n and m respectively from two bivariate populations with finite fourth moments. Then under H_0 (see Section 2), as n and m tend to infinity in constant ratio,*

$$(\varphi(s) - \varphi(t))'(n^{-1}\hat{\Gamma}_1^* + m^{-1}\hat{\Gamma}_2^*)^{-1}(\varphi(s) - \varphi(t)) \to_{\mathscr{L}} \chi_3^2 ,$$

where $\hat{\Gamma}_1^*$ is a consistent estimate of the asymptotic covariance matrix of $\varphi(s)$ obtained by substituting sample quantities for population moments, and $\hat{\Gamma}_2^*$ is similarly defined.

PROOF. By Theorem 1.3(i) we have that under H_0, as $n \to \infty$,

$$n^{\frac{1}{2}}(\varphi(s) - \varphi(\mu)) \to_{\mathscr{L}} N(0, A'\Gamma A) ,$$

where Γ is the asymptotic covariance matrix (1.1) of s and A is the matrix of first partials of φ evaluated at Σ:

$$A = \begin{bmatrix} \mu_{20}^{-1} & 0 & -\rho/2\mu_{20}(1-\rho^2) \\ 0 & \mu_{02}^{-1} & -\rho/2\mu_{02}(1-\rho^2) \\ 0 & 0 & 1/(\mu_{20}\mu_{02})^{\frac{1}{2}}(1-\rho^2) \end{bmatrix} .$$

(Since

$$A = \operatorname{diag}(\mu_{20}^{-1}, \mu_{02}^{-1}, (\mu_{20}\mu_{02})^{-\frac{1}{2}}) \begin{bmatrix} 1 & 0 & -\rho/2(1-\rho^2) \\ 0 & 1 & -\rho/2(1-\rho^2) \\ 0 & 0 & 1/(1-\rho^2) \end{bmatrix} ,$$

the matrix $A'\Gamma A$ depends only on ρ and the λ_{ij}—cf. Section 1). Similarly, under H_0,

$$m^{\frac{1}{2}}(\varphi(t) - \varphi(\mu)) \to_{\mathscr{L}} N(0, A'\Gamma A) .$$

Hence, letting $A'\Gamma A = \Gamma^*$, we see by Theorem 1.1 that

$$n^{\frac{1}{2}}(\varphi(s) - \varphi(t)) \to_{\mathscr{L}} N(0, (1 + n/m)\Gamma^*) .$$

An application of Theorem 1.2 completes the proof. □

The extension to the p-variate case is immediate, with χ_3^2 replaced by $\chi^2_{p(p+1)/2}$.

3.3. *A generalization of Box's test.* The following test is a generalization of that proposed by Box [4] for the equality of variances. Split each sample randomly into groups, each of size $k \geqq 2$ (we assume n and m are divisible by k). Compute $\varphi(s^{(j)})$, $j = 1, \cdots, n/k$, where $s^{(j)}$ is the vector of second-order moments computed from the jth group of the first sample. Compute the same quantities for the m/k groups of the second sample. The $n' = n/k$ vectors computed from the first sample, and the $m' = m/k$ vectors from the second

sample, have approximately the multivariate normal distribution with equal covariance matrices under H_0. That being so, we may take the statistic

$$U = n'm'(n' + m')^{-1}\bar{X}'\hat{\Lambda}^{-1}\bar{X}$$

to have approximately Hotelling's T^2 distribution with 3 and $n' + m' - 2$ df. Here $\bar{X} = n'^{-1}\sum_{j=1}^{n'}\varphi(s^{(j)}) - m'^{-1}\sum_{j=1}^{m'}\varphi(t^{(j)})$ and $\hat{\Lambda}$ is the pooled estimate (on $n' + m' - 2$ df) of the common covariance matrix. (Of course, under H_0, $U \to_{\mathscr{L}} \chi_3^2$ as n', m' tend to infinity.) It may be preferable in practice to use a non-pooled estimate of the covariance matrix of $\bar{X}$—cf. Sub-section 3.2.

In the p-variate case, for an analogously defined statistic U, we have $U \to_{\mathscr{L}} \chi^2_{p(p+1)/2}$.

