

A NOTE ON BOX'S GENERAL METHOD OF APPROXIMATION FOR THE NULL DISTRIBUTIONS OF LIKELIHOOD CRITERIA*

LEON JAY GLESER AND INGRAM OLKIN

(Received Aug. 1, 1972; revised Jan. 28, 1974)

1. Introduction and preliminary results

Many multivariate test statistics (such as the likelihood ratio test statistic for MANOVA and the Bartlett modification of the likelihood ratio test statistic for testing the equality of covariance matrices) have null distributions whose h th moment M_h is of the form

$$(1.1) \quad M_h = K \left(\frac{\prod_{j=1}^J (y_j)^{y_j}}{\prod_{i=1}^I (x_i)^{x_i}} \right)^h \frac{\prod_{i=1}^I \Gamma(x_i(1+h) + \xi_i)}{\prod_{j=1}^J \Gamma(y_j(1+h) + \eta_j)}, \quad h=0, 1, 2, \dots,$$

where K is a constant (such that $M_0=1$), the x 's and y 's are positive numbers, and $\sum_{i=1}^I x_i = \sum_{j=1}^J y_j$. Such statistics have a range of variation from 0 to 1, so that the moments M_h , $h=0, 1, 2, \dots$, determine the null distribution.

For any statistic W , $0 \leq W \leq 1$, whose moments are of the form (1.1), Box [4] has proposed an asymptotic expansion for the cumulative distribution function (c.d.f.). This expansion provides an accurate method for determining the critical constants defining rejection regions for the multivariate tests mentioned above. One form in which this expansion is often given (see Anderson [1], p. 207) is the following:

$$(1.2) \quad P \{-2 \log W \leq t\} = (1 - \phi) P \{\chi^2 \leq \rho t\} + \phi P \{\chi^2_{f+4} \leq \rho t\} + R(\delta).$$

Here, $\delta = \sum_{i=1}^I x_i - \sum_{j=1}^J y_j$,

$$(1.3) \quad f = -2 \left[\sum_{i=1}^I \xi_i - \sum_{j=1}^J \eta_j - \frac{1}{2}(I - J) \right],$$

* Work supported in part by the Educational Testing Service, the Air Force Office of Scientific Research (Contract F44620-70-C-0066) at Johns Hopkins University, and the National Science Foundation (Grant GP-32326X) at Stanford University.

$$(1.4) \quad \phi = \frac{1}{-6\rho^2} \left[\sum_{i=1}^I x_i^{-2} B_3((1-\rho)x_i + \xi_i) - \sum_{j=1}^J y_j^{-2} B_3((1-\rho)y_j + \eta_j) \right],$$

ρ is the solution of the equation

$$(1.5) \quad \sum_{i=1}^I x_i^{-1} B_2((1-\rho)x_i + \xi_i) - \sum_{j=1}^J y_j^{-1} B_2((1-\rho)y_j + \eta_j) = 0,$$

and $R(\delta)$ is a remainder term. (Note: ρ is chosen so that Anderson's ω_i is 0.) Also $B_1(u)$, $B_2(u)$, and $B_3(u)$ are the Bernoulli polynomials of degree 1, 2, and 3 defined by:

$$(1.6) \quad B_1(u) = u - \frac{1}{2}, \quad B_2(u) = u^2 - u + \frac{1}{6}, \quad B_3(u) = u^3 - \frac{3}{2}u^2 + \frac{1}{2}u.$$

If $\lim_{\delta \rightarrow \infty} \delta^{-1}x_i > 0$ and $\lim_{\delta \rightarrow \infty} (1-\rho)x_i$ exists, $i=1, 2, \dots, I$, and if $\lim_{\delta \rightarrow \infty} \delta^{-1}y_j > 0$ and $\lim_{\delta \rightarrow \infty} (1-\rho)y_j$ exists, $j=1, 2, \dots, J$, then Anderson ([1], pp. 203-207) sketches an argument which shows that the remainder term $R(\delta)$ in (1.2) is $O(\delta^{-3})$ as $\delta \rightarrow \infty$. In the case of the likelihood ratio test for MANOVA, Anderson [1] gives examples showing the high accuracy provided by the approximation (1.2).

