ON CERTAIN TWO-SAMPLE NONPARAMETRIC TESTS FOR VARIANCES[1]

By Balkrishna V. Sukhatme

Indian Council of Agricultural Research, New Delhi

Introduction. Let $X_1, X_2, \cdots, X_m$ and $Y_1, Y_2, \cdots, Y_n$ be two samples of independent observations drawn from two populations with cumulative distribution functions $F(x)$ and $G(x)$, respectively. We will assume in what follows that F and G are absolutely continuous and that they are the same in all respects except that they differ in the scale parameter. The problem considered here is that of testing the hypothesis

$$H: F = G,$$

$$A: F \neq G.$$

If the X's and the Y's come from normal populations, the usual test of significance for testing the hypothesis H is the variance ratio F-test, which is the most commonly used statistical test for comparing variances. Usually however, since little is known about the populations from which the samples are drawn, this test is used as if the assumption of normality could be ignored. It appears, however, that such is not the case. This was first pointed out by E. S. Pearson [1], who conducted certain experimental investigations. His findings were later confirmed by several other authors, especially by Geary [2] and Gayen [3]. They showed that the F-test is particularly sensitive to changes in kurtosis from the normal theory value of zero. Now, it is easy to see that the F statistic, when suitably normalised, is asymptotically distribution free. More recently, Box and Andersen [4] and [5] have studied this problem in great detail and have shown on the basis of extensive sampling experiments that the F statistic so normalised is insensitive to departures from normality, at least for large samples. Very recently attempts have also been made to construct nonparametric tests, particularly by Mood [6] and Lehmann [7].

The test proposed by Mood is similar to the variance ratio F-test with ranks replacing the original observations. He has also computed the asymptotic relative efficiency of the test with respect to the F-test for normal alternatives. In this paper, we will derive a general formula for the asymptotic relative efficiency of Mood's test with respect to the F-test for scalar alternatives but almost arbitrary continuous distributions.

The test proposed by Lehmann is essentially of the Wilcoxon-Mann-Whitney type (see [8] and [9]) applied to all possible differences between the X's and

Received January 29, 1956.

[1] This research was performed while the author was at the Statistical Laboratory, University of California, Berkeley, and was supported by the Office of Ordnance Research, United States Army, under Contract DA-04-200-ORD-171.

the Y's. As pointed out by Mood, the test is not distribution free. However, he computed the asymptotic relative efficiency of Lehmann's test with respect to the F-test on the assumption that the asymptotic variance of the test statistic is distribution free. It will be shown in this paper that the asymptotic variance of the test statistic suggested by Lehmann is not independent of the form of the distribution, even under the null hypothesis $F = G$.

Lastly, we will propose a new nonparametric test for comparing variances and obtain a general formula for its asymptotic relative efficiency with respect to the F-test for scalar alternatives but almost arbitrary continuous distributions.

1. Mood's test. This is a dispersion test based on the statistic

$$M = \sum_{i=1}^{n} \left(r_i - \frac{m+n+1}{2} \right)^2, \tag{1.1}$$

where r_i is the rank of Y_i in the combined sample of $m + n$ observations. We reject the hypothesis if M is too large. Then, as shown by Mood, under the null hypothesis

$$E(M) = \frac{n(s+1)(s-1)}{12}, \tag{1.2}$$

$$\operatorname{var}(M) = \frac{mn(s+1)(s+2)(s-2)}{180}, \tag{1.3}$$

and under the alternative

$$\begin{aligned} E(M) = {} & \frac{n}{12} \{3(s+1)^2 - 6(n+1)(s+1) + 2(n+1)(2n+1)\} \\ & - mn\left\{2(m-1)\int FG\,dF + (n-1)\int G^2\,dF - (s-2)\int G\,dF\right\}, \end{aligned} \tag{1.4}$$

where, for short, we write s for $m + n$.