3.4. *A test using the jackknife.* A discussion of the jackknife and its applicability to testing problems is contained, for example, in Miller [11]. The following theorem, which is an extension of a theorem in Miller [12], provides the justification for the use of the jackknife in the problem we are considering. It is proved for a group size of 1, but is valid for larger group sizes.

Suppose we have a sample of size n from a bivariate distribution with covariance matrix

$$\Sigma = \begin{bmatrix} \mu_{20} & \mu_{11} \\ \mu_{11} & \mu_{02} \end{bmatrix}$$

and finite fourth moments. Let $g: E^3 \to E^k$, $1 \leqq k \leqq 3$, be a function whose second partial derivatives are bounded in an interval about μ. Let $g(s) = (g_1(s), \cdots, g_k(s))'$, and denote the jackknife estimator of $g_j(\mu)$ by

$$\hat{g}_j(\mu) = ng_j(s) - (n-1)n^{-1}\sum_{i=1}^{n} g_j(s_{(i)}),$$

and the ith pseudo-value by

$$\hat{g}_j{}^i(\mu) = ng_j(s) - (n-1)g_j(s_{(i)}),$$

where $s_{(i)}$ is the vector of second-order moments computed from the $n-1$ observations obtained by excluding the ith. Let

$$\hat{g}(\mu) = (\hat{g}_1(\mu), \cdots, \hat{g}_k(\mu))',$$

and let

$$\hat{g}^i(\mu) = (\hat{g}_1{}^i(\mu), \cdots, \hat{g}_k{}^i(\mu))'.$$

THEOREM 3.2. (i) $n^{\frac{1}{2}}(\hat{g}(\mu) - g(\mu)) \to_{\mathscr{L}} N(0, A'\Gamma A)$, *where* Γ *is the asymptotic covariance matrix* (1.1) *of* s, *and* A *is the matrix of first partials of* g *evaluated at* Σ.

(ii) $(n-1)^{-1}\sum_{i=1}^{n}(\hat{g}^i(\mu) - \hat{g}(\mu))(\hat{g}^i(\mu) - \hat{g}(\mu))' \to_{\mathscr{L}} A'\Gamma A$.

PROOF. (i) Since

$$n^{\frac{1}{2}}(\hat{g}(\mu) - g(\mu)) = n^{\frac{1}{2}}(g(s) - g(\mu)) - (n-1)n^{-1}n^{\frac{1}{2}}\sum_{i=1}^{n}(g(s_{(i)}) - g(s)),$$

and $n^{\frac{1}{2}}(g(s) - g(\mu)) \to_{\mathscr{L}} N(0, A'\Gamma A)$ by Theorem 1.3 (i), we need only show

that $n^{\frac{1}{2}} \sum_{i=1}^{n} (g(s_{(i)}) - g(s))$ converges to zero in probability. The fourth and lower order sample moments converge in probability to the corresponding population moments. As a consequence of a theorem of Pratt [14], it suffices to show that if $\{(x_i, y_i)\}$ is a sequence of (nonrandom) vectors such that the sequences $\{n^{-1} \sum_{i=1}^{n} |x_i|\}$, $\{n^{-1} \sum_{i=1}^{n} x_i^2\}$, $\{n^{-1} \sum_{i=1}^{n} x_i^4\}$, etc., converge, then $n^{\frac{1}{2}} \sum_{i=1}^{n} (g_j(s_{(i)}) - g_j(s)) \to 0$, where $g_j(s_{(i)})$ and $g_j(s)$ are computed from the first n terms of the sequence $\{(x_i, y_i)\}$ according to the formulas given earlier. We first note that $\max_{1 \leq k \leq n} \{(x_k - \bar{x})^2/n\} \to 0$, which implies $s_{20(i)} \to \mu_{20}$ uniformly in i, since

$$s_{20(i)} = s_{20} - n(n-1)^{-1}(n-2)^{-1}((x_i - \bar{x})^2 - n^{-1} \sum_{j=1}^{n} (x_j - \bar{x})^2) .$$

