A GENERAL DISTRIBUTION THEORY FOR A CLASS OF LIKELIHOOD CRITERIA

By G. E. P. BOX

Imperial Chemical Industries, Dyestuffs Division Headquarters, Blackley, Manchester

1. Introduction

The likelihood ratio method of Neyman & Pearson (1928) has been used by many different workers for the derivation of criteria appropriate for the testing of a large variety of hypotheses. Plackett (1946), in a recent survey of literature on testing the equality of variances and covariances, lists, on this problem alone, criteria for the testing of no less than thirty-one hypotheses investigated at different times by workers in this field. Most of the criteria either have been or can be arrived at by the likelihood ratio method. In the preface to his survey Plackett says: 'Generally speaking the difficulties in testing such hypotheses lie not so much in deriving criteria—but in finding their exact distributions when the hypotheses are true and determining the best critical region to adopt.'

Although in many cases the exact distributions cannot be obtained in a form which is of practical use, it is usually possible to obtain the moments, and these may be used to obtain approximations. In some cases, for instance, a suitable power of the likelihood statistic has been found to be distributed approximately in the type I form, and good approximations have been obtained by equating the moments of the likelihood statistic to this curve. For example, in the original paper on the L_1 test for homogeneity of variances, Neyman & Pearson (1931) suggested that the distribution could be approximately represented in this way, and later Bishop & Nair (1939) showed that the significance points obtained by Nayer (1936), using this method, were in excellent agreement with the true values. The fitting of the type I curve is simple once the moments are obtained, but these moments, being the products of Γ-functions, are usually rather troublesome to calculate. To overcome this difficulty, Bishop (1939), working on the distribution of the multivariate equivalent of the L_1 test (the test for constancy of variances and covariances in k p-variate samples), derived empirical expressions for the parameters of the appropriate type I curve, thus avoiding the troublesome intermediate step of calculating moments. Bishop mentions that Nair succeeded in finding similar expressions on a theoretical basis, and Tukey & Wilks (1946) give a more general theoretical method to find approximations of this kind.

A different line of approach was adopted by Bartlett (1937). Neyman & Pearson had pointed out in their original paper that, if N' is the total sample size, $-N' \log_e L_1$ would be asymptotically distributed as χ^2. From considerations of sufficiency Bartlett obtained what was in effect a modified form (which, following Hartley & Pearson (1946), we shall call M) of this logarithmic statistic. From the moments of the modified likelihood statistic he was able to develop a scale factor C, which was related to the effective sizes of samples and which approached the value unity as the sample sizes became large. The distribution of M/C was then very well represented by χ^2 even when the samples were small. Bartlett later (1938) used the same method to obtain an approximation for the test of significance in multivariate analysis. In 1940 Hartley, starting from the moments of the modified likelihood statistic, obtained an asymptotic series of χ^2 integrals for the logarithmic statistic M which agreed very closely with the exact distribution. In 1941 Wald & Brookner, investigating an entirely different problem, the distribution of Wilks's statistic for testing independence of k sets of variates, again starting from the moments of the likelihood statistic λ, eventually obtained an expression for the distribution of a logarithmic statistic (a negative multiple of $\log_e \lambda$) in the form of an asymptotic χ^2 series. This was later modified by Rao (1948) in the important special case of two groups, when it corresponds to the test of significance in multivariate analysis previously referred to. Neither Wald & Brookner nor Rao investigated the accuracy of these series.

It is possible therefore to distinguish two definite lines of approach, which have been used in certain cases where the moments of the likelihood criteria are known but the exact distributions are not. On the one hand the moments have been used to fit the Pearson-type curve. This usually gives an adequate approximation, but owing to the amount of labour involved in the calculation of the moments it would not be attractive for routine significance testing unless methods such as Bishop's could be used to obtain the parameters of the fitted curve directly, or the results from the method could be tabled. On the other hand, the general expression for the moments of the likelihood statistic has been used in certain cases to obtain for the distribution of the logarithmic statistic M, a χ^2 approximation and an asymptotic χ^2 series. It will be the object of the present paper to investigate in some detail this second line of attack.

The method will be investigated in particular for two general criteria:

(1) The test of constancy of variance and covariance of k sets of p-variate samples. This includes, as an important special case when $p = 1$, the test for constancy of variance in k samples.

(2) Wilks's test for the independence of k sets of residuals, the lth set having p_l variates. When $k = 2$ this corresponds to the test of significance used in multivariate regression and analysis of variance and covariance, and when $k = 2$ and p_1 or p_2 is unity, it gives the corresponding well-known univariate tests. In the latter case, of course, the exact distributions are known.

We shall refer to these two criteria as generalized tests for homoscedasticity and independence, respectively. The assumption of normality or multinormality for the distributions of the original observations will be made throughout this paper.

Taking for our test function M, a negative multiple of the natural logarithm of the likelihood statistic (or some modification of it), we shall obtain in each case,

(*a*) a series solution which, we shall demonstrate, agrees very closely with the exact distribution,

(*b*) an approximate solution using a single χ^2 distribution,

(*c*) a rather better approximation using a single F distribution.

The accuracy of the various methods and the relation of the results to those of other workers will be discussed.

2. The generalized test of homoscedasticity

The univariate statistic. The L_1 statistic of Neyman & Pearson for testing the homogeneity of a set of variances, takes the form of the ratio of a weighted geometric mean of variances to a weighted arithmetic mean, where the weights are the sample numbers. Welch (1935,

1936) generalized the test to cover the case when residuals from a fitted regression equation were tested for homoscedasticity, and derived the moments for a modified criterion in which the weights could have any values whatever. In 1936 Nayer tabled the approximate significance points for L_1 in the cases of equal sample numbers by fitting type I curves to the distributions by the method of moments, as suggested in the original memoir by Neyman & Pearson.

The statistic proposed by Bartlett (1937) which we shall call M is given by

$$M = N \log_e s - \sum_l \nu_l \log_e s_l,$$

where

$$s = (\Sigma \nu_l s_l)/N,$$

and s_l is the usual unbiased estimate of the variance in the lth group, $l = 1, 2, \ldots, k$, based on sums of squares having ν_l degrees of freedom, and $N = \Sigma \nu_l$. It was later shown (Brown, 1939; Pitman, 1939; Bishop & Nair, 1939) that this criterion is unbiased in the sense used by Neyman & Pearson (1936, 1938). Nair (1939) derived a series solution for the distribution of the likelihood statistic in the case of equal sample numbers; his solution is very involved, but has been used as a standard to check approximations. Bishop & Nair (1939) used this series to check the accuracy of the type I approximation used in Nayer's table. They also checked the Bartlett (1937) approximation and found that both methods were fairly good except when the degrees of freedom were small. In the case of unequal samples, however, Nayer's tables were not available, and in view of the labour involved in the type I fit, the χ^2 method of Bartlett's was preferred.

Hartley's (1940) asymptotic series depended to the degree of approximation used, on two parameters c_1 and c_3 which varied with the effective sample size and relative composition of the groups

$$c_1 = \Sigma \frac{1}{\nu_l} - \frac{1}{N}, \quad c_3 = \Sigma \frac{1}{\nu_l^3} - \frac{1}{N^3}.$$

The first is related to Bartlett's scale factor C, in fact

$$C = 1 + \frac{c_1}{3(k-1)}.$$

Tables were afterwards computed by Thompson & Merrington (1946) from Hartley's formula, and comparisons were made with the values calculated by Bishop & Nair.

The multivariate statistic. In the multivariate case Wilks (1932) derived the likelihood ratio test and obtained the moments of the criterion, which is a generalized form of that used in the univariate test, the determinants of estimated variances and covariances replacing the variances. Bishop (1939) took as his criterion l_1, the $1/N'$th power of the likelihood statistic, N' being the total number of observations. He gave reasons for believing that this criterion could be approximately represented by a type I curve

$$p(l_1) = \text{constant}\, l_1^{m_1-1}(1-l_1)^{m_2-1}, \tag{1}$$

by choosing the value of m_1 and m_2 so that the first two moments of the Pearson curve agreed with those of the criterion. His arguments were supported by the agreement found in a number of trials between the higher moments of the fitted type I curve and those of the criterion. Only in the case of two groups and either one or two variates was it possible to obtain a check against the exact distribution, but in these cases the agreement was very good. Unfortunately the labour involved in the calculation of the first two moments of the criterion was too great to allow this method to be recommended for routine use. Bishop therefore proceeded as follows:

(*a*) For the case of equal sample sizes he obtained, empirically, expressions for m_1 and m_2 in terms of the number of observations n in each group, the number of variates p, and the number of groups k

$$\left.\begin{aligned} m_1 &= k(n-p)-0{\cdot}01(k-1)(90-39p+9p^2),\\ m_2 &= 0{\cdot}25(k-1)p(p+1). \end{aligned}\right\} \tag{2}$$

(*b*) For unequal sample sizes he proposed approximating to $-2N'\log_e l_1$ by means of a χ^2 distribution using a scale factor G in a similar way to that adopted by Bartlett in the univariate case.

He showed that $-2N'\log_e l_1$ is approximately distributed as $G\chi^2$, where

$$G = 1+\frac{1}{f}\left[\sum_{l=1}^{k}\sum_{i=1}^{p}\{i^2/2n_l+i/(n_l-i)+n_l/3(n_l-i)^2\}\right.$$
$$\left.-\sum_{i=1}^{p}\{(k+i-1)^2/2N'+(k+i-1)/(N'-k+1-i)+N'/3(N'-k+1-i)^2\}\right]. \tag{3}$$

n_l is the number of observations in the lth group, $\sum_l n_l = N'$ and χ^2 is distributed with $f=\frac{1}{2}(k-1)p(p+1)$ degrees of freedom.

We shall refer to these methods as Bishop's methods (*a*) and (*b*). Bishop remarks that the scale factor G is rather troublesome to calculate unless $n_l = n$. George (1945) was able to evaluate the exact distribution in a number of simple cases. She used her results to test the accuracy of Bishop's approximations and found, in the cases she considered, that (*b*) was superior to (*a*).

Plackett (1947) suggested that in view of the unsatisfactory position with regard to the distribution of this criterion that it might be better to abandon it in favour of an alternative test derived by him which had the advantage that at least when $p = 1$ or 2, and for certain other special cases, the exact distribution was known. Plackett's test, however, has the disadvantages that the results depend on the particular arrangement of observations chosen, and that the samples must be of equal sizes.

2·1. *The present approach*

Suppose s_{ijl} is the usual unbiased estimate of the variance or covariance A^{ijl} between the ith and jth variable in the lth sample based on sums of squares and products having ν_l degrees of freedom, and suppose there are k such samples and s_{ij} is the average variance or covariance $\left(\sum_l \nu_l s_{ijl}\right)\Big/N$, where $N=\sum_l \nu_l$. We take as our criterion a generalized form of Bartlett's criterion

$$M = N\log_e|s_{ij}|-\sum_l(\nu_l\log_e|s_{ijl}|) \tag{4}$$

$$= -N\log_e L_1', \quad \text{where} \quad L_1' = \prod_{l=1}^{k}\left\{\frac{|s_{ijl}|}{|s_{ij}|}\right\}^{\nu_l/N}. \tag{5}$$

When the degrees of freedom are equal, M and L_1' are related with Bishop's l_1 as follows:

$$M = -2N\log_e l_1 \quad \text{and} \quad L_1' = l_1^2. \tag{6}$$

When the degrees of freedom are unequal, L_1' will differ from the likelihood statistic in weighting. When $p=1$, M is the criterion derived by Bartlett and later used by Hartley.

We proceed to obtain the moments of L_1' when the null hypothesis is true. If c_{ijl} are the sums of squares and products based on ν_l degrees of freedom corresponding to the s_{ijl}, we have $c_{ijl} = \nu_l s_{ijl}$, $c_{ij} = N s_{ij}$, so that $c_{ij} = \sum_l c_{ijl}$. The joint probability density of the c_{ijl} for the lth sample is given by the distribution discovered by Wishart (1928):

$$p(c_{11l}, c_{12l}, \ldots, c_{ppl}) = K(\nu_l) \,|\, c_{ijl} \,|^{\frac{1}{2}(\nu_l-p-1)} \exp\{-\tfrac{1}{2}\sum_{ij} A_{ijl} c_{ijl}\}, \tag{7}$$

where

$$\{K(\nu_l)\}^{-1} = 2^{\frac{1}{2}(\nu_l p)} \pi^{\frac{1}{4}p(p-1)} \prod_{j=0}^{p-1} \Gamma\left(\frac{\nu_l - j}{2}\right) |\, A_l \,|^{-\frac{1}{2}\nu_l}, \tag{8}$$

and A_l is the matrix of the A_{ijl}, the inverse of the matrix of the A^{ijl}.

