A LOWER BOUND TO THE PROBABILITY OF VARIANCE RATIO

A. K. P. C. SWAIN

(Received Nov. 13, 1963; revised Sept. 17, 1964)

1. Introduction

F. N. David (1949) derived approximations to the moments of Z (where $Z=\frac{1}{2}\log_e\frac{k_2}{k_2'}$, k_2 and k_2' being the second sample cumulants), when the two estimates of variances involved are based on independent samples from two populations. The underlying assumptions made about the two populations are the existence of cumulants up to any order desired. Using these moments David discussed the effect of skewness on the distributions of Z-criterion. Geary (1947) discussed the effect of kurtosis on the mentioned criterion. Finch (1950) noted the combined effect of skewness and kurtosis on the distribution of Z, under the assumption that the hypothesis tested is true.

The present paper is based on finding a lowerbound to the probability of variance ratio, which uses only the kurtosis of the given populations. The method has been derived from an inequality given by Uspensky.

2. Uspensky's inequality

Let x be a stochastic variable with mean m and variance σ^2. Then Prob $\{x-m\geqq t\}\geqq t^2/(\sigma^2+t^2)$ if $t<0$. Now putting $x-m-t=v$, we have

$$\begin{aligned}\text{Prob}\{v\geqq 0\} &\geqq \frac{t^2}{\sigma^2+t^2}\\ &\geqq \{E(v)\}^2/[V(v)+\{E(v)\}^2]\\ &\geqq \{E(v)\}^2/E(v^2).\end{aligned}$$

Suppose we are having n_1 samples from a population having population variance σ_1^2, and n_2 samples from another population having population variance σ_2^2. The hypothesis to test is $\sigma_1^2=\sigma_2^2$. Let s_1^2 and s_2^2 be the sample variances from the mentioned populations respectively.

$$E(s_1^2)=\sigma_1^2,\qquad E(s_2^2)=\sigma_2^2.$$

Define $v=s_2^2-Fs_1^2$. Under the hypothesis,

$$E(v)=\sigma^2-F\sigma^2$$
$$=\sigma^2(1-F)$$
$$=k_2(1-F),$$

k_2 being the second cumulant in the populations,

$$E(v^2)=E\{s_2^4+F^2s_1^4-2Fs_1^2s_2^2\}$$
$$=E(s_2^4)+F^2E(s_1^4)-2FE(s_1^2)E(s_2^2)$$
$$=\left[\frac{k_4'}{n_1}+\frac{2k_2^2}{n_1-1}+k_2^2\right]+F^2\left[\frac{k_4}{n_2}+\frac{2k_2^2}{n_2-1}+k_2^2\right]-2Fk_2^2,$$

k_4, k_4' being fourth cumulants of the 1st and 2nd populations respectively. Therefore,

$$\text{Prob}\,\{v\geqq 0\}\geqq\frac{k_2^2(1-F)^2}{\left[\frac{k_4'}{n_1}+\frac{2k_2^2}{n_1-1}+k_2^2\right]+F^2\left[\frac{k_4}{n_2}+\frac{2k_2^2}{n_2-1}+k_2^2\right]-2Fk_2^2}$$

$$\geqq\frac{(1-F)^2}{\left[\frac{\beta_2'-3}{n_1}+\frac{2}{n_1-1}+1\right]+F^2\left[\frac{\beta_2-3}{n_2}+\frac{2}{n_2-1}+1\right]-2F}$$

where $\beta_2'=\frac{k_4'}{k_2^2}+3$ and $\beta_2=\frac{k_4}{k_2^2}+3.$

For simplicity taking $n_1=n_2=n$ (say) we have

$$\text{Prob}\left\{\frac{s_2^2}{s_1^2}\geqq F\right\}\geqq\frac{(1-F)^2}{\left[\frac{\beta_2'-3}{n}+\frac{2}{n-1}+1\right]+F^2\left[\frac{\beta_2-3}{n}+\frac{2}{n-1}+1\right]-2F}.$$

The above expression for the lower bound to the probability of variance ratio depends on F, β_2', β_2, and n. For illustration two tables taking $n=4$, $F=2$, $\beta_2'=1, 2, 3, 4, 5, 6, 10$, $\beta_2=1, 2, 3, 4, 5, 6, 10$; and $n=10$, $F=3$, $\beta_2'=1, 2, 3, 4, 5, 6, 10$, $\beta_2=1, 2, 3, 4, 5, 6, 10$ are given below.