Similarly, we see that $s_{02(i)} \to \mu_{02}$ and $s_{11(i)} \to \mu_{11}$, uniformly in i. Now by Taylor's theorem,

$$g_j(s_{(i)}) - g_j(s) = (s_{(i)} - s)'A + \tfrac{1}{2}(s_{(i)} - s)'B_i(s_{(i)} - s) ,$$

where A is the vector of first partials of g_j evaluated at s, and B_i is the symmetric matrix of second partials of g_j evaluated at ζ_i, a point on the line segment joining $s_{(i)}$ and s. We have

$$n^{\frac{1}{2}} \sum_{i=1}^{n} (g_j(s_{(i)}) - g_j(s)) = \tfrac{1}{2} n^{5/2}(n-1)^{-2}(n-2)^{-2} \sum_{i=1}^{n} z_i' B_i z_i$$

(where, for example, $z_{i1} = (x_i - \bar{x})^2 - n^{-1} \sum_{j=1}^{n} (x_j - \bar{x})^2$). Let us consider the expression $(n-1)^{-1} \sum_{i=1}^{n} b_{i;11} z_{i1}^2$ (a similar argument applies to the other terms). Now $\zeta_i \to \mu$ uniformly in i, since $s_{(i)} \to \mu$ uniformly in i. By assumption, the second partials are bounded in a neighborhood of μ, so for n sufficiently large the $b_{i;11}$ are bounded. Also we have $(n-1)^{-1} \sum_{i=1}^{n} z_{i1}^2 \to \mu_{40} - \mu_{20}^2$, so for n sufficiently large $(n-1)^{-1} \sum_{i=1}^{n} b_{i;11} z_{i1}^2$ is bounded. Hence

$$n^{5/2}(n-1)^{-2}(n-2)^{-2} \sum_{i=1}^{n} b_{i;11} z_{i1}^2 \to 0 .$$

(ii)

$$\begin{aligned}
&(n-1)^{-1} \sum_{i=1}^{n} (\hat{g}^i(\mu) - \hat{g}(\mu))(\hat{g}^i(\mu) - \hat{g}(\mu))' \\
&\quad = (n-1)^{-1} \sum (-(n-1)g(s_{(i)}) \\
&\qquad + (n-1)n^{-1} \sum_{j=1}^{n} g(s_{(j)}) - n(n-2)^{-1}A'Z_i + n(n-2)^{-1}A'Z_i) \\
&\qquad \times (-(n-1)g(s_{(i)}) + (n-1)n^{-1} \sum_{j=1}^{n} g(s_{(j)}) \\
&\qquad - n(n-2)^{-1}A'Z_i + n(n-2)^{-1}A'Z_i)' ,
\end{aligned}$$

where $Z_{i1} = (X_i - \bar{X})^2 - n^{-1} \sum_{j=1}^{n} (X_j - \bar{X})^2$, etc. It is easy to verify that $n^3(n-2)^{-2}A'((n-1)^{-1} \sum_{i=1}^{n} Z_i Z_i')A \to_P A'\Gamma A$, so we need only show that the elements of the remaining terms converge to zero in probability. We illustrate the method, again using Pratt's theorem. Since $(g(s_{(i)} - g(s))' = (s_{(i)} - s)'C_i$, where C_i is the matrix of first partials of g evaluated at ξ_i, a point on the line

segment joining $s_{(i)}$ and s, we have, recalling that

$$s_{(i)} - s = -n(n-1)^{-1}(n-2)^{-1}z_i ,$$

$$(n-1)^{-1}\sum_{i=1}^{n}(-(n-1)g(s_{(i)}) + (n-1)n^{-1}\sum_{j=1}^{n} g(s_{(j)}) - n(n-2)^{-1}A'z_i)$$
$$\times(-(n-1)g(s_{(i)}) + (n-1)n^{-1}\sum_{j=1}^{n} g(s_{(j)}) - n(n-2)^{-1}A'z_i)'$$
$$= n^2(n-2)^{-2}(n-1)^{-1}\sum_{i=1}^{n}((C_i - A)'z_i - n^{-1}\sum_{j=1}^{n}(C_j - A)'z_j)$$
$$\times((C_i - A)'z_i - n^{-1}\sum_{j=1}^{n}(C_j - A)'z_j)' .$$

