ON THE TWO SAMPLE PROBLEM: A HEURISTIC METHOD FOR CONSTRUCTING TESTS[1]

By V. P. Godambe

Science College, Nagpur, India

1. Introduction. The two-sample problem arises as follows. We are given two independent samples from populations A and B respectively and are required to investigate whether the population A could be considered as identical with B. In the usual terminology of hypothesis testing: Given two independent samples $x_1, \cdots, x_m$ and $x_{m+1}, \cdots, x_{m+n}$ from populations with unknown cumulative distribution functions F and G respectively, the problem is to test the composite hypothesis

$$H_0: F = G$$

against the alternatives

$$H_1: F \neq G,$$

F and G being completely or partially unspecified.

In the following lines we shall discuss a method (subsequently called the V-method), for testing H_0 against H_1, when F and G are partially specified (the exact meaning of this will be clear later). A test for the situation where F and G are completely unspecified is also put forward.

2. Notation. Suppose $F(x)$ and $G(x)$ to be two cumulative distribution functions on the real axis, $-\infty < x < \infty$, such that their frequency functions exist everywhere. Let $x_1, \cdots, x_m$ and $x_{m+1}, \cdots, x_{m+n}$ denote independent samples from F and G respectively. Now the combined sample from F and G can be represented as a point

$$\mathbf{x} = (x_1, \cdots, x_m, x_{m+1}, \cdots, x_{m+n}) \tag{2.1}$$

in the $m + n$ dimensional Euclidean space $\mathfrak{X}$ of all such points. It follows from the existence of the frequency functions that the probability measure of the set of points $\mathbf{x}$ in $\mathfrak{X}$ defined by $x_i = x_j$ for $i \neq j$ is zero. Next we define on $\mathfrak{X}$ a vector-valued function $\boldsymbol{\gamma}$,

$$\boldsymbol{\gamma}(\mathbf{x}) = (\gamma_1(\mathbf{x}), \cdots, \gamma_i(\mathbf{x}), \cdots, \gamma_{m+n}(\mathbf{x})), \tag{2.2}$$

where $\gamma_i(\mathbf{x})$ is the total number of the components of $\mathbf{x}$ less than or equal to x_i. Thus $\gamma_i(\mathbf{x})$ is the rank of x_i in the combined sample $\mathbf{x} = (x_1, \cdots, x_{m+n})$. Further we arrange the last n components of $\mathbf{x}$, that is $x_{m+1}, \cdots, x_{m+n}$, accord-

Received July 24, 1957; revised June 9, 1959.

[1] This work was submitted as a part of a Ph.D. thesis of the London University. The paper was supported (in part) by funds provided under contract AF-18 (600) 456 with the School of Aviation Medicine, USAF, Randolph Air Force Base, Texas.

ing to their magnitudes as $-\infty < y_1 < \cdots < y_n < \infty$ to define another vector-valued function **a**,

$$\mathbf{a}(\mathbf{x}) = (a_1(\mathbf{x}), \cdots, a_i(\mathbf{x}), \cdots, a_{n+1}(\mathbf{x})), \tag{2.3}$$

where $a_i(\mathbf{x})$ is the total number of the first m components of $\mathbf{x}$ lying between y_{i-1} and y_i, y_0 denoting $-\infty$ and y_{n+1} denoting $+\infty$ for convenience. In addition we define

$$\mathbf{b}(\mathbf{x}) = (b_1(\mathbf{x}), \cdots, b_i(\mathbf{x}), \cdots, b_{m+1}(\mathbf{x})), \tag{2.4}$$

where $b_i(\mathbf{x})$ denotes the number of individuals out of $x_{m+1}, \cdots, x_{m+n}$ lying between the i − 1st and ith ordered individuals from $x_1, \cdots, x_m$; $b_1(\mathbf{x})$ and $b_{m+1}(\mathbf{x})$ being defined analogously to $a_1(\mathbf{x})$ and $a_{n+1}(\mathbf{x})$ in (2.3). Now it is important to note that, given $\mathbf{a}(\mathbf{x})$ in (2.3), $\mathbf{b}(\mathbf{x})$ in (2.4) is uniquely determined and conversely.