From Anderson's discussion (*ibid.*) and the statistical literature, it might appear that to obtain the constants ρ and ϕ , Equation (1.5) must be solved *ab initio* for ρ in each particular case. However, it can be shown by use of (1.6) and algebraic manipulation that solution of (1.5) yields

$$(1.7) \quad \rho = 1 - \frac{1}{f} \left[\sum_{i=1}^I x_i^{-1} B_2(\xi_i) - \sum_{j=1}^J y_j^{-1} B_2(\eta_j) \right].$$

Substitution of (1.7) into (1.4) then gives us

$$(1.8) \quad \phi = \frac{1}{6\rho^2} \left[\sum_{i=1}^I x_i^{-2} B_3(\xi_i) - \sum_{j=1}^J y_j^{-2} B_3(\eta_j) + \frac{3}{2}(1-\rho)^2 f \right].$$

In the important special case when the h th moment of W has the form (1.1) with $I=J \equiv L$, and

$$(1.9) \quad \begin{aligned} x_1 = x_2 = \dots = x_I = y_1 = y_2 = \dots = y_J = z, \\ \xi_i = a + bi, \quad \eta_i = c + di, \quad i=1, \dots, L, \end{aligned}$$

then (1.3), (1.7), and (1.8), respectively, simplify to the following:

$$(1.10) \quad f = -2L \left[(a-c) + \frac{1}{2}(b-d)(L+1) \right],$$

$$(1.11) \quad \rho = 1 - \frac{L}{fz} \left[(a^2 - c^2) + (ab - cd)(L+1) \right]$$

$$(1.12) \quad \phi = -\frac{1}{6z^2\rho^2} \left[A - \frac{3}{2}(1-\rho)zf + \frac{1}{2}f + \frac{3}{2}(1-\rho)^2z^2f \right],$$

where

$$A = (a^3 - c^3)L + \frac{3}{2}(a^2b - c^2d)(L)(L+1) + \frac{1}{2}(ab^2 - cd^2)(L)(L+1)(2L+1) + \frac{1}{4}(b^3 - d^3)(L^2)(L+1)^2.$$

Examples of test statistics W whose moments (under their respective null hypotheses) satisfy (1.1) with the restrictions (1.9) are the likelihood ratio test statistics for MANOVA, and various special cases of the likelihood ratio test statistics for independence of sets of variates.

Formulas for the constants f , ρ , and ϕ in the expansion (1.2) are known for a great many statistics used in multivariate analysis. In many cases a new multivariate test statistic W can be represented distributionally as a function of independent statistics W_1, W_2, \dots, W_G , each of which has an expansion (1.2) for its c.d.f. for which the coefficients f_g, ρ_g, ϕ_g have known values (or for which these values can be easily obtained), $g=1, 2, \dots, G$. If W also has a c.d.f. which can be expanded as in (1.2), the coefficients f, ρ, ϕ of this expansion can, of course, be obtained by appropriate substitution in (1.3), (1.7), and (1.8). However, since W is a function of W_1, \dots, W_G , one would expect that f, ρ , and ϕ are functions of the known quantities f_g, ρ_g , and $\phi_g, g=1, 2, \dots, G$, thus offering an alternative and often more convenient way of obtaining the coefficients f, ρ , and ϕ . Formulas for doing this are given for the important case where $W = \prod_{g=1}^G W_g$; two applications of the results are provided.

2. Approximation for the distribution of the product of independent statistics whose moments are of the form (1.1)

Let the independent random variables $W_g, 0 \leq W_g \leq 1$, be independent, with h th moments of the form

$$(2.1) \quad E((W_g)^h) = K_g \left(\frac{\prod_{j=1}^{J_g} (y_{gj})^{y_{gj}}}{\prod_{i=1}^{I_g} (x_{gi})^{x_{gi}}} \right)^h \frac{\prod_{i=1}^{I_g} \Gamma(x_{gi}(1+h) + \xi_{gi})}{\prod_{j=1}^{J_g} \Gamma(y_{gj}(1+h) + \eta_{gj})},$$