Since $s_{(i)} \to \mu$ uniformly in i, and the first partials are continuous in a neighborhood of μ, we have $C_i \to A$ uniformly in i. Consider $d_{n,\alpha\beta}$, the $\alpha\beta$ element of $D_n = (n-1)^{-1}\sum_{i=1}^{n}(C_i - A)'z_i z_i'(C_i - A)$. For $\varepsilon > 0$ and n sufficiently large

$$|d_{n,\alpha\beta}| < \varepsilon^2 \sum_{j=1}^{3}\sum_{k=1}^{3}(n-1)^{-1}\sum_{i=1}^{n}|z_{ij}z_{ik}| .$$

Since $(n-1)^{-1}\sum_{i=1}^{n}|z_{ij}z_{ik}|$ converges by assumption, and ε is arbitrary, we have $d_{n,\alpha\beta} \to 0$. A similar analysis shows that the $\alpha\beta$ elements of all other matrix terms go to zero. □

Since s is a multivariate U-statistic, this theorem could also be proved using an extension of a theorem of Arvesen [2].

We now apply the theorem to the function

$$\varphi(\mu) = (\log \mu_{20}, \log \mu_{02}, \tanh^{-1}(\mu_{11}(\mu_{20}\mu_{02})^{-\frac{1}{2}}))' .$$

Suppose we have two samples of size n and m from bivariate populations, and we wish to test H_0. Let

$$\hat{\Gamma}_1^* = (n-1)^{-1}\sum_{i=1}^{n}(\hat{\varphi}_1^{\,i}(\mu) - \hat{\varphi}_1(\mu))(\hat{\varphi}_1^{\,i}(\mu) - \hat{\varphi}_1(\mu))'$$

and

$$\hat{\Gamma}_2^* = (m-1)^{-1}\sum_{i=1}^{m}(\hat{\varphi}_2^{\,i}(\mu) - \hat{\varphi}_2(\mu))(\hat{\varphi}_2^{\,i}(\mu) - \hat{\varphi}_2(\mu))' ,$$

where $\hat{\varphi}_j^{\,i}(\mu)$ and $\hat{\varphi}_j(\mu)$ are calculated from the jth sample. Then, as $n, m \to \infty$ in constant ratio,

$$(\hat{\varphi}_1(\mu) - \hat{\varphi}_2(\mu))'(n^{-1}\hat{\Gamma}_1^* + m^{-1}\hat{\Gamma}_2^*)^{-1}(\hat{\varphi}_1(\mu) - \hat{\varphi}_2(\mu)) \to_{\mathscr{L}} \chi_3^2$$

(cf. Sub-section 3.2).

The theorem extends in an obvious fashion to p-variate distributions, with asymptotic degrees of freedom $p(p+1)/2$.

4. Asymptotic comparisons.

4.1. *Pitman efficiency.* Though Pitman's [13] criterion of asymptotic relative efficiency is commonly applied to test statistics which have limiting normal distributions, it can also be applied to cases where the limiting distributions of the two test statistics are the same, but not normal, provided the two power functions can be made asymptotically equal by an appropriate choice of sample sizes. The method used here follows Hannan [9].

Suppose we have samples of size n from two bivariate populations with cdf's F and G respectively. Let $\{\Psi_\theta\}$ be a family of cdf's indexed by a real parameter θ. Consider the hypothesis $H_0: F = G = \Psi_{\theta_0}$ versus the alternative $H_1: F = \Psi_{\theta_0}$, $G = \Psi_\theta$, $\theta \neq \theta_0$, and consider the sequence of alternatives given by $\theta_n = \theta_0 + n^{-\frac{1}{2}}\delta$. Let $T_n' = (T_{1n}, T_{2n}, T_{3n})$ be a statistic computed from the two samples, and let

$$E_\theta T_n = \mu_n(\theta)\,, \qquad \mathrm{Cov}_\theta\, T_n = \Sigma_n(\theta)\,, \qquad \mathrm{Cov}_\theta(T_{in}, T_{jn}) = \sigma_{in}(\theta)\sigma_{jn}(\theta)\rho_{ijn}(\theta)\,,$$

and

$$c_i = \lim_{n\to\infty}\left(\frac{d\mu_{in}(\theta)}{d\theta}\Big|_{\theta=\theta_0}\right)\Big/ n^{\frac{1}{2}}\sigma_{in}(\theta_0)\,.$$