When the null hypothesis is true, A_l is the same for each of the samples, $A_l = A$. $l = 1, 2, \ldots, k$, and the gth moment of $|\, c_{ij} \,|$ is

$$\int \prod_{l=1}^{k} \left\{ K(\nu_l) \,|\, c_{ijl} \,|^{\frac{1}{2}(\nu_l-p-1)} \,|\, c_{ij} \,|^{g/k} \exp\left(-\tfrac{1}{2}\sum_{ij} A_{ij} c_{ijl}\right)\right\} dc_{111}\, dc_{121} \ldots dc_{ppk}; \tag{9}$$

it is also given by

$$\int K(N) \,|\, c_{ij} \,|^{\frac{1}{2}(N-p-1)} \,|\, c_{ij} \,|^{g} \exp\left\{-\tfrac{1}{2}\sum_{ij} A_{ij} c_{ij}\right\} dc_{11}\, dc_{22} \ldots dc_{pp}. \tag{10}$$

Writing $\nu_l(1+2h)$ for ν_l on both sides of the identity and then taking $g = -Nh$ and integrating over the whole space for which the matrices of the c_{ij}, c_{ijl} are positive definite, we have

$$\mathscr{E}\left[\prod_{l=1}^{k} \left\{\frac{|\, c_{ijl} \,|}{|\, c_{ij} \,|}\right\}^{\nu_l}\right]^h \prod_{l=1}^{k} \left[\frac{K\{\nu_l(1+2h)\}}{K(\nu_l)}\right] = \frac{K\{N(1+2h)\}}{K(N)}. \tag{11}$$

That is

$$\mathscr{E}(L_1')^{Nh} = \mathscr{E}(e^{-M})^h = \frac{K\{N(1+2h)\}}{K(N)} \prod_{l=1}^{k} \left[\frac{K(\nu_l)}{K\{\nu_l(1+2h)\}} \left(\frac{N}{\nu_l}\right)^{p\nu_l h}\right] \tag{12}$$

$$= \prod_{l=1}^{k} \left(\frac{N}{\nu_l}\right)^{h\nu_l p} \prod_{j=0}^{p-1} \left[\frac{\Gamma\{\frac{1}{2}(N-j)\}}{\Gamma\{\frac{1}{2}(N(1+2h)-j)\}} \prod_{l=1}^{k} \frac{\Gamma\{\frac{1}{2}(\nu_l(1+2h)-j)\}}{\Gamma\{\frac{1}{2}(\nu_l - j)\}}\right]. \tag{13}$$

We have first proved equation (13) as an analytic identity for real h; it will, however, be generally valid in the range where the functions are analytic. We can thus obtain an expression for the characteristic function of ρM, where ρ is a constant $\leqslant 1$ at our choice, by replacing h by $-it\rho$ in the above expression. The reason for introducing the constant ρ will appear later. Further, if we write $N = \nu k$ (i.e. ν is the average of the degrees of freedom) and define new quantities μ, μ_l, β, β_l by the relations

$$\mu = \rho\nu_l, \quad \mu = \rho\nu, \quad \nu = \mu + \beta, \quad \nu_l = \mu_l + \beta_l, \tag{14}$$

we obtain the characteristic function of ρM in the form

$$\Phi(t) = \prod_{l=1}^{k} \left(\frac{k\mu}{\mu_l}\right)^{-itp\mu_l} \prod_{j=0}^{p-1} \left[\frac{\Gamma[\frac{1}{2}\{k\mu + k\beta - j\}]}{\Gamma[\frac{1}{2}\{k\mu(1-2it) + k\beta - j\}]} \prod_{l=1}^{k} \frac{\Gamma[\frac{1}{2}\{\mu_l(1-2it) + \beta_l - j\}]}{\Gamma[\frac{1}{2}(\mu_l + \beta_l - j)]}\right], \tag{15}$$

and taking logarithms we have the cumulant generating function in the form

$$\Psi(t) = g(t) - g(0), \tag{16}$$

where

$$g(t) = -\sum_{l=1}^{k} it\mu_l p \log\left(\frac{k\mu}{\mu_l}\right) + \sum_{j=0}^{p-1} \left[\sum_{l=1}^{k} \{\log \Gamma[\tfrac{1}{2}\{\mu_l(1-2it) + \beta_l - j\}]\} - \log \Gamma[\tfrac{1}{2}\{k\mu(1-2it) + k\beta - j\}]\right], \tag{17}$$

and $g(0)$ is a constant independent of t obtained by putting $t = 0$ in the above expression. Now Barnes (1899) was able to generalize Stirling's theorem, and he showed that for all x, real or complex, $\log \Gamma(x+h)$ may be expanded in an asymptotic series:

$$\log \Gamma(x+h) = \log \sqrt{(2\pi)} + (x+h-\tfrac{1}{2})\log x - x - \sum_{r=1}^{n} (-1)^r \frac{B_{r+1}(h)}{r(r+1)\,x^r} + R_{n+1}(x), \tag{18}$$

where $R_m(x)$ is a remainder term such that $|R_m(x)| \leqslant \dfrac{\theta}{|x^m|}$, θ is some constant independent of x and $B_r(h)$ is the Bernoulli polynomial of degree r and order unity defined by

$$\frac{\tau e^{h\tau}}{e^\tau - 1} = \sum_{r=0}^{\infty} \frac{\tau^r}{r!} B_r(h). \tag{19}$$

Expanding each of the Γ-functions in this manner we obtain

$$\Psi(t) = Q - g(0) - \tfrac{1}{4}(k-1)\,p(p+1)\log(1-2it) + \sum_{r=1}^{n} \frac{\alpha_r}{\mu^r}(1-2it)^{-r} + R_{n+1}(\mu, t), \tag{20}$$

where Q does not contain t and is given by

$$Q = \frac{p(k-1)}{2}\log 2\pi + \frac{p}{2}\left[\sum_{l=1}^{k}\left\{\nu_l - \frac{p+1}{2}\right\}\log\frac{\mu_l}{2} - \left\{k\nu - \frac{p+1}{2}\right\}\log\frac{k\mu}{2}\right], \tag{21}$$

$$\alpha_r = \frac{(-1)^{r+1}(2\mu)^r}{r(r+1)} \sum_{j=0}^{p-1}\left[\sum_{l=1}^{k} \frac{B_{r+1}\left(\dfrac{\beta_l - j}{2}\right)}{\mu_l^r} - \frac{B_{r+1}\left(\dfrac{k\beta - j}{2}\right)}{\mu^r k^r}\right], \tag{22}$$

and $R_{n+1}(\mu, t)$ is defined by (17) and (18).

From (20) we have

$$\Phi(t) = K(1-2it)^{-\frac{1}{2}f} \sum_{v=0}^{\infty} \frac{a_v}{\mu^v}(1-2it)^{-v} + R'_{n+1}(\mu, t) \tag{23}$$

where $K = \exp\{Q - g(0)\}$, $f = \tfrac{1}{2}(k-1)\,p(p+1)$, and a_v is the coefficient of μ^{-v} in the expansion of $\exp\left\{\sum_{r=1}^{n} \alpha_r/\mu^r\right\}$.

The probability density function of ρM is then given by

$$p(\rho M) = \frac{1}{2\pi}\int_{-\infty}^{+\infty} e^{-it\rho M}\Phi(t)\,dt \tag{24}$$

$$= K\sum_{v=0}^{\infty}\frac{a_v}{\mu^v}p(\chi^2_{f+2v}) + R''_{n+1}(\mu, t). \tag{25}$$

The probability that a given value M_0 of the criterion is exceeded is therefore

$$\Pr.\{M \geqslant M_0\} = K\sum_{v=0}^{\infty}\frac{a_v}{\mu^v}P_{f+2v} + R'''_{n+1}(\mu, t), \tag{26}$$

where

$$P_{f+2v} = \int_{\rho M_0}^{\infty} p(\chi^2_{f+2v})\,d\chi^2, \tag{27}$$

$$R'''_{n+1}(\mu, t) = \frac{K}{2\pi}\int_{\rho M_0}^{\infty}\int_{-\infty}^{\infty} e^{-it\rho M}\sum_{v=0}^{\infty}\frac{a_v}{\mu^v}(1-2it)^{-\frac{1}{2}(f+2v)}\left(e^{R_{n+1}(\mu,t)} - 1\right)dt\,d(\rho M)$$

and $p(\chi^2_{f+2v})$ is the probability density function of the χ^2 distribution with $f+2v$ degrees of freedom. For all sufficiently large values of μ, $R'''_{n+1}(\mu, t)$ tends to zero and the required probability will be given with sufficient accuracy by taking a suitable number of terms of the series in (26). Putting $t = 0$ in (20) we have

$$Q-g(0) = -\left\{\sum_{r=1}^{n} (\alpha_r/\mu^r) + R_{n+1}(\mu, 0)\right\}. \tag{28}$$

It is found in practice that by taking a few terms of the series (even in difficult cases usually not more than six), $\exp(-\Sigma\alpha_r/\mu_r)$ is so close to $\exp\{Q-g(0)\} = K$ that direct calculation of that constant is unnecessary.

If we expand $Q-g(0)$ as well as $g(t)$ we obtain instead of (20)

$$\Psi(t) = -\tfrac{1}{2}f\log(1-2it) + \sum_{r=1}^{n}\left\{\frac{\alpha_r}{\mu^r}(1-2it)^{-r}-1\right\} + R_{n+1}(\mu, t) - R_{n+1}(\mu, 0). \tag{29}$$

Proceeding as before we obtain

$$\Pr.\{M \geqslant M_0\} = P_f + \alpha_1(P_{f+2}-P_f)\,1/\mu + \left\{\alpha_2(P_{f+4}-P_f) + \frac{\alpha_1^2}{2!}(P_{f+4}-2P_{f+2}+P_f)\right\}1/\mu^2 + \text{etc.} \tag{30}$$

Thus we may use a suitable number of terms of either of the series given in (26) or (30) to obtain the probability of the criterion exceeding a given value. Formula (26) has been used in this paper, (30) being rather unwieldy if a large number of terms have to be taken.

It should be noted that in the derivation we have used two series, first the asymptotic series for the expansion of the Γ-functions, and then the exponential series. In any particular case we have to decide how many terms we need in the asymptotic series to give a sufficiently close representation of the function and then how many terms we shall use in the exponential series. In those cases investigated here, six terms of the exponential series have nearly always proved adequate as judged by the closeness of agreement between $\Sigma\alpha_r/\mu^r$ and $g(0)-Q$ independently calculated; often fewer terms were necessary, terms in higher powers of $1/\mu$ having negligible effect. In the case of the exponential series the number of terms necessary to represent adequately $\exp\sum_{r}^{n}\alpha_r/\mu^r$ is usually not more than eight, but has sometimes been as many as fourteen. It is mainly in order to keep the number of terms required at this stage within manageable limits that the scale factor ρ is introduced, since by suitable choice of this constant, the values of the α's can be kept small and the number of terms required in the exponential series is consequently less. We see that in effect we are fitting a χ^2 series to the statistic M by arranging that, to the order of accuracy chosen in the asymptotic series, the series will have *all* its cumulants identical with those of M. Before we consider the problem of choosing a suitable value for ρ, we shall derive an expression for the α's and hence the a's in a form which is more suitable for computation.

2·2. *Determination of the α's in a form suitable for calculation*

From the well-known properties of Bernoulli polynomials (see, for example, Milne-Thomson, 1933), we may write the symbolic equality

$$B_r(x+y) \doteqdot (B+x+y)^r, \tag{31}$$

where, after expansion, each index of B is to be replaced by the corresponding suffix. Whence

$$B_{r+1}\left(\frac{\beta_i-j}{2}\right) = \sum_{s=0}^{r+1}\binom{r+1}{s}\left(\frac{\beta_i}{2}\right)^{r+1-s}B_s\left(\frac{-j}{2}\right). \tag{32}$$

probability will be given with sufficient accuracy by taking a suitable number of terms of the series in (26). Putting $t = 0$ in (20) we have

$$Q - g(0) = -\left\{\sum_{r=1}^{n} (\alpha_r/\mu^r) + R_{n+1}(\mu, 0)\right\}. \tag{28}$$

It is found in practice that by taking a few terms of the series (even in difficult cases usually not more than six), $\exp(-\Sigma\alpha_r/\mu_r)$ is so close to $\exp\{Q - g(0)\} = K$ that direct calculation of that constant is unnecessary.

If we expand $Q - g(0)$ as well as $g(t)$ we obtain instead of (20)

$$\Psi(t) = -\tfrac{1}{2}f\log(1-2it) + \sum_{r=1}^{n}\left\{\frac{\alpha_r}{\mu^r}(1-2it)^{-r} - 1\right\} + R_{n+1}(\mu, t) - R_{n+1}(\mu, 0). \tag{29}$$

Proceeding as before we obtain

$$\Pr.\{M \geqslant M_0\} = P_f + \alpha_1(P_{f+2} - P_f)\,1/\mu + \left\{\alpha_2(P_{f+4} - P_f) + \frac{\alpha_1^2}{2!}(P_{f+4} - 2P_{f+2} + P_f)\right\}1/\mu^2 + \text{etc.} \tag{30}$$

Thus we may use a suitable number of terms of either of the series given in (26) or (30) to obtain the probability of the criterion exceeding a given value. Formula (26) has been used in this paper, (30) being rather unwieldy if a large number of terms have to be taken.

It should be noted that in the derivation we have used two series, first the asymptotic series for the expansion of the Γ-functions, and then the exponential series. In any particular case we have to decide how many terms we need in the asymptotic series to give a sufficiently close representation of the function and then how many terms we shall use in the exponential series. In those cases investigated here, six terms of the exponential series have nearly always proved adequate as judged by the closeness of agreement between $\Sigma\alpha_r/\mu^r$ and $g(0) - Q$ independently calculated; often fewer terms were necessary, terms in higher powers of $1/\mu$ having negligible effect. In the case of the exponential series the number of terms necessary to represent adequately $\exp\sum_{r}^{n}\alpha_r/\mu^r$ is usually not more than eight, but has sometimes been as many as fourteen. It is mainly in order to keep the number of terms required at this stage within manageable limits that the scale factor ρ is introduced, since by suitable choice of this constant, the values of the α's can be kept small and the number of terms required in the exponential series is consequently less. We see that in effect we are fitting a χ^2 series to the statistic M by arranging that, to the order of accuracy chosen in the asymptotic series, the series will have *all* its cumulants identical with those of M. Before we consider the problem of choosing a suitable value for ρ, we shall derive an expression for the α's and hence the a's in a form which is more suitable for computation.