For simplicity we write γ for $\gamma(\mathbf{x})$, **a** for $\mathbf{a}(\mathbf{x})$, etc. Now $P(\gamma \mid F, G)$ denotes the probability of obtaining $\mathbf{x}$ such that $\gamma(\mathbf{x}) = \gamma$ given F and G. Similarly, we have $P(\mathbf{a} \mid F. G)$, etc.

3. The most powerful rank test. Following the above notation it is easy to see that

$$P(\gamma \mid F, F) = 1/(m + n)!.$$

Hence the most powerful rank test of the hypothesis $H_0: F = G$, against the simple alternative H_1, that the c.d.f.'s are specifically F and G respectively, has the critical region

$$\gamma: P(\gamma \mid F, G) > \text{const.} \tag{3.2}$$

Since hereafter there is no possibility of confusion, we shall write $P(\gamma)$ for $P(\gamma \mid F, G)$, $P(\mathbf{a})$ for $P(\mathbf{a} \mid F, G)$, etc. Now from the definition of

$$\mathbf{a} = (a_1, \cdots, a_{n+1})$$

in Section 2 it follows that

$$P(\mathbf{a}) = m!\, n!\, P(\gamma), \tag{3.3}$$

which of course is also true under the null hypothesis. Thus the most powerful test (3.2) is associated with the critical region

$$\mathbf{a}: P(\mathbf{a}) > \text{const.} \tag{3.4}$$

Suppose further that we have a function θ such that

$$G(x) = \theta(F(x))^{2} \tag{3.5}$$

[2] Here one should avoid the mistake of assuming that F is a uniform distribution on (0, 1).

for every x and

$$\theta'(F) = (\partial/\partial F)\theta(F) \tag{3.6}$$

exists for every F, $0 \leqq F \leqq 1$.

THEOREM 3. *We have*

$$P(\mathbf{a}) = \frac{m!\, n!}{\prod_{i=1}^{n+1} a_i\,!} \int \cdots_D \int \prod_{i=1}^{n+1} p_i^{a_i} \cdot \prod_{i=1}^{n} \theta'(p_1 + \cdots + p_i) \prod_{i=1}^{n} dp_i\,, \tag{3.7}$$

where $p_{n+1} = 1 - p_1 - \cdots - p_n$ *and the domain of integration* D *is*

$$D = \{p_1, \cdots, p_n : 0 \leqq p_i \leqq 1, \sum p_i \leqq 1, \quad i = 1, \cdots, n\}.$$

The theorem can be easily derived from a result of Hoeffding's [1], top of p. 88. It is important to note that, in the above formulae, $P(\mathbf{a})$ depends on F and G only through θ, or rather θ', and so does the corresponding test (3.4) for testing against the alternative $G = \theta(F)$. It is, however, seldom possible to evaluate the integral on the right side of (3.7). For this and other reasons we shall in the next section put forward another rank test which depends on θ' alone.

4. The V-test. Consider the following degenerate case of testing a *simple* hypothesis H_0 against a *simple* alternative H_1. (Note that these are not the same as the hypotheses in Section 1.)

H_0: Both samples $x_1, \cdots, x_m$ and $y_1, \cdots, y_n$ are drawn from a common specified c.d.f. F.

H_1: The sample $x_1, \cdots, x_m$ is drawn from F, while the sample $y_1, \cdots, y_n$ is drawn from another specified c.d.f. G.