$$h = 0, 1, 2, \dots,$$

where K_g is a constant (such that $E(W_g^0)=1$), and $\sum_{i=1}^{I_g} x_{gi} = \sum_{j=1}^{J_g} y_{gj}$, for $g=1, \dots, G$. Let $W = \prod_{g=1}^G W_g$. Note that since W_1, \dots, W_G are independent and $0 \leq W_g \leq 1$, $g=1, \dots, G$, we have $0 \leq W \leq 1$ and $E(W^h) = \prod_{g=1}^G E(W_g^h)$, which is of the form (1.1) with $I=I_1+\dots+I_G$, $J=J_1+\dots+J_G$,

$$\begin{aligned} x_{1i}, & \quad i=1, \dots, I_1, \\ x_{2, i-I_1}, & \quad i=I_1+1, \dots, I_1+I_2, \\ \vdots & \\ x_{G, i-\sum_{k=1}^{G-1} I_k}, & \quad i=\sum_{k=1}^{G-1} I_k+1, \dots, I, \end{aligned}$$

and similar formulas relating ξ_i to the ξ_{gi} 's, y_j to the y_{gj} 's, and η_j to the η_{gj} 's.

Since the moments of W_1, \dots, W_G , and W are all of the form (1.1), it follows from the results of Section 1 that the c.d.f.'s of these variables can be expanded in the form (1.2). By means of Equations (1.3), (1.7), and (1.8), straightforward algebraic manipulation yields the following relationships between the coefficients f_g, ρ_g, ϕ_g of the expansions of the c.d.f.'s of W_g , $g=1, 2, \dots, G$, and the corresponding coefficients f, ρ, ϕ in the expansion (1.2) of the c.d.f. of W .

THEOREM 1. *Let the statistically independent variables W_g , $0 \leq W_g \leq 1$, have moments of the form (2.1), $g=1, 2, \dots, G$. Let f_g, ρ_g, ϕ_g , and $\delta_g = \sum_{i=1}^{I_g} x_{gi} = \sum_{j=1}^{J_g} y_{gj}$ be the constants in the Box expansion (1.2) of the c.d.f. of W_g , $g=1, 2, \dots, G$. Finally, let $W = \prod_{g=1}^G W_g$. Then*

$$(2.2) \quad P\{-2 \log W \leq t\} = (1-\phi) P\{\chi^2_{\delta} \leq \rho t\} + \phi P\{\chi^2_{\delta+t} \leq \rho t\} + R(\delta),$$

where $\delta = \sum_{g=1}^G \delta_g$, $f = \sum_{g=1}^G f_g$, $\rho = \frac{1}{f} \sum_{g=1}^G f_g \rho_g$, and

$$\phi = \frac{1}{\rho^2} \sum_{g=1}^G \rho_g^2 \phi_g + \frac{1}{4\rho^2 f} \sum_{g < h} f_g f_h (\rho_g - \rho_h)^2.$$

Note that if in the Box expansions for W_1, W_2, \dots, W_G , each $R_g(\delta_g)$ is $O(\delta_g^{-3})$ as $\delta_g \rightarrow \infty$, and if $\delta_1, \delta_2, \dots, \delta_G$ are all asymptotically of the same order of magnitude (i.e., $\lim \delta_g \delta_h^{-1} > 0$ as $\delta_g, \delta_h \rightarrow \infty$ for all $g \neq h$), then $R(\delta)$ in (2.2) is $O(\delta^{-3})$. In practical use of Theorem 1, the δ_g 's will usually be asymptotically of the same order of magnitude. If the δ_g 's are not asymptotically of the same order of magnitude, Equation (2.2) is formally correct (when δ, f, ρ , and ϕ are defined as in Theorem 1), but the order of magnitude of the remainder term $R(\delta)$ in δ must be separately investigated.

3. Applications to multivariate hypothesis testing problems

Suppose we are interested in testing whether either the mean vectors and/or covariance matrices of k multivariate normal populations are identical. Suppose that an observation (p dimensional row vector) $x^{(i)}$ from the i th population has a p -variate normal distribution with mean vector $\mu^{(i)}$ and covariance matrix $\Sigma^{(i)}$, $i=1, 2, \dots, k$. Let $x^{(i)}$ be partitioned as $(x_1^{(i)}, x_2^{(i)})$, where $x_1^{(i)}$ is $1 \times q$, and let

$$\mu^{(i)} = (\mu_1^{(i)}, \mu_2^{(i)}), \quad \Sigma^{(i)} = \begin{pmatrix} \Sigma_{11}^{(i)} & \Sigma_{12}^{(i)} \\ \Sigma_{21}^{(i)} & \Sigma_{22}^{(i)} \end{pmatrix},$$

be correspondingly partitioned, $i=1, 2, \dots, k$. Suppose that we observe N_i observations from the i th population, $i=1, 2, \dots, k$.