2·2. *Determination of the α's in a form suitable for calculation*

From the well-known properties of Bernoulli polynomials (see, for example, Milne-Thomson, 1933), we may write the symbolic equality

$$B_r(x+y) \doteqdot (B + x + y)^r, \tag{31}$$

where, after expansion, each index of B is to be replaced by the corresponding suffix. Whence

$$B_{r+1}\left(\frac{\beta_i - j}{2}\right) = \sum_{s=0}^{r+1}\binom{r+1}{s}\left(\frac{\beta_i}{2}\right)^{r+1-s} B_s\left(\frac{-j}{2}\right). \tag{32}$$

Also if $P(x)$ is a polynomial in x and $P'(x)$ is the differential coefficient of $P(x)$ with respect to x,

$$\sum_{x=0}^{p-1} P'(x) \doteqdot P(B+p)-P(B). \tag{33}$$

Thus if

$$P'(j) = B_s\left(\frac{-j}{2}\right),$$

$$P(j) = -\frac{2}{s+1} B_{s+1}\left(\frac{-j}{2}\right) + \text{constant}, \tag{34}$$

and

$$\sum_{j=0}^{p-1} B_s\left(\frac{-j}{2}\right) \doteqdot \frac{-2}{s+1}\left[B_{s+1}\left\{-\frac{B+p}{2}\right\} - B_{s+1}\left\{-\frac{B}{2}\right\}\right]. \tag{35}$$

If we denote the expression in square brackets by δ_s, we obtain from (32) and (35)

$$\sum_{j=0}^{p-1}\sum_{l=1}^{k} B_{r+1}\left(\frac{\beta_l - j}{2}\right) \doteqdot -2\sum_{l=1}^{k}\sum_{s=0}^{r+1}\binom{r+1}{s}\frac{1}{s+1}\left(\frac{\beta_l}{2}\right)^{r+1-s}\delta_s, \tag{36}$$

and in a similar way we find

$$\sum_{j=0}^{p-1} B_{r+1}\left(\frac{k\beta - j}{2}\right) \doteqdot -2\sum_{s=0}^{r+1}\binom{r+1}{s}\frac{k^{r+1-s}}{s+1}\left(\frac{\beta}{2}\right)^{r+1-s}\delta_s. \tag{37}$$

Whence from (22) we obtain α_r as a polynomial of degree r in β

$$\alpha_r = \frac{(-1)^r k}{r(r+1)(r+2)}\sum_{s=1}^{r+1}\binom{r+2}{s+1} 2^s D_s \beta^{r+1-s}, \tag{38}$$

where

$$D_s = \delta_s \gamma_s,$$

$$\delta_s \doteqdot B_{s+1}\left\{-\frac{B+p}{2}\right\} - B_{s+1}\left\{-\frac{B}{2}\right\}, \tag{39}$$

and

$$\gamma_s = \frac{1}{k}\sum_{l=1}^{k}\left(\frac{\nu}{\nu_l}\right)^{s-1} - \frac{1}{k^s}. \tag{40}$$

It is interesting to note the relation between these quantities and the c's defined by Hartley; if

$$c_s = \sum_l \frac{1}{\nu_l^s} - \frac{1}{N^s}, \quad \gamma_s = k^{-1}\nu^{s-1}c_{s-1}. \tag{41}$$

In the special case when the samples have equal degrees of freedom

$$\gamma_s = 1 - 1/k^s.$$

The values for δ_s for $s = 0, 1, \ldots, 7$, found from equation (39), are given below:

$$\left.\begin{array}{ll}
s & \delta_s \\
0 & -\frac{1}{2}p, \\
1 & \frac{1}{4}p(p+1), \\
2 & -\frac{1}{16}p(2p^2+3p-1), \\
3 & \frac{1}{16}p(p-1)(p+1)(p+2), \\
4 & -\frac{1}{192}p(6p^4+15p^3-10p^2-30p+3), \\
5 & \frac{1}{128}p(p-1)(p+1)(p+2)(2p^2+2p-7), \\
6 & -\frac{1}{768}p(6p^6+21p^5-21p^4-105p^3+21p^2+147p-5), \\
7 & \frac{1}{768}p(p-1)(p+1)(p+2)(3p^4+6p^3-23p^2-26p+62),
\end{array}\right\} \tag{42}$$

and the values for the first six α's from equation (38) are

$$\left.\begin{aligned}
\alpha_1 &= -\tfrac{1}{3}k\{3D_1\beta+2D_2\},\\
\alpha_2 &= \tfrac{1}{6}k\{3D_1\beta^2+4D_2\beta+2D_3\},\\
\alpha_3 &= -\tfrac{1}{15}k\{5D_1\beta^3+10D_2\beta^2+10D_3\beta+4D_4\},\\
\alpha_4 &= \tfrac{1}{60}k\{15D_1\beta^4+40D_2\beta^3+60D_3\beta^2+48D_4\beta+16D_5\},\\
\alpha_5 &= -\tfrac{1}{105}k\{21D_1\beta^5+70D_2\beta^4+140D_3\beta^3+168D_4\beta^2+112D_5\beta+32D_6\},\\
\alpha_6 &= \tfrac{1}{42}k\{7D_1\beta^6+28D_2\beta^5+70D_3\beta^4+112D_4\beta^3+112D_5\beta^2+64D_6\beta+16D_7\}.
\end{aligned}\right\} \quad (43)$$

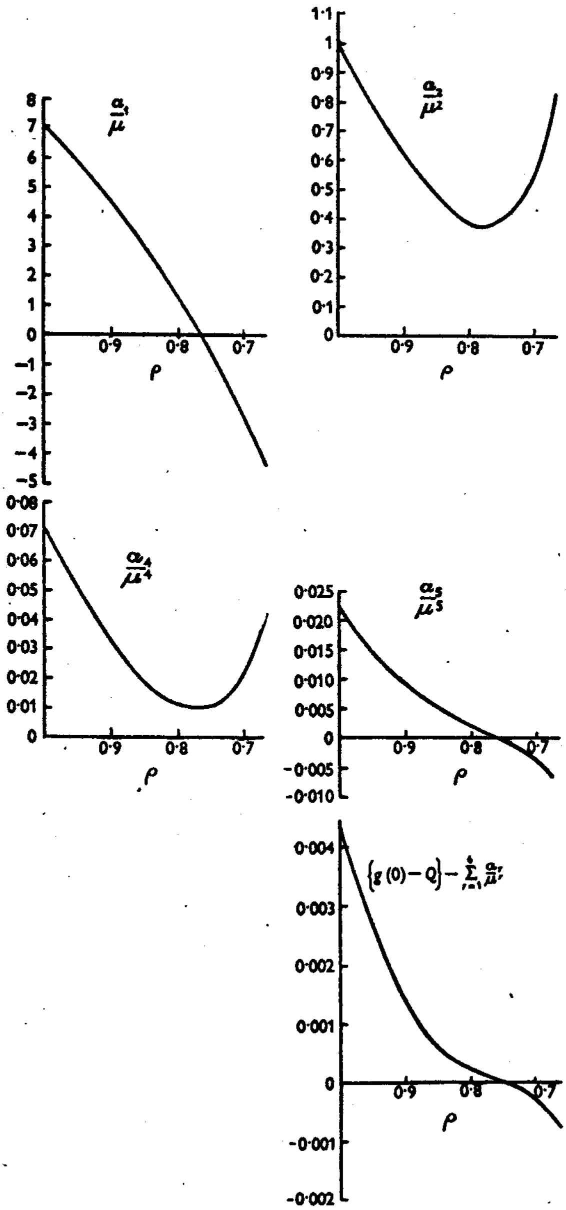

Fig. 1

2·3. *Choice of the value of ρ*

In order that the series should be of practical utility, it must be possible to represent $\exp\{\Sigma\alpha_r\mu^{-r}(1-2it)^{-r}\}$ adequately by a reasonable number of terms of the exponential series; this can be done only if the coefficients α are fairly small. In the *univariate* case, these coefficients will be small even if $\rho = 1$, and in fact if we put $p = 1$ and $\rho = 1$ in equation (26) the series we obtain corresponds exactly with that found by Hartley (1940) using rather a different method of approach. The accuracy of Hartley's series, using only three terms in the asymptotic series, was demonstrated (Hartley & Pearson, 1946) by comparison with the significance levels obtained from Nair's exact expansion; the agreement obtained was good even when the degrees of freedom were as low as three. In the multivariate case, a much more satisfactory series can be obtained if ρ is less than unity.

A typical set of curves showing the values of α_1/μ, α_2/μ^2, ..., α_6/μ^6, and the closeness of agreement between $Q-g(0)$ and $-\sum_{r=1}^{6}\alpha_r/\mu^r$ for varying values of ρ, are plotted in the figure for the case $p = 5$, $k = 5$, $\nu = 9$. The curves all have minima or cross the zero line between $\rho = 0{\cdot}7$ and $\rho = 0{\cdot}8$. The value of ρ which makes α_1 zero is $\rho = 0{\cdot}76296$.

In the calculations carried out here, ρ was chosen so that $\alpha_1 = 0$, since this not only resulted in the other coefficients being small, but the absence of α_1 made the calculation of the a's much easier. Putting $\alpha_1 = 0$ we obtain

$$\rho = 1-\frac{(2p^2+3p-1)}{6(p+1)(k-1)}\left(\Sigma\frac{1}{\nu_l}-\frac{1}{N}\right). \tag{44}$$

2·4. *Example of a calculation using the series*

To check his two working approximations (a) and (b), Bishop used as a standard of reference the values obtained by exact fitting of type I curves to the first two moments of the criterion l_1. In the case $p = 4$, $k = 5$, $\nu = 9$ Bishop found for the 5 % point a value corresponding to $M = 70{\cdot}281$.

To obtain from the series the probability associated with this value, we calculate

$$f = 40,\quad \rho = 0{\cdot}808{,}889,\quad \rho M = 56{\cdot}849{,}5,\quad \mu = \rho\nu = 7{\cdot}28.$$

$r\ (=v)$	α_r/μ^r	a_v/μ^v	P_{f+2v}
0	—	1·000,000	0·040,742
1	0·000,000	0·000,000	—
2	0·143,702	0·143,702	0·092,597
3	0·003,675	0·003,675	0·131,138
4	0·001,793	0·012,118	0·178,763
5	0·000,094	0·000,622	0·235,161
6	0·000,032	0·000,791	0·304,909
7	—	0·000,059	0·369,563
8	—	0·000,044	0·446,178

$\sum_1^6 \alpha_r/\mu^r$	0·149,296	$\sum_1^8 a_v/\mu^v$	1·161,011
$g(0)-Q$	0·149,305	$\exp\{\sum_1^6 \alpha_r/\mu^r\}$	1·161,016
Difference	0·000,009	Difference	0·000,005

$$K = \exp\{Q-g(0)\} = 0{\cdot}861{,}306,\quad \exp\left\{-\sum_1^6\alpha_r\Big/\mu^r\right\} = 0{\cdot}861{,}314$$

$$\Pr.\{M > 70{\cdot}281\} = K\Sigma\{a_v/\mu^v\}P_{f+2v} = 0{\cdot}0492.$$

To illustrate the accuracy with which the asymptotic series represents the function, independent calculations of $g(0)-Q$ have been made. As has already been indicated, however, in practice this rather laborious calculation would not be necessary, K being taken as

$$\exp\{-\Sigma\alpha_r/\mu^r\}.$$

2·5. *Some comparisons between the series and the exact distribution*

For the cases $p = 1, k = 2$ and $p = 2, k = 2$, the exact distribution is known for all values of ν; for $p = 1$ the criterion will simply be a function of the variance ratio, and when $p = 2$ the exact distribution has been found by Pearson & Wilks (1933). Table 1 enables the probabilities obtained from these exact distributions to be compared with those found using the series with scale factor ρ and up to four terms in the asymptotic and exponential series, higher terms having negligible effect. The table shows the values of M corresponding to the 5% and 1% points obtained by Bishop by fitting a type I curve to the first two moments of l_1. The exact probabilities corresponding to these points and those obtained using the series are shown below the values of M.

Table 1. *Comparison of the series with the exact distribution*

		$\nu=9$	$\nu=27$	$\nu=79$
$p=1$, $k=2$	5% point (type I)	4·0499	3·9042	3·8794
	Probability: exact	0·05005	0·05009	0·05002
	series	0·05005	0·05009	0·05003
	1% point (type I)	6·9902	6·7461	6·6991
	Probability: exact	0·00998	0·01001	0·01002
	series	0·00998	0·01001	0·01002
$p=2$, $k=2$	5% point (type I)	8·8801	8·1191	8·0018
	Probability: exact	0·05005	0·04997	0·04979
	series	0·05005	0·04997	0·04979
	1% point (type I)	12·8969	11·7844	11·6074
	Probability: exact	0·00999	0·01000	0·00997
	series	0·00999	0·01000	0·00997

The agreement between the series and the exact values is remarkably good, the series giving five-decimal accuracy in almost every case tested. The more difficult cases, however, are those where p and k are larger, especially when ν is small. For these, the closeness with which $\Sigma\alpha_r/\mu^r$ approaches $g(0)-Q$ and the adequacy, when ρ is suitably chosen, of the exponential series as judged by the comparison of $\exp\{\Sigma(\alpha_r/\mu^r)\}$ and $\Sigma(a_v/\mu^v)$, support belief in the accuracy of this solution. For example, the case $p = 4, k = 5$ which we have used to illustrate the calculation of probabilities from the series, is not a particularly favourable one. It appears, however, that six terms of the asymptotic series and eight of the exponential series will be adequate; in less severe cases of course fewer terms are necessary. Further evidence is supplied later for the accuracy of this type of solution, for in tests of independence to be discussed in § 6, exact distributions are available for comparison, in cases where the series is not favoured, and excellent agreement is found.