Let $G(x) = F(x\theta)$. Then, proceeding as in [6],

$$\left.\frac{dE(M)}{d\theta}\right|_{\theta=1} = -2(s-2)\left\{\int xF(x)f(x)\,dF(x) - \frac{1}{2}\int xf(x)\,dF(x)\right\}. \tag{1.5}$$

The efficacy of the M-test is therefore equal to

$$\frac{180(s-2)^2\left\{2\int xF(x)f^2(x)\,dx - \int xf^2(x)\,dx\right\}^2}{mn(s+1)(s+2)(s-2)}. \tag{1.6}$$

Also, the efficacy of the variance ratio F-test is

$$\frac{4mn}{(m+n)(\beta_2 - 1)}. \tag{1.7}$$

Hence, the asymptotic relative efficiency of the M-test with respect to the F-test is given by

$$e_M = 45(\beta_2 - 1)\left\{2\int xF(x)f^2(x)\,dx - \int xf^2(x)\,dx\right\}^2, \tag{1.8}$$

where

$$\beta_2 = \frac{\int (x - EX)^4\,dF(x)}{\left\{\int (x - EX)^2\,dF(x)\right\}^2}.$$

From the formula (1.8), it is obvious that depending on $f(x)$, $0 < e_M < \infty$. Thus, considering

$$f(x) = \frac{1-\alpha}{2\cdot a^{1-\alpha}}\cdot\frac{1}{|x|^\alpha} \qquad -a \leqq x \leqq a,\quad \alpha < 1, \tag{1.9}$$

we find after some computations that

$$e_M = \frac{5(1-\alpha)}{(5-\alpha)}, \tag{1.10}$$

which tends to zero as α tends to unity. Thus the asymptotic relative efficiency of the M-test with respect to the F-test can be made as small as we please. Similarly, taking $f(x)$ to be a Pearson Type VII density function, it can be shown that the asymptotic efficiency can be made as large as we please. In particular, if $f(x)$ is the standard normal density function with mean zero and variance unity, $e_M = 0.76$. If $f(x)$ is equal to one on the unit interval about the origin and zero otherwise, then, $e_M = 1$.

2. Lehmann's test. The test consists in forming all the $\binom{m}{2}$ positive differences between the m X's and the $\binom{n}{2}$ positive differences between the n Y's. The test is then based on the statistic

$$L = \binom{m}{2}^{-1}\binom{n}{2}^{-1}\sum_{\substack{i<j\\k<l}}\varphi(|X_i - X_j|, |Y_k - Y_l|), \tag{2.1}$$

where

$$\varphi(u, v) = 1 \quad \text{if } u < v,$$
$$= 0 \quad \text{otherwise.}$$

Clearly, this is a generalised U statistic in the sense of Lehmann [10] and hence it follows that if $m = Np$ and $n = Nq$, with $p + q = 1$, $\sqrt{N}(L - EL)$ is asymptotically normally distributed with asymptotic variance σ^2 given by

(2.2) $$\sigma^2 = 4\left\{\frac{\xi_{10}}{p} + \frac{\xi_{01}}{q}\right\},$$

where

(2.3) $$\begin{aligned}\xi_{10} = E[\varphi(|X_1 - X_2|, |Y_1 - Y_2|)\varphi(|X_1 - X_3|, |Y_3 - Y_4|)] \\ - E^2\varphi(|X_1 - X_2|, |Y_1 - Y_2|)\end{aligned}$$

and

(2.4) $$\begin{aligned}\xi_{01} = E[\varphi(|X_1 - X_2|, |Y_1 - Y_2|)\varphi(|X_3 - X_4|, |Y_1 - Y_3|)] \\ - E^2\varphi(|X_1 - X_2|, |Y_1 - Y_2|).\end{aligned}$$

Now, we have

(2.5) $$E\varphi(|X_1 - X_2|, |Y_1 - Y_2|) = P(|X_1 - X_2| \leqq |Y_1 - Y_2|) = \tfrac{1}{2}$$

under hypothesis.

To compute ξ_{10} and ξ_{01}, we first compute the following, all under the hypothesis $F = G$. Let

$$|X_1 - X_2| = U_1, |X_1 - X_3| = U_2, |Y_1 - Y_2| = V_1, |Y_3 - Y_4| = V_2.$$

Then,

(2.6) $$K(v_1) = P(V_1 \leqq v_1) = \int [F(x + v_1) - F(x - v_1)]\, dF(x).$$

Also, we have

(2.7) $$\begin{aligned}H(u_1, u_2) &= P(U_1 \leqq u_1, U_2 \leqq u_2) \\ &= \int [F(x + u_1) - F(x - u_1)][F(x + u_2) - F(x - u_2)]\, dF(x).\end{aligned}$$