Then assuming that T_n is asymptotically multivariate normal under θ_0 and under θ_n, and that certain regularity conditions are satisfied, the quantity $nc'R^{-1}(\theta_0)c$ (where $c' = (c_1\, c_2,\, c_3)$, $R(\theta_0) = (\rho_{ij}(\theta_0))$ and $\rho_{ij}(\theta_0) = \lim_{n\to\infty}\rho_{ijn}(\theta_0)$) is the limiting Pitman efficacy for a test based on the statistic $(T_n - \mu_n(\theta_0))'\times \Sigma_n^{-1}(\theta_0)(T_n - \mu_n(\theta_0))$. If we have two sequences of test statistics $T_n^{(1)}$ and $T_n^{(2)}$ satisfying the regularity conditions, the ratio

$$c_1'R_1^{-1}(\theta_0)c_1/c_2'R_2^{-1}(\theta_0)c_2$$

is the Pitman relative efficiency of $T^{(1)}$ with respect to $T^{(2)}$.

Now suppose Ψ_θ is the $N(0, \theta B)$ cdf, where

$$B = \begin{bmatrix}\mu_{20} & \mu_{11}\\ \mu_{11} & \mu_{02}\end{bmatrix},$$

and let $\theta_0 = 1$. Let $T_n^{(1)} = \varphi(s) - \varphi(t)$. Then $\mu_n^{(1)}(\theta) = (-\log\theta,\, -\log\theta,\, 0)'$, and the statistic of Sub-section 3.2, $nT_n^{(1)\prime}(\hat{\Gamma}_1^* + \hat{\Gamma}_2^*)^{-1}T_n^{(1)}$, is asymptotically equivalent to the statistic $T_n^{(1)}\Sigma_n^{(1)-1}(1)T_n^{(1)}$. We assume that the necessary regularity conditions hold and that $n\Sigma_n^{(1)}(1) \to 2\Gamma^*$, where $2\Gamma^*$ is the asymptotic covariance matrix of $n^{\frac{1}{2}}T_n^{(1)}$ under H_0, i.e.

$$2\Gamma^* = \begin{bmatrix}4 & 4\rho^2 & 2\rho\\ & 4 & 2\rho\\ & & 2\end{bmatrix}.$$

We find $c_{11} = c_{12} = -\frac{1}{2}$, $c_{13} = 0$ and

$$R_1^{-1}(1) = \begin{bmatrix}1 + \frac{1}{2}\rho^2(1-\rho^2)^{-1} & -\frac{1}{2}\rho^2(1-\rho^2)^{-1} & -2^{-\frac{1}{2}}\rho\\ & 1 + \frac{1}{2}\rho^2(1-\rho^2)^{-1} & -2^{-\frac{1}{2}}\rho\\ & & 1+\rho^2\end{bmatrix},$$

so that $c_1'R_1^{-1}(1)c_1 = \frac{1}{2}$. If we let $T_n^{(2)} = \hat{\varphi}_1(\mu) - \hat{\varphi}_2(\mu)$ (see Sub-section 3.4), we similarly find that $c_2'R_2^{-1}c_2 = \frac{1}{2}$, so that the Pitman relative efficiency of the jackknife test to the standard error test is unity. This statement is true

for any sequence of alternatives for which the regularity conditions hold—the two procedures are asymptotically equivalent. Let $T_n^{(3)}$ be the statistic $\bar{X}$ of the Box test (Sub-section 3.3). We have $\mu_n^{(3)}(\theta) = (-\log\theta, -\log\theta, 0)'$, and the asymptotic ($n' \to \infty$, k fixed) covariance matrix of $(n')^{\frac{1}{2}}T_n^{(3)}$ under H_0 is 2Δ, where Δ is the covariance matrix of $\varphi(s^{(j)})$. Now the covariance matrix of $s^{(j)}$ under H_0 is

$$A'\begin{bmatrix} 2/(k-1) & \rho^2(2/(k-1)+1/k(k-1)) & \rho(2/(k-1)-1/k(k-1)) \\ & 2/(k-1) & \rho(2/(k-1)-1/k(k-1)) \\ & & (1+\rho^2)/(k-1) \end{bmatrix} A,$$