APPROXIMATIONS

The series we have found is of rather too complicated a character for routine use; as an alternative, approximations were sought which were relatively simple.

3. Approximations using a single χ^2 distribution

We have for the cumulant generating function of M (putting $\rho = 1$ in equation (20))

$$\Psi(t) = Q - g(0) - \frac{f}{2}\log(1-2it) + \sum_{r=1}^{\infty} \frac{\alpha_r'}{\nu^r}(1-2it)^{-r}, \tag{45}$$

where $f = \frac{1}{2}p(p+1)(k-1)$ and α_r' is obtained by putting $\rho = 1$ (i.e. $\beta = 0$) in equation (43).

Expanding this expression in powers of t we obtain

$$\Psi(t) = \sum_{j=1}^{\infty} \frac{(it)^j}{j!} 2^{j-1}(j-1)!f\left\{1 + \sum_{r=1}^{\infty}\binom{j+r-1}{r}\frac{2r\alpha_r'}{\nu^r f}\right\}. \tag{46}$$

The jth cumulant of M is then given by

$$\kappa_j = 2^{j-1}(j-1)!f\left\{1 + jA_1 + \frac{j(j+1)}{2}A_2 + \dots\right\}, \tag{47}$$

where
$$A_r = \frac{2r\alpha_r'}{\nu^r f}, \tag{48}$$

and in particular for the generalized test for homoscedasticity which we are considering,

$$\left.\begin{aligned} A_1 &= \frac{2p^2+3p-1}{6(k-1)(p+1)}\left(\sum_{l=1}^{k}\frac{1}{\nu_l} - \frac{1}{N}\right), \\ A_2 &= \frac{(p-1)(p+2)}{6(k-1)}\left(\sum_{l=1}^{k}\frac{1}{\nu_l^2} - \frac{1}{N^2}\right). \end{aligned}\right\} \tag{49}$$

3·1. *The choice of a scale factor in the χ^2 approximation*

Now $2^{j-1}(j-1)!f$ is the jth cumulant of χ^2 with f degrees of freedom. Thus, to order ν^{-1}, (47) is identical with the jth cumulant of $C\chi^2$, where C is either $1+A_1$ or $(1-A_1)^{-1}$. If A_2 were zero then $C = 1 + A_1$ would give the first cumulant κ_1 to order ν^{-2} and the remaining cumulants would clearly be less in error than if C were taken as $(1-A_1)^{-1}$. However, if $A_2 = A_1^2$, it would be preferable to put $C = (1-A_1)^{-1}$, since here this form would give agreement to order ν^{-2}.

Clearly this would also be the better form to use if A_2 were near to or greater than A_1^2. In the univariate case $A_2 = 0$ and C should therefore be taken as

$$C = 1 + A_1 = 1 + \frac{1}{3(k-1)}\left(\Sigma\frac{1}{\nu_l} - \frac{1}{N}\right)$$

as has been shown by Bartlett (1937).

For the generalized test for homoscedasticity we find

$$A_2 - A_1^2 = \left(\frac{k}{k-1}\right)^2 \frac{\gamma_2^2}{36(p+1)^2\nu^2}\left\{6(p-1)(p+1)^2(p+2)\left(\frac{\gamma_3}{\gamma_2^2}\frac{k-1}{k}\right) - (2p^2+3p-1)\right\}, \tag{50}$$

where γ_s is defined by equation (40). For $p = 1$, $A_2 = 0$, and consequently this quantity is negative for all values of k. When $p > 1$ it is positive, except in the particular case when $p = 2$ and $k = 2$ and the ν's are equal, when the quantity in curled brackets is equal to -1, and A_1^2 is almost exactly equal to A_2; if the ν's are not equal, this quantity is greater than -1, and it is positive for all larger values of p and k.

For the multivariate statistic, $p > 1$; we therefore take M/C to be approximately distributed as χ^2 with $f = \frac{1}{2}(k-1)p(p+1)$ degrees of freedom and

$$\frac{1}{C} = (1-A_1) = 1 - \frac{(2p^2+3p-1)}{6(p+1)(k-1)}\left(\sum_{l=1}^{k}\frac{1}{\nu_l} - \frac{1}{N}\right);$$

if the degrees of freedom are equal this becomes

$$\frac{1}{C} = 1 - \frac{(2p^2+3p-1)(k+1)}{6(p+1)k\nu}. \tag{51}$$

We note that $1/C$ is the same as the value ρ chosen as scale factor (44) in the series solution. In the case of samples with equal degrees of freedom, the statistic M is equivalent to Bishop's criterion l_1, so that the multivariate scale factor C proposed here is comparable with the scale factor G proposed by Bishop and given in equation (3). Table 2 shows a number of comparisons for the significance levels, together with the values for the probabilities given by the series.

Table 2. χ^2 *approximation; comparisons of scale factors. Significance points for* M *with probability given by series*

			$p=2$		$p=4$		$p=6$	
$k=5$, $\nu=9$	5 %	Bishop (*b*)	23·06	0·0531	67·38	0·0742	142·19	0·1633
		Box	23·27	0·0503	68·93	0·0597	148·30	0·1041
	1 %	Bishop (*b*)	28·82	0·0107	77·01	0·0173	156·35	0·0533
		Box	29·01	0·0101	78·74	0·0135	163·16	0·0286
$k=5$, $\nu=19$	5 %	Bishop (*b*)	22·13	0·0486	61·36	0·0511	121·29	0·0660
		Box	22·03	0·0501	61·31	0·0515	122·82	0·0556
	1 %	Bishop (*b*)	27·34	0·0104	69·95	0·0106	133·60	0·0144
		Box	27·47	0·0100	70·03	0·0105	135·13	0·0116

It appears that, not only is the factor suggested here very much simpler than Bishop's, but that it also gives a better approximation. However, it appears that even with the scale factor C this approximation fails when p is large and ν is small.

4. Approximations using the F distribution

The χ^2 approximation becomes less and less satisfactory as p and k are made larger and ν is made smaller. We know, however, that for all finite p and k, M/C will tend to a type III curve as ν becomes large. When ν is not large we might expect the point corresponding to the distribution of M in the β_1, β_2 plane to lie near the type III line, in either the type I or type VI regions. We shall see that the use of these curves rather than the type III will enable us to absorb a further term in the cumulant series, corresponding to the extra adjustable parameter available with type I and type VI curves, and thus ensure agreement in the cumulants to order ν^{-2}. Although percentage points of the B-function have been tabled (Thompson, 1941), tables of the function F are usually more readily available. For this reason results which occur in the B-function form will be inverted, so that only tables of the F distribution will be required in using these approximations, and they will be referred to as F approximations.

4·1. *Choice of relevant type of curve*

The 'start' of the probability density function for M is at zero. For the Pearson system of frequency curves in which the restriction is made that the start of the curve is at zero, the relation between the cumulants

$$\frac{\kappa_3\kappa_1}{\kappa_2^2} = 2\tau \tag{52}$$

corresponds with Pearson's type III curve when $\tau = 1$. If τ slightly exceeds unity the curve falls in the type VI region, if it is slightly less than unity it falls in the type I region. Substituting the values for the cumulants of the criterion M, using equation (47), we obtain, ignoring terms of order ν^{-3},

$$\tau = \frac{1+4A_1+7A_2+3A_1^2}{1+4A_1+6A_2+4A_1^2}. \tag{53}$$

Thus for all sufficiently large values of ν the region into which the curve will fall is given by

$$\left.\begin{array}{ccc} A_2 > A_1^2 & A_2 = A_1^2 & A_2 < A_1^2 \\ \tau > 1 & \tau = 1 & \tau < 1 \\ \text{Type VI} & \text{Type III} & \text{Type I} \end{array}\right\} \tag{54}$$

For example, from equation (50) obtained in the case of the generalized test for homoscedasticity, it is clear that for $p = 1$ the curve will be in the type I region, and for nearly all other cases, when p is greater than 1, it will be in the type VI region.

4·2. *Type VI*

The F distribution with $2P$ and $2Q$ degrees of freedom is defined by

$$p(F) = \text{constant } F^{P-1}(PF+Q)^{-(P+Q)}. \tag{55}$$

The rth moment of a quantity bF, where b is a constant, is given by

$$\mu_r'(bF) = \left(b\frac{Q}{P}\right)^r \frac{\Gamma(P+r)\,\Gamma(Q-r)}{\Gamma(P)\Gamma(Q)}, \tag{56}$$

from which, after some algebraic reduction, we obtain the first four cumulants of bF as

$$\begin{aligned}
\kappa_1(bF) &= P(b/P)\,(1-1/Q)^{-1},\\
\kappa_2(bF) &= P(b/P)^2\,(1+(P-1)/Q)\,(1-1/Q)^{-2}\,(1-2/Q)^{-1},\\
\kappa_3(bF) &= 2P(b/P)^3\,(1+(P-1)/Q)\,(1+(2P-1)/Q)\,(1-1/Q)^{-3}\,(1-2/Q)^{-1}\,(1-3/Q)^{-1},\\
\kappa_4(bF) &= 6P(b/P)^4\,\{(P/Q)^2\,(5/Q-11/Q^2)+(1-1/Q)^2\,(1-3/Q+2/Q^2+6P/Q-13P/Q^2)\}\\
&\quad\times(1-1/Q)^{-4}\,(1-2/Q)^{-2}\,(1-3/Q)^{-1}\,(1-4/Q)^{-1}.
\end{aligned} \tag{57}$$

Now we have seen that M is approximately distributed as $C\chi^2$, so that if τ is greater than unity we would expect to be able to find values b, P and Q, so that bF would be an even better approximation. Since we already know that the distribution is close to type III, we would further expect that Q will be large compared with P since this will be so for type VI curves close to the type III line.

If then we ignore terms of order $(P/Q)^2$, we find

$$\left.\begin{aligned}\kappa_1(bF) &= P(b/P)\{1 + 1/Q\},\\ \kappa_2(bF) &= P(b/P)^2\{1+P/Q+3/Q\},\\ \kappa_3(bF) &= 2P(b/P)^3\{1+3P/Q+6/Q\},\\ \kappa_4(bF) &= 6P(b/P)^4\{1+6P/Q+10/Q\}.\end{aligned}\right\} \tag{58}$$

Now put $\quad 2P=f_1=f, \quad 2Q=f_2=\dfrac{f_1+2}{A_2-A_1^2} \quad$ and $\quad b=\dfrac{f_1}{1-A_1-f_1/f_2}$,

then we obtain approximately

$$\left.\begin{aligned}\kappa_1(bF) &= f\{1+A_1+A_2\},\\ \kappa_2(bF) &= 2f\{1+2A_1+3A_2\},\\ \kappa_3(bF) &= 8f\{1+3A_1+6A_2\},\\ \kappa_4(bF) &= 48f\{1+4A_1+10A_2\},\end{aligned}\right\} \tag{59}$$

which are identical to order ν^{-2} with the cumulants of M given by equation (47). Thus M/b will be distributed approximately as F with f_1 and f_2 degrees of freedom, where

$$f_1=f, \quad f_2=\frac{f_1+2}{A_2-A_1^2}, \quad b=\frac{f_1}{1-A_1-f_1/f_2}. \tag{60}$$

4·3. *Type I*

We define a quantity X distributed in a type I form with parameters P and Q,

$$p(X) = \text{constant}\, X^{P-1}(1-X)^{Q-1}. \tag{61}$$

The rth moment of bX, where b is a constant, is given by

$$\mu_r'(bX) = b^r\frac{\Gamma(P+r)\,\Gamma(P+Q)}{\Gamma(P)\,\Gamma(P+Q+r)}, \tag{62}$$

from which we find the first four cumulants of bX to be

$$\left.\begin{aligned}\kappa_1(bX) &= P(b/Q)(1+P/Q)^{-1},\\ \kappa_2(bX) &= P(b/Q)^2(1+P/Q)^{-2}(1+(P+1)/Q)^{-1},\\ \kappa_3(bX) &= 2P(b/Q)^3(1-P/Q)(1+P/Q)^{-3}(1+(P+1)/Q)^{-1}(1+(P+2)/Q)^{-1},\\ \kappa_4(bX) &= 6P(b/Q)^4\{1+1/Q-2P/Q-4P/Q^2-2P^2/Q^2+P^2/Q^3+P^3/Q^3\}\\ &\quad\times(1+P/Q)^{-4}(1+(P+1)/Q)^{-2}(1+(P+2)/Q)^{-1}(1+(P+3)/Q)^{-1}.\end{aligned}\right\} \tag{63}$$

As before if Q is large compared with P, so that terms of order $(P/Q)^2$ may be ignored, we obtain

$$\left.\begin{aligned}\kappa_1(bX) &= P(b/Q)\{1-P/Q\},\\ \kappa_2(bX) &= P(b/Q)^2\{1-3P/Q+1/Q\},\\ \kappa_3(bX) &= 2P(b/Q)^3\{1-6P/Q-3/Q\},\\ \kappa_4(bX) &= 6P(b/Q)^4\{1-10P/Q-6/Q\},\end{aligned}\right\} \tag{64}$$

and putting $2P = f_1 = f,\quad 2Q = f_2 = \dfrac{f_1+2}{A_1^2 - A_2}$ and $b = \dfrac{f_2}{1-A_1+2/f_2}$

we again obtain approximately the values given in (59) which to the order of approximation ν^{-2} are the cumulants of M.