From the Tables 1 and 2, it is seen that the lower limit decreases to 0.0764 and 0.3025 respectively, when the kurtosis of the populations in question are as high as 10. This is very small in comparison with 0.2308 and 0.6429, when the populations are normal.

Table 1. with $n=4$, $F=2$

	1	2	3	4	5	6	10
1	0.5454	0.4800	0.4285	0.3871	0.3529	0.3243	0.2449
2	0.3529	0.3243	0.3000	0.2791	0.2608	0.2449	0.1967
3	0.2608	0.2449	0.2308	0.2182	0.2069	0.1967	0.1644
4	0.2069	0.1967	0.1875	0.1791	0.1714	0.1644	0.1412
5	0.1714	0.1644	0.1579	0.1519	0.1463	0.1412	0.1237
6	0.1463	0.1412	0.1364	0.1319	0.1276	0.1237	0.1101
10	0.0923	0.0902	0.0882	0.0863	0.0845	0.0828	0.0764

Table 2. with $n=10$, $F=3$

	1	2	3	4	5	6	10
1	0.9474	0.9255	0.9046	0.8846	0.8654	0.8471	0.7809
2	0.7809	0.7660	0.7516	0.7377	0.7244	0.7115	0.6642
3	0.6642	0.6534	0.6429	0.6327	0.6228	0.6133	0.5779
4	0.5779	0.5696	0.5616	0.5539	0.5463	0.5389	0.5114
5	0.5114	0.5049	0.4986	0.4925	0.4865	0.4806	0.4586
6	0.4586	0.4534	0.4483	0.4434	0.4385	0.4337	0.4157
10	0.3246	0.3220	0.3194	0.3169	0.3144	0.3120	0.3025

3. Conclusion

The lower bound has been derived on the basis of the assumption that the cumulants exist up to fourth order for both the populations and can be applied to any type of parent population whether continuous or discontinuous. As it has been seen that the lower bound depends on the knowledge of the kurtosis of the populations in question and in case of non availability, an approximate idea can be obtained by using the upper bound of β_2's of the populations.

Remark. For the practical utility I shall also give the values of F, which make the lower bound constant for the values of β_2' and β_2.

β_2 \ β_2'	1	2	3	4	5	6	10
1	1.0485	1.0854	1.1118	1.1328	1.1539	1.1698	1.2224
2	1.0912	1.1177	1.1388	1.1600	1.1759	1.1918	1.2395
3	1.1236	1.1449	1.1662	1.1822	1.1982	1.2035	1.2567
4	1.1511	1.1725	1.1865	1.2046	1.2153	1.2313	1.2742
5	1.1788	1.1950	1.2111	1.2218	1.2326	1.2487	1.2918
6	1.2014	1.2176	1.2230	1.2393	1.2555	1.2663	1.3042
10	1.2833	1.2943	1.3054	1.3165	1.3275	1.3331	1.3773

The table so constructed is given for the case, $n=10$, $P=0.05$ and β_2', $\beta_2=1, 2, 3, 4, 5, 6, 10$.

The above table has been constructed from the equation $f(F, \beta_2', \beta_2, n) = 0.05$, where n, β_2' and β_2 are known. Since the equation is a quadratic one, I have taken the root, greater than unity.

RIVER RESEARCH INSTITUTE, CALCUTTA, INDIA

REFERENCES

[1] F. N. David, "The moments of the Z and F distributions," *Biometrika*, 36 (1949), 394-403.

[2] D. J. Finch, "The effects of non-normality on the Z test, when used to compare the variances in two populations," *Biometrika*, 37 (1950), 186-189.

[3] R. C. Geary, "Testing for normality," *Biometrika*, 34 (1947), 209-242.

[4] J. V. Uspensky, *Introduction to Mathematical Probability*, McGraw Hill Book Company, 1937.