A modification to Paulson's approximation to the variance ratio distribution

By T. Ashby*

Paulson's approximation to the variance ratio distribution is poor or useless for less than five degrees of freedom in the denominator. However, it can be modified in this range to give results correct to three significant figures.

(First received November 1967)

Tables of percentage points of the variance ratio distribution F are readily available (e.g. Fisher and Yates, 1963), but are not convenient to use with a digital computer. On the other hand it is not practical to calculate regressions or analyses of variance involving many variables without using a computer, and significance testing is naturally required in such programs. Consequently some way of calculating values of F is required, and an approximation is generally used.

Kendall and Stuart (1958 edition) discuss three such approximations. The third of these is due to Paulson (1942) whose reasoning is first that F can be written as the ratio of two chi-squared variates divided by their degrees of freedom. Secondly, Wilson and Hilferty (1931) have shown that $(X^2/n)^{1/3}$ is nearly normally distributed, with mean $1 - 2/9n$ and variance $2/9\,n$. Thirdly, Fieller (1932) has shown that if x and y are normally distributed with means μ_x and μ_y and standard deviations σ_x and σ_y then a function

$$R = \frac{\frac{y}{x}\mu_x - \mu_y}{\left(\left(\frac{y}{x}\right)^2 \sigma_x^2 + \sigma_y^2\right)^{1/2}}$$

is approximately normally distributed with zero mean and unit variance. Replacing y/x in this last expression by $F^{1/3}$ we obtain

$$U = \frac{(1 - 2/9n_2)F^{1/3} - (1 - 2/9n_1)}{(2F^{2/3}/9n_2 + 2/9n_1)^{1/2}}$$

where F is the variance ratio and n_1 and n_2 are the degrees of freedom in the numerator and the denominator. This function is approximately normally distributed with zero mean and unit variance.

This approximation can of course be rearranged to give approximations to the percentage points of F in terms of normal deviates U and n_1 and n_2. We shall call this approximation P. If we write f_1 for $2/9n_1$ and f_2 for $2/9n_2$ and solve Paulson's equation as a quadratic in $P^{1/3}$ we obtain

$$P^{1/3} = \frac{(1-f_1)(1-f_2) \pm [u^2(f_1+f_2+f_1 f_2(f_1+f_2-u^2-4))]^{1/2}}{(1-f_2)^2 - u^2 f_2}.$$

Computer programs store the important values of normal deviates u, and from these calculate P. The approximation is good for values of n_2 above 10 but for smaller values is poorer, and for n_2 up to 3 produces useless and even negative results, as can be seen by comparing **Tables 1** and **2**.

This difficulty can be overcome, however, by applying a linear transformation $G = mP + c$, where m and c are constants depending on the significance level and on n_2, the degrees of freedom of the denominator. Suitable values are stored in the computer and P is adjusted whenever n_2 is below a critical value. This modified approximation gives results which are generally correct to three significant figures, as can be seen by comparing **Tables 1** and **3**.

Table 1

Variance ratio distribution, F
(Fisher & Yates)

n_2 \ n_1	95%			99%			99·9%		
	2	6	24	2	6	24	2	6	24
1	199·5	234·0	249·0	4999	5859	6234	5000*	5859*	6235*
2	19·00	19·33	19·45	99·00	99·33	99·46	999·0	999·3	999·5
3	9·55	8·94	8·64	30·82	27·91	26·60	148·5	132·8	125·9
4	6·94	6·16	5·77	18·00	15·21	13·93	61·25	50·53	45·77
5	5·79	4·95	4·53	13·27	10·67	9·47	37·12	28·84	25·14
10	4·10	3·22	2·74	7·56	5·39	4·33	14·91	9·92	7·64

* Multiply these values by 100.

* *Central Operational Research Department, Richard Thomas and Baldwins Ltd., Pye Corner, Nash, Near Newport, Mon.* (now with *MAS Branch, British European Airways, Bealine House, South Ruislip, Middlesex.*)

P approximates F well above $n_2 = 10$ and is quite good for n_2 down to 5. A program testing significance at the 5%, 1% and 0·1% levels and using the modification for n_2 up to 5 needs to store 30 extra numbers.

Table 4 gives values of m and c for n_2 up to 10, derived empirically. The modified approximation has been programmed in Autocode for the Elliott 803 and in NELIAC for the Univac 490.

Table 2

Paulson's approximation, P

n_2 \ n_1	95% 2	95% 6	95% 24	99% 2	99% 6	99% 24	99·9% 2	99·9% 6	99·9% 24
1	5281*	6706*	7300*	−14·58	−24·96	−29·33	−0·246	−2·108	−2·913
2	23·92	24·86	25·20	560·20	640·3	673·3	−181·0	−292·5	−339·3
3	10·09	9·54	9·25	42·42	40·86	40·11	1381	1544	1612
4	7·08	6·33	5·94	20·17	17·66	16·50	110·7	103·4	100·2
5	5·83	5·02	4·60	13·98	11·51	10·36	47·79	40·11	36·76
10	4·08	3·22	2·74	7·58	5·44	4·39	15·32	10·32	8·074

* Multiply these values by 10,000.

Table 3

Modified approximation

n_2 \ n_1	95% 2	95% 6	95% 24	99% 2	99% 6	99% 24	99·9% 2	99·9% 6	99·9% 24
1	198·5	234·2	249·1	4988	5863	6232	4990*	5859*	6235*
2	19·00	19·33	19·44	99·01	99·34	99·48	999·0	999·4	999·6
3	9·53	8·94	8·63	30·78	27·94	26·58	148·6	132·8	126·2
4	6·93	6·16	5·77	18·01	15·22	13·94	61·28	50·50	45·77
5	5·77	4·96	4·54	13·27	10·67	9·47	37·17	28·82	25·17
10	4·08	3·22	2·74	7·56	5·38	4·32	14·90	9·90	7·65

* Multiply these values by 100.

Table 4

Constants for modified approximation

n_2	95% points m	95% points c	99% points m	99% points c	99·9% points m	99·9% points c
1	0·00000251	65·9	−84·33	3758	−46667	487600
2	0·3461	10·723	0·0042	96·65	−0·0034	998·4
3	1·066	−1·23	1·826	−46·67	−0·0972	282·9
4	1·024	−0·316	1·11	−4·38	1·488	−103·4
5	1	−0·06	1·05	−1·41	1·087	−14·78
6	1	−0·03	1·035	−0·68	1·038	−4·96
7	1	−0·01	1·027	−0·40	1·0225	−2·28
8	1	0	1·020	−0·25	1·012	−1·19
9	1	0	1·015	−0·167	1·005	−0·67
10	1	0	1·013	−0·125	1	−0·42

References

FISHER, R. A., and YATES, F. (1963). *Statistical Tables for Biological Agricultural and Medical Research*, Edinburgh: Oliver & Boyd.

KENDALL, M. G., and STUART, A. (1958). *The Advanced Theory of Statistics*, Vol. I, London: Charles Griffin.

PAULSON, E. (1942). An approximate normalisation of the analysis of variance distribution, *Annals of Math. Stat.*, Vol. 13, pp. 233–235.

WILSON, E. B., and HILFERTY, M. M. (1931). The distribution of Chi-square, *National Acad. Sc. Proc.*, Vol. 17, pp. 684–688.

FIELLER, E. C. (1932). The distribution of the index in a normal bivariate population, *Biometrika*, Vol. 24, pp. 428–440.