where $A = \operatorname{diag}(\mu_{20}, \mu_{02}, (\mu_{20}\mu_{02})^{\frac{1}{2}})$. Expanding $\varphi(s^{(j)})$ about $\mu = (\mu_{20}, \mu_{02}, \mu_{11})'$, and ignoring terms of order k^{-2}, we find that $\Delta \approx (k-1)^{-1}\Gamma^*$. Hence $c_3 \approx (k/(k-1))^{\frac{1}{2}}c_1$, and since $R_3^{-1}(1) = R_1^{-1}(1)$, we have $c_3'R_3^{-1}(1)c_3 \approx (k-1)/2k$, and the approximate Pitman relative efficiency of the Box test to the standard error test is $(k-1)/k$. This result reflects the loss of information incurred in dividing the data into groups.

If Ψ_θ is the cdf of a nonnormal bivariate distribution with covariance matrix

$$\theta\begin{bmatrix} \mu_{20} & \mu_{11} \\ \mu_{11} & \mu_{02} \end{bmatrix}$$

and finite fourth moments, the efficacies are not simple expressions, and the relative efficiency will depend on ρ and the λ_{ij}. However if the distribution has i.i.d. components with common kurtosis γ, we readily find that the approximate Pitman relative effiiciency of the Box test to the standard error test is $(2+\gamma)/(2k/(k-1)+\gamma)$.

The meaning of asymptotic relative efficiency in a multivariate situation is somewhat clouded by the fact that one can define different sequences of alternatives converging at the same rate to the same null hypothesis but yielding different relative efficiencies. For example, let Ψ_θ be the cdf of a bivariate distribution with covariance matrix

$$\begin{bmatrix} 1+\theta & \alpha\theta \\ \alpha\theta & 1+\theta \end{bmatrix}$$

and finite fourth moments. Let $\theta_0 = 0$, and suppose Ψ_0 corresponds to i.i.d. components with common kurtosis γ. We find that the approximate Pitman relative efficiency of the Box test to the standard error test is

$$((2k/(k-1)+\gamma)^{-1} + \alpha^2(k-1)/2k)/((2+\gamma)^{-1} + \tfrac{1}{2}\alpha^2),$$

which corresponds to the expression given in the previous paragraph only if $\alpha = 0$.

4.2. *Approximate Bahadur efficiency.* Suppose $\{T_n^{(i)}\}$, $i = 1, 2$, are sequences of test statistics, and that the test rule is to reject H_0 if $T_n^{(i)}$ is "large." Also

suppose that under H_0, $F_{T_n^{(i)}} \to F^{(i)}$, where $F_{T_n^{(i)}}$ is the cdf of $T_n^{(i)}$. If x is the observed outcome of the experiment, the approximate "attained level" of the test based on $T_n^{(i)}$ is $L_n^{(i)}(x) = 1 - F^{(i)}(T_n^{(i)}(x))$. The experimenter will reject H_0 for "small" values of $L_n^{(i)}(x)$, so if an alternative $\omega \in H_A$ is true, $T_n^{(i)}$ is better than $T_n^{(2)}$ for given x if $L_n^{(1)}(x) < L_n^{(2)}(x)$. If $\operatorname{plim}(\log L_n^{(1)}(X)/\log L_n^{(2)}(X))$ is a constant $\varphi_{1,2}(\omega)$, this quantity is a possible measure of asymptotic relative efficiency at ω, with $\varphi_{1,2} > 1$ indicating that $\{T_n^{(1)}\}$ is the superior sequence of test statistics. The measure $\varphi_{1,2}$, known as approximate Bahadur efficiency, was introduced by Bahadur [3] and generalized by Gleser (1964); they give sufficient conditions for $\operatorname{plim}(\log L_n^{(1)}(X)/\log L_n^{(2)}(X))$ to be a constant. They also point out that there is a serious difficulty in the interpretation of $\varphi_{1,2}$ as a measure of asymptotic relative efficiency; that is, it is possible for two equivalent test statistics not to have a relative efficiency of 1 everywhere in H_A. This difficulty would not arise if instead of $L_n^{(i)}(x)$ we used the exact attained levels, $1 - F_{T_n^{(i)}}(T_n^{(i)}(x))$, but to do so may be difficult or impossible. The justification for considering approximate Bahadur efficiency is that it is easy to compute, may be the only measure readily available, and in some cases can be shown to be a good approximation to the measure based on exact levels, at least for alternatives close to H_0.