Thus M/b will be distributed as X in expression (61) with $2P = f_1$ and $2Q = f_2$ and

$$f_1 = f,\quad f_2 = \frac{f_1+2}{A_1^2 - A_2},\quad b = \frac{f_2}{1-A_1+2/f_2}. \tag{65}$$

Alternatively, $\dfrac{f_2 M}{f_1(b-M)}$ will be distributed as F with f_1 and f_2 degrees of freedom.

We note that although M can vary from 0 to ∞, bX can vary only between the limits 0 and b, so that we are fitting a curve with limited range to one with infinite range. In practice, however, this presents no difficulty (see, for example, the comparisons of Tables 3, 4 and 5), for since the distribution of M will be near to type III, f_2 will be large compared with f_1; consequently b will be large compared with f_1. The mean for such curves will be approximately equal to f_1, so that the range will be large compared with the mean, and the part of the curve ignored by the truncation will be negligible.

4·4. *Application of the F approximation in tests of homoscedasticity*

From (50) we know that when $p = 1$, $A_2 - A_1^2$ is negative, and hence the type I form of the approximation is appropriate. When $p \geqslant 2$ we have seen that, except for the case $p = 2, k = 2$, when to this degree of approximation the curve is almost type III, $A_2 - A_1^2$ is positive and the type VI form is appropriate.

4·41. *Univariate test* ($p = 1$)

When $p = 1$, A_2 is zero, so that to carry out the test we calculate in turn

$$A_1 = \frac{1}{3(k-1)}\left(\Sigma\frac{1}{\nu_t} - \frac{1}{N}\right),\quad f_1 = (k-1),\quad f_2 = \frac{k+1}{A_1^2},\quad b = \frac{f_2}{1-A_1+2/f_2}, \tag{66}$$

and refer $\dfrac{f_2 M}{f_1(b-M)}$ to tables of the F distribution with f_1 and f_2 degrees of freedom.

In the special case when the degrees of freedom are equal

$$A_1 = \frac{k+1}{3\nu k}. \tag{67}$$

To test the accuracy of the approximation we will compare the values it gives for the 5% and 1% points of M, with those obtained from (1) Bartlett's approximation, (2) Bishop & Nair's (1939) values and (3) the χ^2 series given by Hartley and corresponding to equation (26) with $p = 1$ and $\rho = 1$. Tables 3 and 4 are adapted from those given by Pearson & Hartley (1946) with the value of Bartlett's approximation and the present approximation added. In Table 3 a number of comparisons are made for the special case where the degrees of freedom are equal, and Table 4 shows a few comparisons for the case of five estimates of variance with unequal degrees of freedom.

If the accuracy is judged by the closeness of agreement with the values obtained by Bishop & Nair, it appears that the F approximation is an improvement upon that suggested by and putting $2P = f_1 = f, \quad 2Q = f_2 = \dfrac{f_1+2}{A_1^2 - A_2}$ and $b = \dfrac{f_2}{1 - A_1 + 2/f_2}$

we again obtain approximately the values given in (59) which to the order of approximation ν^{-2} are the cumulants of M.

Thus M/b will be distributed as X in expression (61) with $2P = f_1$ and $2Q = f_2$ and

$$f_1 = f, \quad f_2 = \frac{f_1+2}{A_1^2 - A_2}, \quad b = \frac{f_2}{1 - A_1 + 2/f_2}. \tag{65}$$

Alternatively, $\dfrac{f_2 M}{f_1(b-M)}$ will be distributed as F with f_1 and f_2 degrees of freedom.

We note that although M can vary from 0 to ∞, bX can vary only between the limits 0 and b, so that we are fitting a curve with limited range to one with infinite range. In practice, however, this presents no difficulty (see, for example, the comparisons of Tables 3, 4 and 5), for since the distribution of M will be near to type III, f_2 will be large compared with f_1; consequently b will be large compared with f_1. The mean for such curves will be approximately equal to f_1, so that the range will be large compared with the mean, and the part of the curve ignored by the truncation will be negligible.

4·4. *Application of the F approximation in tests of homoscedasticity*

From (50) we know that when $p = 1$, $A_2 - A_1^2$ is negative, and hence the type I form of the approximation is appropriate. When $p \geqslant 2$ we have seen that, except for the case $p = 2$, $k = 2$, when to this degree of approximation the curve is almost type III, $A_2 - A_1^2$ is positive and the type VI form is appropriate.

4·41. *Univariate test* ($p = 1$)

When $p = 1$, A_2 is zero, so that to carry out the test we calculate in turn

$$A_1 = \frac{1}{3(k-1)}\left(\Sigma\frac{1}{\nu_t} - \frac{1}{N}\right), \quad f_1 = (k-1), \quad f_2 = \frac{k+1}{A_1^2}, \quad b = \frac{f_2}{1 - A_1 + 2/f_2}, \tag{66}$$

and refer $\dfrac{f_2 M}{f_1(b-M)}$ to tables of the F distribution with f_1 and f_2 degrees of freedom.

In the special case when the degrees of freedom are equal

$$A_1 = \frac{k+1}{3\nu k}. \tag{67}$$

To test the accuracy of the approximation we will compare the values it gives for the 5% and 1% points of M, with those obtained from (1) Bartlett's approximation, (2) Bishop & Nair's (1939) values and (3) the χ^2 series given by Hartley and corresponding to equation (26) with $p = 1$ and $\rho = 1$. Tables 3 and 4 are adapted from those given by Pearson & Hartley (1946) with the value of Bartlett's approximation and the present approximation added. In Table 3 a number of comparisons are made for the special case where the degrees of freedom are equal, and Table 4 shows a few comparisons for the case of five estimates of variance with unequal degrees of freedom.

If the accuracy is judged by the closeness of agreement with the values obtained by Bishop & Nair, it appears that the F approximation is an improvement upon that suggested by

Table 3. *Comparison of approximations. Significance points for M (equal degrees of freedom, p = 1)*

k	ν	5% Bartlett (χ^2)	5% Box (F)	5% Hartley (series)	5% Bishop & Nair	1% Bartlett (χ^2)	1% Box (F)	1% Hartley (series)	1% Bishop & Nair
3	2	7·32	7·20	7·05	7·11*	11·26	10·85	10·57	10·74*
	3	6·88	6·83	6·79	6·80†	10·57	10·41	10·32	10·43†
	4	6·66	6·63	6·61	6·62*	10·23	10·14	10·10	10·13*
	9	6·29	6·29	6·28	6·30†	9·67	9·64	9·64	9·67†
5	2	11·39	11·23	11·01	11·09*	15·93	15·52	15·15	15·32*
	3	10·75	10·69	10·62	10·67†	15·05	14·88	14·76	14·91†
	4	10·44	10·39	10·37	10·38*	14·60	14·42	14·46	14·47*
	9	9·91	9·89	9·90	9·93†	13·87	13·85	13·84	13·86†
10	2	20·02	19·68	19·45	19·62*	25·68	25·22	24·65	24·90*
	3	18·99	18·91	18·79	18·82†	24·31	24·12	23·97	24·09†
	4	18·47	18·41	18·38	18·42*	23·65	23·54	23·49	23·34*
	9	17·61	17·61	17·60	17·64†	22·69	22·58	22·53	22·48†

* Calculated from Nair's exact distribution. † Calculated by fitting type I curve to L_1.

Table 4. *Comparison of approximations. Significance points for M (unequal degrees of freedom, p = 1, k = 5)*

N	ν_1	ν_2	ν_3	ν_4	ν_5	5% Bartlett (χ^2)	5% Box (F)	5% Hartley (series)	5% Bishop & Nair	1% Bartlett (χ^2)	1% Box (F)	1% Hartley (series)	1% Bishop & Nair
20	6	6	4	2	2	10·70	10·65	10·54	10·59	14·97	14·82	14·62	14·80
45	16	16	9	2	2	10·45	10·40	10·30	10·35	14·62	14·51	14·31	14·46
20	5	5	4	3	3	10·49	10·46	10·41	10·43	14·68	14·58	14·51	14·59
45	14	14	9	4	4	10·07	10·05	10·04	10·05	14·09	14·05	14·03	14·05

Bartlett and is about as accurate as Hartley's series, whilst it requires no special tables and involves only simple calculations.

Since the approximations proposed by Bartlett, Hartley and the present author are essentially asymptotic, it is to be expected that for small values of ν, and particularly when $\nu = 1$, the approximations will break down. This does in fact happen to a certain extent with all of them, but it seems least serious with the present F approximation; for example, when $k = 4$, $\nu = 1$, we have

Approximation	5% point	1% point
Bartlett (χ^2)	11·1	16·1
Hartley (series)	9·0	11·8
Box (F)	10·3	14·6
Nair's expansion	10·0	14·1

For the case $\nu = 1$, Table 5 compares, for a number of values of k, the 5% and 1% levels given by Bartlett's approximation and by the present method with values obtained by Bishop & Nair (1939) using Nair's expansion.

Table 5. *Comparison of the approximations when $\nu = 1$. Significance points for M*

Value of k		2	3	4	5	6	7	8	9	10
5% point	Bartlett (χ^2)	5·8	8·7	11·1	13·3	15·4	17·4	19·3	21·3	23·1
	Box (F)	5·1	7·9	10·3	12·6	14·6	16·7	18·6	20·5	22·4
	Nair's expansion	5·1	7·7	10·0	12·0	14·1	15·9	17·9	19·6	21·3
1% point	Bartlett (χ^2)	10·0	13·3	16·1	18·6	21·0	23·2	25·4	27·5	29·6
	Box (F)	7·9	11·3	14·6	17·1	19·2	21·5	23·7	25·8	27·9
	Nair's expansion	8·3	11·5	14·0	16·5	18·9	21·0	23·1	25·2	27·2

4·42. *Multivariate test $p \geqslant 2$*

To carry out the test we calculate the quantities

$$\left.\begin{aligned} A_1 &= \frac{2p^2+3p-1}{6(k-1)(p+1)}\left(\Sigma\frac{1}{\nu_l}-\frac{1}{N}\right), \quad A_2 = \frac{(p-1)(p+2)}{6(k-1)}\left(\Sigma\frac{1}{\nu_l^2}-\frac{1}{N^2}\right), \\ f_1 &= \tfrac{1}{2}(k-1)p(p+1), \quad f_2 = \frac{f_1+2}{A_2-A_1^2}, \quad b = \frac{f_1}{1-A_1-f_1/f_2}, \end{aligned}\right\} \tag{68}$$

and refer M/b to the tables of the F distribution with f_1 and f_2 degrees of freedom.

When the degrees of freedom are equal

$$A_1 = \frac{(p^2+3p-1)(k+1)}{6(p+1)k\nu}, \quad A_2 = (p-1)(p+2)\frac{(k^2+k+1)}{6k^2\nu^2}. \tag{69}$$

George (1945) was able to evaluate the exact distribution of the generalized L_1 statistic in simple cases, although, when the value of p and k are not very small, the method becomes unmanageable. She used her exact distribution to check Bishop's approximations. Table 6 is taken for George's Table 1 and shows the equivalent value of M obtained by Bishop's empirical formula, method (*a*), for the 5% point, together with the exact value of the probability obtained by George by direct integration. The probability corresponding to this value of M has also been calculated by the χ^2 and F approximations suggested here. Thus the closeness with which exact probability approaches 0·0500 indicates the accuracy of Bishop's method, and the closeness with which the probabilities for the χ^2 and F approximations coincide with the exact probability measures the accuracy of these approximations.

We see that the values given by the F approximation are in excellent agreement with the exact probabilities, and even the χ^2 approximation is considerably better than Bishop's method. Unfortunately, no exact values are available in the cases where p and k are larger, when approximation to the curve is more difficult. For these distributions the series given by formula (26), using in most cases up to six* terms in the asymptotic series and up to eight*

* When $\nu = 9$, and $p = 5$ and 6, the coefficients α are rather large, and ten and fourteen terms respectively had to be used in the exponential series. When $p = 6$ there is evidence that further terms in the asymptotic series would give closer agreement.

Table 6. *5 % points for M given by Bishop's empirical approximation with their associated probabilities calculated by:* (1) *George's exact method,* (2) *the F approximation,* (3) *the* χ^2 *approximation*

	ν	M	Probabilities: Exact (George)	Probabilities: F (Box)	Probabilities: χ^2 (Box)
$p=2$, $k=2$	9	8·924	0·0492	0·0492	0·0492
	14	7·835	0·0495*	0·0494	0·0496
	19	7·831	0·0496*	0·0496	0·0496
	24	7·828	0·0498*	0·0497	0·0497
$p=2$, $k=3$	9	14·164	0·0491	0·0491	0·0490
	19	13·285	0·0496	0·0497	0·0496
	29	13·031	0·0497	0·0499	0·0499
$p=3$, $k=2$	9	15·740	0·0461	0·0458	0·0446
	14	14·434	0·0475	0·0475	0·0470
	24	13·598	0·0485	0·0486	0·0485
	29	13·416	0·0488	0·0489	0·0487
	39	13·211	0·0488	0·0489	0·0488
$p=3$, $k=3$	14	23·661	0·0481	0·0479	0·0473
	29	22·288	0·0484	0·0480	0·0478
$p=4$, $k=2$	19	20·989	0·0461	0·0461	0·0455
	29	19·946	0·0477	0·0478	0·0476

* These values have been recalculated and do not agree with the values given in George's table.