We consider two tests of hypotheses. The first test compares the null hypothesis

$$H_{mvc} : \mu^{(1)} = \mu^{(2)} = \dots = \mu^{(k)}, \quad \Sigma^{(1)} = \Sigma^{(2)} = \dots = \Sigma^{(k)},$$

against general alternatives. In the second test, we compare the null hypothesis H_{mvc} against the alternative:

$$H_{mvc'} : \mu_1^{(1)} = \mu_1^{(2)} = \dots = \mu_1^{(k)}, \quad \Sigma_{11}^{(1)} = \Sigma_{11}^{(2)} = \dots = \Sigma_{11}^{(k)}.$$

Let $\bar{x}^{(i)} = (\bar{x}_1^{(i)}, \bar{x}_2^{(i)})$ be the sample mean vector, let $N = \sum_{i=1}^k N_i$, $\bar{x} = (\bar{x}_1, \bar{x}_2) = \frac{1}{N} \sum_{i=1}^k N_i (\bar{x}_1^{(i)}, \bar{x}_2^{(i)})$, $V^{(i)}$ be the sample cross-product matrix from the i th population, $i=1, 2, \dots, k$, and A the cross-product matrix of the means:

$$V^{(i)} = \begin{pmatrix} V_{11}^{(i)} & V_{12}^{(i)} \\ V_{21}^{(i)} & V_{22}^{(i)} \end{pmatrix}, \quad A = \begin{pmatrix} A_{11} & A_{12} \\ A_{21} & A_{22} \end{pmatrix} = \sum_{i=1}^k N_i (\bar{x}^{(i)} - \bar{x})' (\bar{x}^{(i)} - \bar{x}).$$

3.1. Test of H_{mvc} versus general alternatives

Anderson [1] suggests testing H_{mvc} against general alternatives by means of the test statistic

$$(3.1) \quad W = \left(\frac{\prod_{i=1}^k |(1/n_i) V^{(i)}|^{n_i/2}}{|(1/n) \sum_{i=1}^k V^{(i)}|^{n/2}} \right) \left(\frac{|(1/n) \sum_{i=1}^k V^{(i)}|}{|(1/n) (\sum_{i=1}^k V^{(i)} + A)|} \right)^{n/2} \equiv W_1 W_2,$$

where $n_i = N_i - 1$, $i=1, 2, \dots, k$, and $n = \sum_{i=1}^k n_i$.

As Anderson [1], shows, the statistics W_1 and W_2 are independent when H_{mvc} holds, the moments of W_1 under H_{mvc} are of the form

(1.1), and

$$f_1 = \frac{1}{2}(k-1)p(p+1), \quad \rho_1 = 1 - \left(\sum_{i=1}^k \frac{1}{n_i} - \frac{1}{n} \right) \frac{2p^2 + 3p - 1}{6(p+1)(k-1)},$$

$$\phi_1 = \frac{p(p+1) \left\{ (p-1)(p+2) \left[\sum_{i=1}^k (1/n_i)^2 - (1/n)^2 \right] - 6(k-1)(1-\rho_1)^2 \right\}}{48\rho_1^2}.$$

(In comparing (3.2) with Anderson's results it should be noted that his q is our k .)

Anderson ([1], p. 207) gives the h th moments of $\lambda = (W_2)^{N/n}$ under $H_{m'vc}$. From his result,

$$E((W_2)^h) = K_2 \frac{\prod_{s=1}^p \Gamma((1/2)n(1+h) + 1/2 - (1/2)s)}{\prod_{t=1}^p \Gamma((1/2)n(1+h) + k/2 - (1/2)t)}.$$

Thus, applying (1.11) through (1.13), with $L=p$, $z=n/2$, $a=1/2$, $b=-1/2$, $c=k/2$, $d=-1/2$, we find that

$$f_2 = p(k-1), \quad \rho_2 = 1 - \frac{p-k+2}{2n}, \quad \phi_2 = \frac{p(k-1)[p^2 + (k-1)^2 - 5]}{48n^2(\rho_2)^2}.$$

To obtain an asymptotic expansion for the c.d.f. of the test statistic W under $H_{m'vc}$, Anderson ([1], p. 255) goes back to the h th moments of W and applies the Box expansion method *ab initio*.