Briefly, the sufficient conditions of Bahadur and Gleser in the present context are as follows: Let ω index the family consisting of pairs of bivariate populations having finite fourth moments. Let $\{T_n^{(1)}\}$, $\{T_n^{(2)}\}$ be two sequences of real-valued test statistics such that

(i) $F_{T_n^{(i)}}(y) \to F^{(i)}(y)$, $\forall \omega \in H_0$ (see Section 2), where $2\log(1 - F^{(i)}(y)) = -a^i y^{r_i}(1 + o(1))$ as $y \to \infty$, $a_i > 0$, $r_i > 0$;

(ii) $\operatorname{plim}(n^{-t_i} T_n^{(i)}) = h_i(\omega)$, $0 < h_i(\omega) < \infty$, $\forall \omega \in H_A$;

(iii) $t_1 r_1 = t_2 r_2$.

Under these conditions $\varphi_{1,2}(\omega) = a_1(h_1(\omega))^{r_1}/a_2(h_2(\omega))^{r_2}$. The quantity $a_i(h_i(\omega))^{r_i}$ is called the slope of the sequence $\{T_n^{(i)}\}$ at ω.

We note that if F is the χ_k^2 cdf for any k, then F satisfies the condition of (i) above with $a = 1$, $r = 1$ [3]. With this information we can compute some relative efficiencies for the tests considered in previous sections.

First consider the normal-theory likelihood-ratio criterion for H_0, when both samples are of size n: $-2\log\lambda = -n\log(\prod_{i=1}^{2} 4\hat{\theta}_i(\hat{\theta}_i + 1)^{-2})$. Consider the alternative

$$\Sigma_1 = \begin{bmatrix} \mu_{20} & \mu_{11} \\ \mu_{11} & \mu_{02} \end{bmatrix}, \quad \Sigma_2 = \eta\Sigma_1, \qquad \eta > 0.$$

The roots of $|\Sigma_1\Sigma_2^{-1} - \theta I| = 0$ are $\theta_1 = \theta_2 = \eta^{-1}$, and $\operatorname{plim}((-2\log\lambda)/n) = 2\log((1+\eta)^2/4\eta)$, so condition (ii) is satisfied with $t = 1$. Since under H_0 and the normality assumption $-2\log\lambda$ has a limiting χ^2 distribution, condition

(i) is satisfied with $a = 1$, $r = 1$, and the slope of the likelihood-ratio test at the given alternative is $2\log((1+\eta)^2/4\eta)$. The standard error test statistic is $n(\varphi(s) - \varphi(t))'(\hat{\Gamma}_1^* + \hat{\Gamma}_2^*)^{-1}(\varphi(s) - \varphi(t))$. For the given alternative, and under the normality assumption,

$$\Gamma_1^* = \Gamma_2^* = \begin{bmatrix} 2 & 2\rho^2 & \rho \\ & 2 & \rho \\ & & 1 \end{bmatrix},$$

so

$$\begin{aligned} &\operatorname{plim}(\varphi(s) - \varphi(t))'(\hat{\Gamma}_1^* + \hat{\Gamma}_2^*)^{-1}(\varphi(s) - \varphi(t)) \\ &\quad = (-\log\eta, -\log\eta)\tfrac{1}{4}\begin{bmatrix} 1 + \rho^2/2(1-\rho^2) & -\rho^2/2(1-\rho^2) \\ -\rho^2/2(1-\rho^2) & 1 + \rho^2/2(1-\rho^2) \end{bmatrix}\begin{bmatrix} -\log\eta \\ -\log\eta \end{bmatrix} \\ &\quad = \tfrac{1}{2}(\log\eta)^2, \end{aligned}$$

so condition (ii) is satisfied with $t = 1$. The statistic has a limiting χ^2 distribution under H_0, so condition (i) is satisfied with $a = 1$ and $r = 1$. The approximate Bahadur efficiency of the likelihood-ratio test to the standard error test is thus $4\log((1+\eta)^2/4\eta)/(\log\eta)^2$. The limit of this expression as $\eta \to 1$ is 1, and the limit as $\eta \to \infty$ (or as $\eta \to 0$) is 0. For $\eta = 2$, the efficiency is .981, and for $\eta = 4$ is .929.