Table 7. *Comparisons of approximations. Significance points for M obtained by various methods, with probabilities given by series* (26)

			$p=2$		$p=3$		$p=4$		$p=5$		$p=6$	
$k=5$, $\nu=9$	5 %	Bishop (a)	23·40	0·0485	—	—	71·07	0·0434	—	—	173·17	0·0105
		Bishop (b)	23·06	0·0531	—	—	67·38	0·0742	—	—	142·19	0·1633
		Box (χ^2)	23·27	0·0503	42·56	0·0532	68·93	0·0597	103·65	0·0673	148·30	0·1041
		Box (F)	23·30	0·0500	42·83	0·0506	69·84	0·0524	106·40	0·0545	153·36	0·0692
		Type I	23·26	0·0504	42·88	0·0502	70·28	0·0492	107·15	0·0500	157·38	0·0488
	1 %	Bishop (a)	29·19	0·0096	—	—	81·35	0·0082	—	—	192·34	0·0010
		Bishop (b)	28·82	0·0107	—	—	77·01	0·0173	—	—	156·35	0·0533
		Box (χ^2)	29·01	0·0101	50·24	0·0111	78·74	0·0135	113·54	0·0225	163·16	0·0286
		Box (F)	29·05	0·0100	50·58	0·0102	79·84	0·0105	118·56	0·0122	168·94	0·0165
		Type I	29·07	0·0099	50·59	0·0102	80·45	0·0097	120·20	0·0098	173·78	0·0097
$k=5$, $\nu=19$	5 %	Bishop (a)	22·14	0·0486	—	—	61·63	0·0489	—	—	124·25	0·0469
		Bishop (b)	22·13	0·0486	—	—	61·36	0·0511	—	—	121·29	0·0660
		Box (χ^2)	22·03	0·0501	39·09	0·0506	61·31	0·0515	89·08	0·0532	122·82	0·0556
		Box (F)	22·04	0·0500	39·14	0·0501	61·47	0·0502	89·47	0·0505	123·61	0·0508
		Type I	21·92	0·0516	39·10	0·0505	61·36	0·0511	89·59	0·0496	123·70	0·0503
	1 %	Bishop (a)	27·55	0·0098	—	—	70·50	0·0095	—	—	137·09	0·0088
		Bishop (b)	27·34	0·0104	—	—	69·95	0·0106	—	—	133·60	0·0144
		Box (χ^2)	27·47	0·0100	46·14	0·0102	70·03	0·0105	99·56	0·0109	135·13	0·0116
		Box (F)	27·48	0·0100	46·20	0·0100	70·23	0·0101	100·01	0·0101	136·03	0·0103
		Type I	27·55	0·0098	46·24	0·0099	70·27	0·0100	100·24	0·0098	136·32	0·0099

terms in the exponential series, may be used as a standard for comparison. Table 7 sh
significance points for M obtained by five different methods together with the prob
calculated from the series. The methods are: Bishop's empirical approximation (*a*), B
approximation (*b*), the χ^2 and F approximations suggested in this paper, and the fi
a type I curve by exact calculation of the first two moments. The values for M for B
approximations and the type I approximation have been calculated from Bishop's sign
points for l_1 given in his Tables 9 and 10.

If we take the series solution as supplying essentially accurate values, we confirm B
suggestion that the type I curve, fitted exactly to the first two moments of l_1, prov
exceedingly good approximation. Of the working approximations, the F approxi
suggested here appears to be the best and the χ^2 approximation with the generalize
factor C will be fairly satisfactory if p and k are not greater than five and ν is not les
say, twenty.

Table 8 supplies a few comparisons with equal and unequal degrees of freedom.

Table 8. *Significance points for M from χ^2 and F approximations for some equal and u groupings, when $p = 4$ and $k = 5$, with associated probability given by series (26*

N	ν_1	ν_2	ν_3	ν_4	ν_5		5%		1%	
95	19	19	19	19	19	χ^2 F	61·31 61·47	0·0515 0·0502	70·03 70·23	0·0 0·0
95	9	9	19	29	29	χ^2 F	63·22 63·99	0·0578 0·0521	72·33 73·14	0·0 0·0
95	9	9	9	9	59	χ^2 F	66·32 67·39	0·0627 0·0535	75·76 77·07	0·0 0·0
45	9	9	9	9	9	χ^2 F	68·93 69·84	0·0597 0·0524	78·74 79·84	0·0 0·0

It appears, at least for unequal samples with none of the degrees of freedom less t
that the F approximation will be fairly satisfactory.

5. Generalization of the procedure

The method we have developed has so far been illustrated in the case of the univariat
multivariate tests of homoscedasticity; its application is, however, more genera
fact, the method can be used whenever, by choosing a suitable power of the original crit
we can obtain a statistic W which has its hth moment of the form

$$\mathscr{E}(W)^h = \text{constant} \times \left[\frac{\prod_{j=1}^{k}(y_j^{y_j})}{\prod_{i=1}^{m}(x_i^{x_i})}\right]^h \frac{\prod_{i=1}^{m}[\Gamma\{x_i(1+h)+\xi_i\}]}{\prod_{j=1}^{k}[\Gamma\{y_j(1+h)+\eta_j\}]},$$

where $$\sum_{i=1}^{m} x_i = \sum_{j=1}^{k} y_j.$$

(The constant will be obtained of course by putting $h = 0$ in (70) and taking the reciprocal.) Many of the tests in Plackett's review, referred to in the introduction to this paper, fall into this category. We have already seen that the generalized L_1 statistic is of this type; others are Wilks's test for the independence of k groups of variates (which has some important special cases; and will be considered in detail in the next section); the generalized test for constancy of means, variances and covariances for k samples given by Wilks (1932); and the tests for 'compound symmetry' of variance-covariance matrices discussed by Votaw (1948).

Another group of criteria, which has been studied by Mauchly (1940) and Wilks (1946), arises from tests made on a single sample of n p-variate observations. Mauchly's criterion tests the hypothesis that the variances of the variates are all equal and that the covariances between the variates are all zero. Wilks considered criteria for testing three further hypotheses:

(*a*) That the p means, p variances, and $\frac{1}{2}p(p-1)$ covariances for the variates have respectively the same unknown values.

(*b*) That the variances are the same and the covariances are the same irrespective of what values the means have.

(*c*) That the means are the same (assuming (*b*) true).

It is hoped to consider some of these tests rather more closely in a later paper. Here we shall merely note that, except for Wilks's third criterion (which is always distributed exactly in type I form), the exact distribution of the test function is, in general, not exactly known. The expression for the hth moment, however, is in each case of the form of equation (70) and, as is shown below, our previous approach will provide approximations in all these cases. Tukey & Wilks (1946) have considered this class of statistics and have pointed out that they all possess in common the property that, when the null hypothesis is true, they are distributed as a product of independent components, each component being distributed in type I form.

Consider the expression (70) for the hth moment of any statistic W of this type. If we take $M = -2\log W$ as our working statistic, and write $(1-\rho)x_i = \beta_i$, $(1-\rho)y_j = \epsilon_j$, where ρ is a constant $\leqslant 1$ at our choice, we find for the cumulant generating function of ρM

$$\Psi(t) = g(t) - g(0),$$

where $$g(t) = 2it\rho\left[\sum_{i=1}^{m} x_i \log x_i - \sum_{j=1}^{k} y_j \log y_j\right]$$
$$+\sum_{i=1}^{m} \log\Gamma\{\rho x_i(1-2it)+\beta_i+\xi_i\} - \sum_{j=1}^{k}\log\Gamma\{\rho y_j(1-2it)+\epsilon_j+\eta_j\}, \tag{71}$$

and $g(0)$ is independent of t and is obtained by writing $t = 0$ in (71). Expanding the logarithms of the Γ-functions by (18), we obtain the cumulant generating function of ρM in the form

$$\Psi(t) = Q - g(0) - \frac{f}{2}\log(1-2it) + \sum_{r=1}^{\infty} \omega_r (1-2it)^{-r}, \tag{72}$$

where
$$f = -2\left\{\sum_{i=1}^{m} \xi_i - \sum_{j=1}^{k} \eta_j - \tfrac{1}{2}(m-k)\right\}, \tag{73}$$

$$\omega_r = \frac{(-1)^{r+1}}{r(r+1)}\left\{\sum_{i=1}^{m} \frac{B_{r+1}(\beta_i+\xi_i)}{(\rho x_i)^r} - \sum_{j=1}^{k} \frac{B_{r+1}(\epsilon_j+\eta_j)}{(\rho y_j)^r}\right\}, \tag{74}$$

$$Q = \tfrac{1}{2}(m-k)\log 2\pi - \frac{f}{2}\log\rho + \sum_{i=1}^{m}(x_i+\xi_i-\tfrac{1}{2})\log x_i - \sum_{j=1}^{k}(y_j+\eta_j-\tfrac{1}{2})\log y_j. \tag{75}$$

From the cumulant generating function (72), the asymptotic χ^2 series corresponding to (26) and (30) are immediately obtainable. Alternatively, we may obtain approximations in the manner given in § 3; the method outlined there is clearly perfectly general for this whole class of statistics. We need the quantities $A_1 = 2\omega_1'/f$ and $A_2 = 4\omega_2'/f$, where ω_r' is the value taken on by ω_r when $\rho = 1$. The scale factor C for the χ^2 approximation will be $1+A_1$ or $(1-A_1)^{-1}$; a decision between the two alternative forms can be reached by the considerations set out in § 3·1. Then, to this order of approximation, M/C will be distributed as χ^2. If greater accuracy is required we may use the F type approximations described in § 4. The particular form is decided by the sign of the quantity $A_2 - A_1^2$. If this quantity is positive, the curve of best fit will be type VI. Putting

$$f_1 = f, \quad f_2 = \frac{f_1+2}{A_2-A_1^2}, \quad b = \frac{f_1}{1-A_1-f_1/f_2}, \tag{76}$$

M/b is distributed approximately as the variance ratio F with f_1 and f_2 degrees of freedom. Alternatively, if $A_2 - A_1^2$ is negative the best fitting curve will be type I, and if we put

$$f_1 = f, \quad f_2 = \frac{f_1+2}{A_1^2-A_2}, \quad b = \frac{f_2}{1-A_1+2/f_2}, \tag{77}$$

then approximately $\dfrac{f_2 M}{f_1(b-M)}$ will be distributed as F with f_1 and f_2 degrees of freedom.

There are thus a number of possible levels of approximation as measured by the order of agreement between the cumulants of the statistic and those of the fitted curve.

(1) Ignoring terms of order x_i^{-1}, y_j^{-1}, M is distributed as χ^2.

(2) Ignoring terms of order x_i^{-2}, y_j^{-2}; by a technique originally used by Bartlett and here generalized, a quantity C can be found such that M is distributed as $C\chi^2$.

(3) Ignoring terms of order x_i^{-3}, y_j^{-3}, a function of M can be obtained which is distributed as the variance ratio F.

(4) Finally, for very precise work and for checking other approximations, a χ^2 series solution may be used and here agreement with the cumulants (as represented by their asymptotic expansions) of the statistic can be obtained to as great an order as seems profitable.

In practice method (4) is sometimes rather long, although it has been found very accurate, but (3) involves very little labour and will often be sufficiently precise.

As a second example of the application of this technique we consider Wilks's generalized test of independence.

6. The generalized test for independence

Wilks (1935) considered the following problem: suppose we have a sample of $\nu+u$ observations for a kp variate normal population and we have some *a priori* reason for dividing the variates into k groups containing $p_1, \ldots, p_n, \ldots, p_l, \ldots, p_k$ variates (where $\sum_l p_l = kp$ and p is thus the average size of the groups and is not necessarily integer). It is required to test

where
$$f = -2\left\{\sum_{i=1}^{m}\xi_i - \sum_{j=1}^{k}\eta_j - \tfrac{1}{2}(m-k)\right\}, \tag{73}$$

$$\omega_r = \frac{(-1)^{r+1}}{r(r+1)}\left\{\sum_{i=1}^{m}\frac{B_{r+1}(\beta_i+\xi_i)}{(\rho x_i)^r} - \sum_{j=1}^{k}\frac{B_{r+1}(\epsilon_j+\eta_j)}{(\rho y_j)^r}\right\}, \tag{74}$$

$$Q = \tfrac{1}{2}(m-k)\log 2\pi - \frac{f}{2}\log\rho + \sum_{i=1}^{m}(x_i+\xi_i-\tfrac{1}{2})\log x_i - \sum_{j=1}^{k}(y_j+\eta_j-\tfrac{1}{2})\log y_j. \tag{75}$$

From the cumulant generating function (72), the asymptotic χ^2 series corresponding to (26) and (30) are immediately obtainable. Alternatively, we may obtain approximations in the manner given in § 3; the method outlined there is clearly perfectly general for this whole class of statistics. We need the quantities $A_1 = 2\omega_1'/f$ and $A_2 = 4\omega_2'/f$, where ω_r' is the value taken on by ω_r when $\rho = 1$. The scale factor C for the χ^2 approximation will be $1+A_1$ or $(1-A_1)^{-1}$; a decision between the two alternative forms can be reached by the considerations set out in § 3·1. Then, to this order of approximation, M/C will be distributed as χ^2. If greater accuracy is required we may use the F type approximations described in § 4. The particular form is decided by the sign of the quantity $A_2 - A_1^2$. If this quantity is positive, the curve of best fit will be type VI. Putting

$$f_1 = f, \quad f_2 = \frac{f_1+2}{A_2-A_1^2}, \quad b = \frac{f_1}{1-A_1-f_1/f_2}, \tag{76}$$

M/b is distributed approximately as the variance ratio F with f_1 and f_2 degrees of freedom. Alternatively, if $A_2 - A_1^2$ is negative the best fitting curve will be type I, and if we put

$$f_1 = f, \quad f_2 = \frac{f_1+2}{A_1^2-A_2}, \quad b = \frac{f_2}{1-A_1+2/f_2}, \tag{77}$$

then approximately $\dfrac{f_2 M}{f_1(b-M)}$ will be distributed as F with f_1 and f_2 degrees of freedom.