Then, we see that

(2.8) $$\xi_{10} = P(U_1 \leqq V_1, U_2 \leqq V_2) - \tfrac{1}{4} = \iint H(v_1, v_2)k(v_1)k(v_2)\, dv_1\, dv_2 - \tfrac{1}{4}.$$

Exactly, in the same manner, we find that

(2.9) $$\xi_{01} = \iint K(t_1)K(t_2)h(t_1, t_2)\, dt_1\, dt_2 - \tfrac{1}{4},$$

where

$$h(t_1, t_2) = \int [f(x + t_1) + f(x - t_1)][f(x + t_2) - f(x - t_2)]\, dF(x).$$

Taking

$$\begin{aligned} f(x) &= 1 \qquad -\tfrac{1}{2} \leqq x \leqq \tfrac{1}{2}, \\ &= 0 \qquad \text{otherwise,}\end{aligned}$$

it turns out that

$$\sigma^2/N = 1/180 \cdot (1/m + 1/n). \tag{2.10}$$

On the other hand, if

$$f(x) = \tfrac{1}{2} \cdot e^{-|x|},$$
$$\sigma^2/N = \frac{107}{2^6 \cdot 3^4} \cdot (1/m + 1/n). \tag{2.11}$$

It follows that the asymptotic variance depends essentially on the form of the distribution function $F(x)$, even when $F = G$. Hence, the test based on the statistic L is not asymptotically distribution free.

From the above results, it seems that Mood's test is reasonably efficient for normal alternatives and highly efficient for some non-normal alternatives. The test, however, presupposes knowledge about the relative location of the two populations, which is not always present. If it is not, the test can be modified by applying the test to the deviations from the sample medians rather than to the observations themselves. The modified test is essentially the same test, and we would expect the modified test to behave nicely, at least for large samples. It will be shown in another paper that the modified test is not asymptotically distribution free in the sense that the asymptotic distribution of the test statistic is not independent of the original population from which the samples are drawn under the null hypothesis. In the next section, we therefore propose another nonparametric test for comparing variances, especially constructed with this object in view. As will be seen in the next section, the test is not so efficient as the one proposed by Mood. This test also assumes knowledge about the relative location of the two populations. It will be shown in another paper that under certain regularity conditions, the proposed test after modification is asymptotically distribution free.

3. The proposed T-test. The test statistic may be defined as

$$T = \frac{1}{mn} \sum_{i=1}^{m} \sum_{j=1}^{n} \psi(X_i, Y_j), \tag{3.1}$$

where

$$\psi(X, Y) = 1 \quad \text{if} \begin{cases} \text{either } 0 < X < Y \\ \text{or} \quad Y < X < 0, \end{cases}$$
$$= 0 \quad \text{otherwise.}$$

We reject the hypothesis if T is either too large or too small. We shall now find the mean and the variance of T both under the hypothesis and the alternative.

Mean and variance of T under the hypothesis.

$$E(T) = E\psi(X_i, Y_j) = P(0 < X_i < Y_j) + P(Y_j < X_i < 0) = \tfrac{1}{4}. \tag{3.2}$$

Squaring and taking expectations and noting that

$$E[\psi(X_i, Y_j)\psi(X_i, Y_k)] = E[\psi(X_i, Y_j)\psi(X_h, Y_j)] = \tfrac{1}{12},$$

it follows easily that

$$\operatorname{var}(T) = \frac{m + n + 7}{48mn}. \tag{3.3}$$

Mean and variance of T under the alternative.

$$E(T) = \int_0^\infty (1 - G)\, dF + \int_{-\infty}^0 G\, dF. \tag{3.4}$$

To find the variance under the alternative, it is easily seen that

$$E[\psi(X_i, Y_j)\psi(X_i, Y_k)] = \int_0^\infty (1 - G)^2\, dF + \int_{-\infty}^0 G^2\, dF, \tag{3.5}$$

$$E[\psi(X_i, Y_j)\chi(X_k, Y_j)] = \int F^2\, dG - \int F\, dG + \tfrac{1}{4}. \tag{3.6}$$