Alternatively, using Theorem 1, we find that the constants f , ρ , and ϕ in the Box expansion (1.2) of the c.d.f. of W are given by

$$(3.3) \quad f = \frac{1}{2}(k-1)p(p+3),$$

$$\rho = 1 - \left(\sum_{i=1}^k \frac{1}{n_i} - \frac{1}{n} \right) \frac{2p^2 + 3p - 1}{6(p+3)(k-1)} - \frac{p-k+2}{n(p+3)},$$

$$\phi = \frac{p}{288\rho^2} \left\{ 6 \left[\sum_{i=1}^k \frac{1}{(n_i)^2} - \frac{1}{n^2} \right] (p^2 - 1)(p+2) - 36(p+3)(k-1)(1-\rho)^2 \right. \\ \left. - \frac{12(k-1)}{n^2} (-2k^2 + 7k + 3pk - 2p^2 - 6p - 4) \right\}.$$

3.2. Test of $H_{m'vc}$ versus $H_{m'vc'}$

Gleser and Olkin [5] show that the likelihood ratio test statistic for testing $H_{m'vc}$ against the alternative $H_{m'vc'}$, modified along the lines suggested by Bartlett [2], is

$$U_2 = \frac{\left(\prod_{i=1}^k |(1/n_i)V_{22 \cdot 1}^{(i)}|^{n_i/2} \right) \left| (1/n) \left(\sum_{i=1}^k V_{11}^{(i)} + A_{11} \right) \right|^{n/2}}{\left| (1/n) \left(\sum_{i=1}^k V^{(i)} + A \right) \right|^{n/2}}.$$

where $V_{22 \cdot 1}^{(i)} = V_{22}^{(i)} - V_{21}^{(i)}(V_{11}^{(i)})^{-1}V_{12}^{(i)}$, $n_i = N_i - 1$, $i = 1, 2, \dots, k$, and $n = \sum_{i=1}^k n_i$. (There is more than one way to modify the likelihood ratio test statistic along the lines suggested by Bartlett [2]. One way is given here; another, and possibly preferable, way is considered in Gleser and Olkin [5].)

To obtain the c.d.f. of U_2 let U_1 be a similar modification of the likelihood ratio test statistic for testing hypothesis $H_{m'vc}$ against *general* alternatives. Gleser and Olkin [5] have derived the likelihood ratio test statistic. From their result, we find that

$$U_1 = \frac{\prod_{i=1}^k |(1/n_i)V_{11}^{(i)}|^{n_i/2}}{\left| (1/n) \left(\sum_{i=1}^k V_{11}^{(i)} + A_{11} \right) \right|^{n/2}}.$$

Comparing the statistic W defined in (3.1) with U_1U_2 , and recalling that $|V^{(i)}| = |V_{11}^{(i)}| |V_{22 \cdot 1}^{(i)}|$ for $i = 1, 2, \dots, k$, we see that $W = U_1U_2$.

Since under $H_{m'vc}$ the statistics $\bar{x}_2^{(i)} - \bar{x}_1^{(i)}(V_{11}^{(i)})^{-1}V_{12}^{(i)}$, $(V_{11}^{(i)})^{-1}V_{12}^{(i)}$, $V_{22 \cdot 1}^{(i)}$, $i = 1, 2, \dots, k$, \bar{x}_1 , and $\left(\sum_{i=1}^k V_{11}^{(i)} + A_{11} \right)$ are complete and sufficient, and since the distribution of U_1 is the same for all values of the parameters $\mu^{(1)}, \mu^{(2)}, \dots, \mu^{(k)}$, $\Sigma^{(1)}, \Sigma^{(2)}, \dots, \Sigma^{(k)}$ obeying $H_{m'vc}$, it follows from a theorem of Basu [3] that U_1 and U_2 are statistically independent under $H_{m'vc}$ (and thus under H_{mvc}).