There are thus a number of possible levels of approximation as measured by the order of agreement between the cumulants of the statistic and those of the fitted curve.

(1) Ignoring terms of order x_i^{-1}, y_j^{-1}, M is distributed as χ^2.

(2) Ignoring terms of order x_i^{-2}, y_j^{-2}; by a technique originally used by Bartlett and here generalized, a quantity C can be found such that M is distributed as $C\chi^2$.

(3) Ignoring terms of order x_i^{-3}, y_j^{-3}, a function of M can be obtained which is distributed as the variance ratio F.

(4) Finally, for very precise work and for checking other approximations, a χ^2 series solution may be used and here agreement with the cumulants (as represented by their asymptotic expansions) of the statistic can be obtained to as great an order as seems profitable.

In practice method (4) is sometimes rather long, although it has been found very accurate, but (3) involves very little labour and will often be sufficiently precise.

As a second example of the application of this technique we consider Wilks's generalized test of independence.

6. The generalized test for independence

Wilks (1935) considered the following problem: suppose we have a sample of $\nu+u$ observations for a kp variate normal population and we have some *a priori* reason for dividing the variates into k groups containing $p_1, \ldots, p_n, \ldots, p_l, \ldots, p_k$ variates (where $\sum_l p_l = kp$ and p is thus the average size of the groups and is not necessarily integer). It is required to test the hypothesis that the k groups of residuals, obtained after fitting u independent constants to each of the variates, are mutually independent.

If $|c_{ij}|$ is the $kp \times kp$ determinant of sums of squares and products of residuals for the kp variates and $|c_{ij}|_l$ is the $p_l \times p_l$ determinant of sums of squares and products of residuals of the lth group, then the likelihood ratio criterion obtained by Wilks is

$$\lambda = \frac{|c_{ij}|}{\prod_{l=1}^{k} |c_{ij}|_l} = \frac{|r_{ij}|}{\prod_{l=1}^{k} |r_{ij}|_l}, \tag{78}$$

where $|r_{ij}|$ and $|r_{ij}|_l$ are the corresponding determinants of sample correlation coefficients having ν degrees of freedom. Wilks obtained the moments, and also, for special sets of values of k and p_l, the exact distribution of his criterion which generalizes a very large class of statistical tests. Problems in which there are more than two groups of variates, i.e. where $k > 2$, occur for example in educational research; we may have some prior reason for believing that a battery of, say, ten different tests applied to pupils may be divided up into a number of groups, each group concerned with some distinct ability, and may wish therefore to test the hypothesis that, when the means are eliminated, the selected groups are independent of each other.

When $k = 2$, we consider only two groups of variates containing p_1 in the first and p_2 in the second. Since the criterion and its distribution will be unaffected if the set of p_2 variates are fixed independent variables and the set of p_1 variates 'dependent' variables distributed in a p_1-variate normal distribution, the function is then appropriate for testing the general multivariate linear hypothesis (see, for example, Bartlett, 1934, 1938, 1947). If, in addition, $p_1 = 1$, then the likelihood criterion is $\lambda = 1 - R^2$, where R is the coefficient of multiple correlation between the single dependent variate and the p_2 independent variates. A second special case of Wilks's statistic which is of some interest, and is considered more fully later in this section, occurs when there is only one variate in each of the k groups. The statistic then supplies an overall test for independence between k variates. For the general statistic Wald & Brookner (1941), using a rather different technique from that of Wilks, were able to extend the catalogue of values of k and p_l for which the distribution of λ is exactly known in terms of elementary functions, to include all cases where at most one group contains an odd number of variates. These distributions, although exact, are rather complicated in character. As an alternative and to cover the remaining cases, these authors obtained a series solution and Rao (1948) modified this series in the important case where $k = 2$ to provide an improved test in problems of multivariate analysis. These series will later appear as special cases of that which we are now investigating.

6·1. *Derivation of the series*

The hth moment of λ is given by Wilks as

$$\prod_{l=1}^{k} \prod_{j=0}^{p_l-1} \left\{ \frac{\Gamma\left(\frac{\nu-j}{2}\right)}{\Gamma\left(\frac{\nu-j}{2}+h\right)} \right\} \prod_{m=0}^{kp-1} \left\{ \frac{\Gamma\left(\frac{\nu-m}{2}+h\right)}{\Gamma\left(\frac{\nu-m}{2}\right)} \right\}. \tag{79}$$

So that if we write $W = \lambda^{\frac{1}{2}\nu}$, the hth moment of W will be in the form given in equation (70); taking as our logarithmic statistic $M = -2 \log W$ we obtain

$$M = -\nu \log \lambda. \tag{80}$$

To obtain the series, we begin as before by defining the relationship between a quantity ρ (less than unity) and quantities μ and β by the equations $\rho = \mu/\nu$, $\nu = \mu+\beta$. It is also convenient to define a set of quantities

$$\Sigma_s = \left(\sum_l p_l\right)^s - \sum_l p_l^s, \tag{81}$$

which appear in the solution in much the same way as the quantities $\Sigma \frac{1}{\nu_l^s} - \frac{1}{N^s}$ appear in the tests for homoscedasticity. Then, as before, we obtain equation (72) for the cumulant generating function of the ρM, and the constants are available by direct substitution in (73), (74) and (75),

$$f = \tfrac{1}{2}\Sigma_2, \tag{82}$$

$$\omega_r = \frac{\alpha_r}{\mu^r} = \frac{(-2)^r}{r(r+1)\mu^r} \sum_{l=2}^{k} \sum_{j=0}^{p_l-1} \left\{ B_{r+1}\left(\frac{\beta-j}{2}\right) - B_{r+1}\left(\frac{\beta - \sum_{n=1}^{l-1} p_n - j}{2}\right)\right\}, \tag{83}$$

$$Q = -\frac{f}{2}\log\frac{\mu}{2}. \tag{84}$$

The calculation of α_r from formula (83) would clearly be extremely laborious for all but small numbers of variates; we therefore seek an alternative simpler form. Using relations (32) and (35), we find

$$\alpha_r = \frac{(-1)^{r+1}}{r(r+1)(r+2)} \sum_{s=0}^{r+1} \binom{r+2}{s+1} 2^s \sum_{l=2}^{k} \left\{ \beta^{r+1-s} - \left(\beta - \sum_{n=1}^{l-1} p_n\right)^{r+1-s}\right\} \delta_s(p_l), \tag{85}$$

where
$$\delta_s(p_l) \doteqdot B_{s+1}\left(-\frac{B+p_l}{2}\right) - B_{s+1}\left(-\frac{B}{2}\right), \tag{86}$$

and the values taken by (86) when $p = 1, 2, \ldots, 7$, are given by putting $p = p_l$ in equation (42). Writing α'_r for the value which α_r has when $\rho = 1$, i.e. when $\beta = 0$, then substituting for $\delta_s(p_l)$ in (85) and summing, we obtain for the first six values of α'_r:

$$\left.\begin{aligned}
\alpha'_1 &= \tfrac{1}{24}\{2\Sigma_3 + 3\Sigma_2\},\\
\alpha'_2 &= \tfrac{1}{48}\{\Sigma_4 + 2\Sigma_3 - \Sigma_2\},\\
\alpha'_3 &= \tfrac{1}{720}\{6\Sigma_5 + 15\Sigma_4 - 10\Sigma_3 - 30\Sigma_2\},\\
\alpha'_4 &= \tfrac{1}{480}\{2\Sigma_6 + 6\Sigma_5 - 5\Sigma_4 - 20\Sigma_3 + 3\Sigma_2\},\\
\alpha'_5 &= \tfrac{1}{840}\{2\Sigma_7 + 7\Sigma_6 - 7\Sigma_5 - 35\Sigma_4 + 7\Sigma_3 + 49\Sigma_2\},\\
\alpha'_6 &= \tfrac{1}{2016}\{3\Sigma_8 + 12\Sigma_7 - 14\Sigma_6 - 84\Sigma_5 + 21\Sigma_4 + 196\Sigma_3 - 10\Sigma_2\},
\end{aligned}\right\} \tag{87}$$

where Σ_s is defined by equation (81), whence we have for the α's

$$\left.\begin{aligned}
\alpha_1 &= \alpha'_1 - (f/2)\beta,\\
\alpha_2 &= \alpha'_2 - \alpha'_1\beta + (f/4)\beta^2,\\
\alpha_3 &= \alpha'_3 - 2\alpha'_2\beta + \alpha'_1\beta^2 - (f/6)\beta^3,\\
\alpha_4 &= \alpha'_4 - 3\alpha'_3\beta + 3\alpha'_2\beta^2 - \alpha'_1\beta^3 + (f/8)\beta^4,\\
\alpha_5 &= \alpha'_5 - 4\alpha'_4\beta + 6\alpha'_3\beta^2 - 4\alpha'_2\beta^3 + \alpha'_1\beta^4 - (f/10)\beta^5,\\
\alpha_6 &= \alpha'_6 - 5\alpha'_5\beta + 10\alpha'_4\beta^2 - 10\alpha'_3\beta^3 + 5\alpha'_2\beta^4 - \alpha'_1\beta^5 + (f/12)\beta^6.
\end{aligned}\right\} \tag{88}$$

As before, from the cumulant generating function we obtain the series corresponding to (26) and (30), and if ρ is chosen so that $\alpha_1 = 0$, we have

$$\beta = 2\alpha_1'/f, \quad \rho = 1 - \frac{1}{12\nu f}(2\Sigma_3 + 3\Sigma_2). \tag{89}$$

Wald & Brookner (1941) derived a χ^2 series for this statistic by a different method from that used here; it is not difficult to show, however, that the series they obtained is equivalent to our series, but with $\rho = 1$. In this form the series is of little practical use for small, or even moderate values of ν because of the difficulty we have noted before of adequately approximating to $\exp \Sigma\alpha_r\mu^{-r}(1-2it)^{-r}$ by means of a series, unless α_r is small or μ is large. By introducing the factor ρ, the size of the coefficients α can be greatly reduced and the series be used even for fairly small values of ν. As an example, consider the case of three groups of variates with two variates in each grouping, $k = 3$, $p_1 = 2$, $p_2 = 2$, $p_3 = 2$, and suppose $\nu = 10$. The values of the coefficients α_r/μ^r are shown below when $\rho = 1$ and also when ρ takes on a value making α_1 zero. When $\rho = 1$, μ is of course equal to ν.

Values of α_r/μ^r

r	$\rho = 1$	$\rho = 0{\cdot}68\dot{3}$
1	1·900,00	0·000,00
2	0·335,00	0·073,17
3	0·086,33	0·003,71
4	0·026,88	0·001,78
5	0·009,40	0·000,23
6	0·003,55	0·000,07
Total	2·361,16	0·078,96
$g(0)-Q$	2·363,61	0·078,98

For the Wald & Brookner series, if ν is small, the coefficients are so large that in practice it would be impossible to represent the exponent adequately by a reasonably small number of terms of the exponential series; by suitably choosing ρ, however, the size of the coefficients are greatly reduced while the agreement between the sum of the terms and $g(0)-Q$ is improved. In the particular example quoted, the exact distribution is known (Wilks, 1935). It appears in rather a complicated form, but has been used here to check the series and the approximations. Table 9 shows the 5% and 1% significance points for the criterion M

Table 9. *Some comparisons for Wilks's statistic*

			χ^2 approximation			F approximation		
			M	Probability		M	Probability	
				Exact	Series		Exact	Series
$k=3$, $p_1=2$, $p_2=2$, $p_3=2$	$\nu=10$	5%	30·770	0·0612	0·0612	31·357	0·0549	0·0548
		1%	38·366	0·0139	0·0139	39·180	0·0117	0·0117
	$\nu=20$	5%	24·982	0·0516	0·0516	25·083	0·0504	0·0504
		1%	31·149	0·0105	0·0105	31·292	0·0101	0·0101

obtained by using the χ^2 and F approximations which are derived in the next section, together with the exact probabilities and the probabilities calculated from the series, using $\rho = 0{\cdot}68\dot{3}$, and six terms in the asymptotic and eight in the exponential series. Agreement to four places of decimals is usually obtained between the series and the exact value for the probability.

χ^2 approximation. Following the previous procedure, we find that M/C is distributed approximately as χ^2 with f degrees of freedom, where

$$\frac{1}{C} = 1-\frac{1}{12\nu f}(2\Sigma_3+3\Sigma_2) \quad \text{and} \quad f = \tfrac{1}{2}\Sigma_2.$$

F approximation. We have

$$f = \tfrac{1}{2}\Sigma_2, \quad A_1 = \frac{1}{12\nu f}(2\Sigma_3+3\Sigma_2), \quad A_2 = \frac{1}{12\nu^2 f}(\Sigma_4+2\Sigma_3-\Sigma_2),$$

from which, using equations (76) and (77), the F type approximation can be easily computed.

The quantities Σ_2, Σ_3 and Σ_4 required in these approximations are given by (81), the calculations of Table 9 give some indication of the accuracy to be expected.

6·2. *Special cases*

We consider two important special cases of the statistic, that in which there are only two groups of variates and that in which there is only one variate in each of the k groups.