Whence, we find that

$$\begin{aligned}\operatorname{var}(T) = 1/mn\Bigg[\int_0^\infty F\, dG - \int_{-\infty}^0 F\, dG + (n - 1)\\ \cdot\left[\int_0^\infty (1 - G)^2\, dF + \int_{-\infty}^0 G^2\, dF\right] + (m - 1)\left[\int F^2\, dG - \int F\, dG + \frac{1}{4}\right]\\ - (m + n - 1)\left\{\int_0^\infty F\, dG - \int_{-\infty}^0 F\, dG\right\}^2\Bigg],\end{aligned} \tag{3.7}$$

which tends to zero as m and n tend to infinity. Thus,

$$T \xrightarrow{P} \int_0^\infty F\, dG - \int_{-\infty}^0 F\, dG \quad \text{as} \quad m \quad \text{and} \quad n \longrightarrow \infty.$$

Hence, the test is consistent.

Asymptotic efficiency of the T test. We observe that T is a modified form of the Wilcoxon-Mann-Whitney statistic. Mann and Whitney proved the asymptotic normality of the Wilcoxon statistic under the hypothesis and Lehmann proved it under the alternative. Using these results, it follows easily that T is asymptotically normally distributed both under the hypothesis and the alternative. It can also be verified that all the conditions of Pitman's [11] theorem are satisfied. We are therefore ready to compute the asymptotic relative efficiency of the T-test with respect to the variance ratio F-test. We have

$$\left.\frac{dE(T)}{d\theta}\right|_{\theta=1} = \int_{-\infty}^{0} xf^2(x)\,dx - \int_{0}^{\infty} xf^2(x)\,dx.$$

Efficacy of the T-test is therefore equal to

$$48mm/(m + n + 7)\left[\int_0^{\infty} xf^2(x)\,dx - \int_{-\infty}^{0} xf^2(x)\,dx\right]^2. \tag{3.8}$$

Whence, we find as before that the asymptotic relative efficiency of the T-test is equal to

$$e_T = 12(\beta_2 - 1)\left[\int_0^{\infty} xf^2(x)\,dx - \int_{-\infty}^{0} xf^2(x)\,dx\right]^2. \tag{3.9}$$

It can be demonstrated as before that the efficiency can be anything from zero to infinity. In particular, if $f(x)$ is the standard normal density function, $e_T = 0.61$. If $f(x) = \frac{1}{2}\cdot e^{-|x|}$ $e_T = 0.94$.

4. Acknowledgment. The author wishes to express his thanks to Professor Erich Lehmann for proposing this investigation and for his interest and helpful suggestions during its progress.

REFERENCES

[1] E. S. Pearson, "The analysis of variance in cases of non-normal variation," *Biometrika*, Vol. 23 (1931), pp. 114–134.
[2] R. C. Geary, "Testing for normality," *Biometrika*, Vol. 34 (1947), pp. 209–241.
[3] A. K. Gayen, "The distribution of the variance ratio in random samples of any size drawn from non-normal universes," *Biometrika*, Vol. 37 (1950), pp. 236–255.
[4] G. E. P. Box, "Non-normality and tests on variances," *Biometrika*, Vol, 40 (1953), pp. 318–335.
[5] G. E. P. Box and S. L. Anderson, "Permutation theory in the derivation of robust criteria and the study of departures from assumption," *J. R. Statist. Soc.*, B, Vol. 17 (1955), pp. 1–34.
[6] A. Mood, "On the asymptotic efficiency of certain nonparametric tests," *Ann. Math. Stat.*, Vol. 25 (1954), pp. 514–522.
[7] E. L. Lehmann, "Consistency and unbiasedness of certain nonparametric tests," *Ann. Math. Stat.*, Vol. 22 (1951), pp. 165–179.
[8] F. Wilcoxon, "Individual comparison by ranking methods," *Biometrics*, Vol. 1 (1945), pp. 80–83.
[9] H. B. Mann and D. R. Whitney, "On a test of whether one of two random variables is stochastically larger than the other," *Ann. Math. Stat.*, Vol. 18 (1947), pp 50–60.
[10] E. L. Lehmann, unpublished.
[11] G. E. Noether, "Asymptotic properties of the Wald-Wolfowitz test of randomness," *Ann. Math. Stat.*, Vol. 21 (1950), pp. 231–246.