Now consider the alternative

$$\Sigma_1 = \begin{bmatrix} \mu_{20} & 0 \\ 0 & \mu_{02} \end{bmatrix}, \quad \Sigma_2 = \eta\Sigma_1, \qquad \eta > 0.$$

Suppose the distributions under H_0 and H_A have independent marginals with finite fourth moments and common kurotosis γ. In this case

$$\Gamma_1^* = \Gamma_2^* = \operatorname{diag}(2+\gamma, 2+\gamma, 1),$$

and the slope of the standard error test is

$$\operatorname{plim}(\varphi(s) - \varphi(t))'(\hat{\Gamma}_1^* + \hat{\Gamma}_2^*)^{-1}(\varphi(s) - \varphi(t)) = (\log\eta)^2/(2+\gamma).$$

The slope of the Box test is similarly found to be approximately $(\log\eta)^2/(2k/(k-1) + \gamma)$, so the approximate relative efficiency of the Box test to the standard error test is $(2+\gamma)/(2k/(k-1) + \gamma)$, which is independent of η and coincides with the approximate Pitman relative efficiency.

Acknowledgments. I am grateful to Rupert Miller, Lincoln Moses and Ingram Olkin for their encouragement and many helpful suggestions in the course of this work.

REFERENCES

[1] ANDERSON, T. W. (1958). *Introduction to Multivariate Statistical Analysis.* Wiley, New York.

[2] ARVESEN, J. N. (1969). Jackknifing U-statistics. *Ann. Math. Statist.* **40** 2076–2100.

[3] BAHADUR, R. R. (1960). Stochastic comparison of tests. *Ann. Math. Statist.* **31** 276–295.
[4] BOX, G. E. P. (1953). Non-normality and tests on variances. *Biometrika* **40** 318–335.
[5] CRAMÉR, H. (1946). *Mathematical Methods of Statistics*. Princeton University Press.
[6] GAYEN, A. K. (1951). Distribution of correlation coefficient from non-normal samples. *Biometrika* **38** 219–247.
[7] GLESER, L. J. (1964). On a measure of test efficiency proposed by R. R. Bahadur. *Ann. Math. Statist.* **35** 1537–1544.
[8] GLESER, L. J. (1966). The comparison of multivariate tests of hypotheses by means of Bahadur efficiency. *Sankhyā Ser. A* **28** 157–174.
[9] HANNAN, E. J. (1956). The asymptotic powers of certain tests based on multiple correlations. *J. Roy. Statist. Soc. Ser. B* **18** 227–233.
[10] HOTELLING, H. (1950). A generalized T-test and measure of multivariate dispersion. *Proc. Second Berkely Symp. Math. Statist. Prob.* **1** 23–41.
[11] MILLER, R. G. JR. (1964). A trustworthy jackknife. *Ann. Math. Statist.* **35** 1594–1605.
[12] MILLER, R. G. JR. (1968). Jackknifing variances. *Ann. Math. Statist.* **39** 567–582.
[13] PITMAN, E. J. G. (1949). *Lecture Notes on Non-Parametric Inference*. Columbia University (Unpublished).
[14] PRATT, J. W. (1959). On a general concept of "in probability." *Ann. Math. Statist.* **30** 549–558.
[15] QUENOUILLE, M. H. (1956). Notes on bias in estimation. *Biometrika* **43** 353–360.
[16] ROY, S. N. and GNANADESIKAN, R. (1960). On tests of certain types of hypotheses—and the associated confidence bounds. *Ann. Math. Statist.* **30** 1277.
[17] TUKEY, J. W. (1962). Data analysis and behavioral science. Unpublished manuscript.
[18] WILKS, S. S. (1962). *Mathematical Statistics*. Wiley, New York.