6·21. *Case $k = 2$*

In this case the expressions for the coefficients in the series simplify considerably. Writing $p_1 = p$, $p_2 = q$, we obtain

$$f = pq, \quad \rho = 1-\frac{p+q+1}{2\nu}, \quad \beta = \tfrac{1}{2}(p+q+1),$$

$$\alpha_1 = 0, \quad \alpha_2 = \frac{pq}{48}(p^2+q^2-5),$$

$$\alpha_3 = 0, \quad \alpha_4 = \frac{pq}{1920}\{3p^4+3q^4+10p^2q^2-50(p^2+q^2)+159\},$$

$$\alpha_5 = 0, \quad \alpha_6 = \frac{pq}{16,128}\{3(p^6+q^6)-105(p^4+q^4)+1,113(p^2+q^2) + (21p^2-350+21q^2)p^2q^2-2,995\}.$$

Putting these values in (26) and (30) we confirm* the series given to terms in μ^{-4} by Rao (1948) for this case, $k = 2$. Bartlett had already (1938) obtained the χ^2 approximation using the scale factor $\dfrac{1}{C} = 1-\dfrac{p+q+1}{2\nu}$ (which is of course the factor given by the present procedure). Rao introduced this scale factor into the Wald & Brookner series, so as to obtain a χ^2 series with Bartlett's χ^2 approximation as the leading term, equivalent to (30). As we have seen, this choice of factor results in this particular case in α_1 and the α's of odd order being zero, so that the calculation of the series is correspondingly simpler.

* There appears to be a misprint in Rao's paper in the expression which corresponds with α_4, where the constant 159 is wrongly given as 150.

χ^2 *(Bartlett) approximation.* $\frac{M}{C} = \left(1 - \frac{p+q+1}{2\nu}\right) M$ is approximately distributed as χ^2 with $f = pq$ degrees of freedom.

F approximation. We find $A_2 - A_1^2 = (p^2+q^2-5)/12\nu^2$; thus for p and $q \geqslant 2$, $A_2 - A_1^2 > 0$, and the type VI form will be appropriate. M/b will be approximately distributed as F with f_1 and f_2 degrees of freedom, where

$$f_1 = pq, \quad f_2 = \frac{12\nu^2(pq+2)}{p^2+q^2-5}, \quad b = \frac{pq}{1 - \frac{p+q+1}{2\nu} - \frac{f_1}{f_2}}.$$

For p or q equal to 1 and 2, the exact distributions are known and provide simple tests (Wilks, 1932, 1935). For these cases λ and $\sqrt{\lambda}$ respectively are distributed in a type I distribution, and the significance test can be made, either by directly entering Thompson's tables of percentage points of the incomplete B-function, or by inversion of the statistic to its equivalent 'variance ratio' form and using tables of F or of Fisher's z (Bartlett, 1934; Rao, 1948). As has been pointed out by Bartlett (1938) if $p = 1$ and $q = 2$ (or $p = 2$ and $q = 1$) $\frac{M}{C} = \frac{\nu - 2}{\nu} M$ is distributed *exactly* as χ^2, and substituting these values for p and q in the expressions for α_r we find that in this case all these coefficients are zero, so providing a useful check. If p and q were both unity, $A_2 - A_1^2$ would be negative and the type I form be appropriate for the F approximation. Of course we shall not need to use the method here because the criterion $\sqrt{(1-\lambda)}$ is the sample correlation coefficient r and the exact distribution is known. The exact distributions are also known in certain other cases (Wilks, 1935; Wald & Brookner, 1941); the form which these take, however, is rather complicated, but they are useful to check approximations. In Table 10 are shown the 5 % significance points of M for a number of combinations of p and q as given by the χ^2 and F methods of approximation. In the cases chosen, the exact distribution is known, and this has been used to calculate the exact probability associated with each of these points. For comparison, the probability given by the series, using terms up to α_6 in the asymptotic series and, for most values, up to a_8 in the exponential series, is also shown.

We see that, providing ν is sufficiently large, Bartlett's approximation is in good agreement with the exact values, and the F approximation, since it involves very little more labour, provides a worth-while improvement. If ν is not large and one is doubtful whether these approximations will be sufficient, a rough but useful indication is provided by comparing the values obtained by the χ^2 and F approximations (in calculating the F approximation one will have already calculated the quantities needed for the χ^2 approximation). If these two approximations give substantially the same value, it may generally be taken as an indication that the approximation is adequate. If they differ markedly, a more accurate value should be calculated from the series.

6·22. *Case* $p_l = 1$, $l = 1, 2, \ldots, k$

If the k groups each contain only one variate, the hypothesis tested is that each of the variates is independent of all the others. The λ criterion then becomes the determinant of the sample correlation matrix, e.g., if $k = 3$,

$$\lambda = \begin{vmatrix} 1 & r_{12} & r_{13} \\ r_{12} & 1 & r_{23} \\ r_{13} & r_{23} & 1 \end{vmatrix},$$

Table 10. 5 % *significance points for M*

p	q	ν	χ^2 (Bartlett)			F (Box)		
			M	Probability		M	Probability	
				Exact	Series		Exact	Series
1	1	9	4·610	0·0494	0·0494	4·592	0·0499	0·0499
1	5	10	17·032	0·0624	0·0624	17·542	0·0555	0·0555
		20	13·419	0·0518	0·0518	13·504	0·0504	0·0504
1	10	20	26·153	0·0666	0·0666	27·022	0·0562	0·0562
		40	21·538	0·0525	0·0525	21·690	0·0505	0·0505
2	2	9	13·137	0·0515	0·0515	13·200	0·0506	0·0506
2	5	10	30·512	0·0737	0·0737	31·654	0·0614	0·0614
		20	22·884	0·0529	0·0529	23·053	0·0507	0·0507
2	10	20	46·534	0·0753	0·0753	48·164	0·0595	0·0595
		40	37·505	0·0535	0·0535	37·775	0·0507	0·0507
4	4	10	47·811	0·0945	0·0940	49·996	0·0735	0·0731*
		20	33·931	0·0542	0·0542	34·216	0·0512	0·0512

* With $\nu = 10$ and $p = q = 4$, six terms were taken in the asymptotic series and twelve in the exponential series; greater accuracy can be obtained by taking more terms in the asymptotic series.

where r_{ij} is the usual sample product moment correlation coefficient between the ith and jth variates. When $k = 2$ the criterion is simply $1 - r_{12}^2$.

The statistic is useful in supplying an overall test of independence between the k variates. For example, when $k = 5$ there will be ten individual correlation coefficients. Even when the null hypothesis, that all the variates are uncorrelated, is true, we shall expect often to come across individual coefficients which are 'significant'. For such a case it will be appropriate to apply the overall test before testing individual correlations. Again, the expressions for the coefficients simplify and we find, choosing ρ so that $\alpha_1 = 0$,

$$f = \tfrac{1}{2}k(k-1), \quad \rho = 1 - \frac{2k+5}{6\nu}, \quad \beta = \frac{2k+5}{6},$$

$$\alpha_1 = 0, \quad \alpha_2 = \frac{k(k-1)}{288}(2k^2 - 2k - 13),$$

$$\alpha_3 = \frac{k(k-1)}{3,240}(k-2)(2k-1)(k+1),$$

$$\alpha_4 = \frac{k(k-1)}{34,560}(16k^4-32k^3-252k^2+268k+1147),$$

$$\alpha_5 = \frac{k(k-1)}{136,080}(k-2)(k+1)(2k-1)(8k^2-8k-97),$$

$$\alpha_6 = \frac{k(k-1)}{7,838,208}(496k^6-1,488k^5-12,576k^4+27,632k^3+137,490k^2-151,554k-562,103).$$

For the χ^2 approximation we find, from the argument of § 3·1, that we should take $\frac{1}{C} = 1-\frac{2k+5}{6\nu}$. Thus $\left\{1-\frac{2k+5}{6\nu}\right\}M$ will be approximately distributed as χ^2 with $\frac{1}{2}k(k-1)$ degrees of freedom.

For the F approximation we have

$$f = \tfrac{1}{2}k(k-1), \quad A_1 = \frac{2k+5}{6\nu}, \quad A_2 = \frac{k^2+3k+2}{6\nu^2}.$$

For $k = 2$ and 3 we use the type I form and for $k \geqslant 4$ the type VI, since $A_2 - A_1^2 = \frac{2k^2-2k-13}{36\nu^2}$ is negative when $k = 2$ or 3 and positive for larger values of k. We then calculate f_1, f_2 and b, required in this approximation, by formulae (77) and (76) respectively.

7. Summary and conclusions

For a particular class of likelihood criteria, whose moments appear as the product of Γ-functions, a general method is described for obtaining probability levels when the null hypothesis is true. A number of statistics whose moments appear in this form are referred to, and a general method developed to obtain:

(*a*) A series which is in close agreement with the exact distribution.

(*b*) An approximate solution, using a single χ^2 distribution, which is sufficiently accurate for moderate or large samples.

(*c*) A rather better approximation, using a single F distribution, giving close agreement even when the samples are rather small.

The method is illustrated for the following two general statistics:

(1) *Tests for constancy of variance and covariance*

(*a*) *Univariate case.* The F approximation is of the same order of accuracy as Hartley's (1940) series solution although it requires very much less calculation, and significance may be judged by consulting tables of the significance points of the variance ratio F alone.

(*b*) *Multivariate case.* The series solution shows remarkably close agreement with the exact distribution when this is known, and is used in other cases to compare approximations. The χ^2 approximation does not correspond with that found by Bishop (1939), but is, in fact, simpler and more accurate.

The series confirms the accuracy of significance points found by fitting a type I curve to the first two moments of l_1. The calculation of the moments involved in this method renders it too laborious for routine use, and Bishop suggested two working approximations; the F approximation developed here is more accurate than these approximations, whilst it involves no more labour and can be used when the sample sizes are unequal.

(2) *Wilks's test for independence of k groups of variates*

The asymptotic series, and χ^2 and F approximations are derived for this case, and the relation of the results with those of Wald & Brookner, Bartlett, and Rao is discussed. The exact distribution is used to assess the accuracy of the proposed methods in a number of cases. The probabilities given by the series are found to be in excellent agreement with the true values, even for fairly small samples. Providing the sample sizes are not too small, the χ^2 and F approximations will be sufficiently accurate, the latter providing the better approximation, and allowing the sample size to be rather smaller than is possible with the χ^2 approximation. When the number of variates in each group is one, we have a test criterion for the hypothesis that k variates are mutually independent, and the same procedure provides the series solution and simple approximations for tests of significance.

In conclusion, I wish gratefully to acknowledge the help and guidance I have received from Dr H. O. Hartley throughout this investigation.

REFERENCES

Barnes, E. W. (1899). *Mess. Math.* p. 64.
Bartlett, M. S. (1934). *Proc. Camb. Phil. Soc.* **30**, 327.
Bartlett, M. S. (1937). *Proc. Roy. Soc.* A, **160**, 268.
Bartlett, M. S. (1938). *Proc. Camb. Phil. Soc.* **34**, 33.
Bartlett, M. S. (1947). *J.R. Statist. Soc. Suppl.* **9**, 176.
Bishop, D. J. (1939). *Biometrika*, **31**, 31.
Bishop, D. J. & Nair, U. S. (1939). *J.R. Statist. Soc. Suppl.* **6**, 89.
Brown, G. W. (1939). *Ann. Math. Statist.* **10**, 119.
George, A. (1945). *Sankhyā*, **1**, 20.
Hartley, H. O. (1940). *Biometrika*, **31**, 249.
Hartley, H. O. & Pearson, E. S. (prefatory note) (1946). *Biometrika*, **33**, 296.
Mauchly, J. W. (1940). *Ann. Math. Statist.* **11**, 204.
Milne-Thomson, L. M. (1933). *The calculus of finite differences.* Macmillan.
Nair, U. S. (1939). *Biometrika*, **30**, 274.
Nayer, P. P. N. (1936). *Statist. Res. Mem.* **1**, 38.
Neyman, J. & Pearson, E. S. (1928). *Biometrika*, **20 A**, 175 and 263.
Neyman, J. & Pearson, E. S. (1931). *Bull. int. Acad. Cracovie*, A, p. 460.
Neyman, J. & Pearson, E. S. (1936). *Statist. Res. Mem.* **1**, 1.
Neyman, J. & Pearson, E. S. (1938). *Statist. Res. Mem.* **2**, 25.
Pearson, E. S. & Wilks, S. S. (1933). *Biometrika*, **25**, 353.
Pitman, E. J. G. (1939). *Biometrika*, **31**, 200.
Plackett, R. L. (1946). *J.R. Statist. Soc.* **109**, 457.
Plackett, R. L. (1947). *Biometrika*, **34**, 311.
Rao, C. R. (1948). *Biometrika*, **35**, 71.
Thompson, C. M. (1941). *Biometrika*, **32**, 151.
Thompson, C. M. & Merrington, M. (1946). *Biometrika*, **33**, 296.
Tukey, J. W. & Wilks, S. S. (1946). *Ann. Math. Statist.* **17**, 318.
Votaw, D. F. (1948). *Ann. Math. Statist.* **19**, 447.
Wald, A. & Brookner, R. J. (1941). *Ann. Math. Statist.* **12**, 137.
Welch, B. L. (1935). *Biometrika*, **27**, 145.
Welch, B. L. (1936). *Statist. Res. Mem.* **1**, 52.
Wilks, S. S. (1932). *Biometrika*, **24**, 471.
Wilks, S. S. (1935). *Econometrica*, **3**, 309.
Wilks, S. S. (1946). *Ann. Math. Statist.* **17**, 257.
Wishart, J. (1928). *Biometrika*, **20 A**, 32.