Note that U_1 has the same form as W , except that U_1 is a function only of $x_1^{(i)}(j)$, $j = 1, 2, \dots, N_i$; $i = 1, 2, \dots, k$. That is, U_1 is a q -dimensional version of W . It has already been noted that the moments of W under H_{mvc} are of the form (1.1). The moments of U_1 under H_{mvc} can be obtained from the formula for the moments of W by everywhere replacing p by q . Since U_1 and U_2 are independent under H_{mvc} , and since $W = U_1U_2$, the h th moment of U_2 under H_{mvc} equals the h th moment of W under H_{mvc} divided by the h th moment of U_1 under H_{mvc} . It follows that the moments of U_2 under H_{mvc} have the form (1.1).

From the preceding discussion and Theorem 1, it follows that the c.d.f.'s of U_1 , U_2 , and W all have asymptotic expansions of the form (1.2). Let f , ρ , and ϕ be the coefficients in the expansion (1.2) for the c.d.f. of W ; these constants are given by (3.2). Let f_g^* , ρ_g^* , and ϕ_g^* be the coefficients in the expansion for the c.d.f. of U_g , $g = 1, 2$. Since U_1 is a q -dimensional version of W , the coefficients f_1^* , ρ_1^* , and ϕ_1^* can

be obtained by substituting q for p in (3.3). Finally, from Theorem 1 we know that

$$(3.4) \quad f = f_1^* + f_2^*, \quad \rho = \frac{f_1^* \rho_1^* + f_2^* \rho_2^*}{f},$$

$$\phi = \frac{(\rho_1^*)^2 \phi_1^* + (\rho_2^*)^2 \phi_2^*}{\rho^2} + \frac{f_1^* f_2^*}{4f} (\rho_1^* - \rho_2^*)^2.$$

Solving for f_2^* , ρ_2^* , and ϕ_2^* in (3.4) yields

$$f_2^* = \frac{1}{2} (k-1)(p-q)(p+q+3),$$

$$\rho_2^* = 1 - \left[\sum_{i=1}^k \frac{1}{n_i} - \frac{1}{n} \right] \left[\frac{2p^2 + 2pq + 2q^2 + 3p + 3q - 1}{6(p+q+3)(k-1)} \right] - \left[\frac{p+q-k+2}{n(p+q+3)} \right],$$

$$\phi_2^* = \frac{1}{288(\rho_2^*)^2} \left\{ 6 \left[\sum_{i=1}^k \frac{1}{(n_i)^2} - \frac{1}{n^2} \right] [(p^2-1)(p^2+2p) - (q^2-1)(q^2+2q)] \right.$$

$$\left. - \left[\frac{(12)(k-1)(p-q)}{n^2} \right] [3(p+q)(k-2) - 2(p^2+pq+q^2) - 2k^2 + 7k - 4] \right.$$

$$\left. - 72f_2^*(1-\rho_2^*)^2 \right\}.$$

We conclude that

$$P\{-2 \log U_2 \leq t\} = (1 - \phi_2^*) P\{\chi_{j_2^*}^2 \leq \rho_2^* t\} + \phi_2^* P\{\chi_{j_2^*+4}^2 \leq \rho_2^* t\} + O(n^{-3}),$$

where f_2^* , ρ_2^* , and ϕ_2^* are given above.

THE JOHNS HOPKINS UNIVERSITY AND EDUCATIONAL TESTING SERVICE*
STANFORD UNIVERSITY AND EDUCATIONAL TESTING SERVICE

REFERENCES

- [1] Anderson, T. W. (1958). *An Introduction to Multivariate Statistical Analysis*, John Wiley and Sons, Inc., New York.
- [2] Bartlett, M. S. (1937). Properties of sufficiency and statistical tests, *Proc. Roy. Soc., A*, **160**, 268-282.
- [3] Basu, D. (1955). On statistics independent of a complete sufficient statistic, *Sankhyā*, **18**, 377-380.
- [4] Box, G. E. P. (1949). A general distribution theory for a class of likelihood criteria, *Biometrika*, **36**, 317-346.
- [5] Gleser, L. J. and Olkin, I. (1973). Multivariate statistical inference under marginal structure, I, *Brit. J. Math. Statist. Psychol.*, **26**, 98-123.

* Now at Purdue University.