The most powerful test (not the most powerful rank test) in this case is independent of the sample from F and in fact is given by the critical region

$$y_1, \cdots, y_n : \frac{g(y_1) \cdots g(y_n)}{f(y_1) \cdots f(y_n)} > \text{const.}, \tag{4.1}$$

where g and f are the frequency functions of G and F respectively. Again, if, as in (3.5), we have $G(x) = \theta(F(x))$ for all x and if $\theta'(F)$ exists for $0 \leqq F \leqq 1$, then the critical region (4.1) can be expressed as

$$F(y_1), \cdots, F(y_n) : \prod_{i=1}^{n} \theta'(F(y_i)) > \text{const.} \tag{4.2}$$

For instance, when F and G are normal distributions with unit variance, the mean of F being 0 and that of G being δ, it can be seen from equation (6.4) of Section 6 that (4.2) can be written as

$$F(y_1), \cdots, F(y_n) : \sum_{i=1}^{n} \psi^{-1}(F(y_i)) > \text{const.},$$

where ψ^{-1} is the functional inverse of the normal integral as defined in (6.1). Since, however, F itself is the standard normal c.d.f., the above critical region is identical with

$$y_1, \cdots, y_n: \sum_1^n y_i > \text{const.}$$

This is the usual optimum test for the normal mean when the variance is known.

Now the critical region (4.2) clearly depends on F in addition to θ'. However, an approximation to (4.2) which depends on θ' alone can be worked out as follows. For a given second sample $(y_1, \cdots, y_n)$ the quantities

$$\left| F(y_i) - \frac{a_1 + \cdots + a_i}{m} \right| \tag{4.3}$$

can simultaneously be made arbitrarily small for $i = 1, \cdots, n$, with as large a probability as we please, by increasing sufficiently the size of the first sample. Hence

$$a_1, \cdots, a_n: \prod_{i=1}^n \theta'\left(\frac{a_1 + \cdots + a_i}{m}\right) > \text{const.} \tag{4.4}$$

is the suggested approximation to (4.2). Note that (4.4) is a non-parametric test, depending only on the order relationships within the sample.

Now for testing the null hypothesis $H_0: G = F$ against the alternative $H_1: G = \theta(F)$, we propose the V-statistic

$$V(\mathbf{a}) = \prod_1^n \theta'\left(\frac{1 + a_1 + \cdots + a_i}{m + 2}\right) \tag{4.5}$$

or a suitable monotonic increasing function of the right side of (4.5), the corresponding V-test being defined by the critical region

$$\mathbf{a}: V(\mathbf{a}) > \text{const.} \tag{4.6}$$

This will also be referred to as the V-method of obtaining tests. The motivation for the V-method is made clear in the preceding paragraph. In fact, (4.6) is obtained from (4.4), with a small modification to prevent the V-statistic from assuming infinitely large values.

About the intuitive appeal of the V-method, it may be said that some tests derived by its application, with a slight difference, have already been proposed by different authors on more or less intuitive grounds. This will be verified in some of the subsequent sections, where V-tests are compared with some of the known tests, including the one given by the statistic (3.7).

Further it would appear from the above discussion that, though the V-method is put forward as a sure method of obtaining tests for simple alternatives, it can in some cases yield tests even for composite alternatives. This can be checked from the illustrations to follow.

5. Lehmann's alternatives. In this case it is assumed that

$$G = \theta(F) = F^k, \tag{5.1}$$

where $k > 1$. From (5.1) we have

$$\theta'(F) = kF^{k-1}. \tag{5.2}$$

Hence from (4.6) the V-test for the present situation is given by the critical region

$$\mathbf{a}: V(\mathbf{a}) > \text{const.}, \tag{5.3}$$

where the V-statistic is defined by

$$V(\mathbf{a}) = \prod_{1}^{n}(1 + a_1 + \cdots + a_i). \tag{5.4}$$

It is interesting to see that the V-test does not depend upon k for $k > 1$. The test for $k < 1$ can be obtained similarly.

I. R. Savage [3] has studied very extensively the alternatives in (5.1). He also has tabulated the probabilities $P(\mathbf{a})$ in (3.7), when $\theta(F) = F^k$, for different values of k and different sample sizes.

In Table 1 we give the 5 **a**'s corresponding to the largest values of $V(\mathbf{a})$ in (5.4). Now it so happens that the same **a**'s are the ones having the largest probabilities $P(\mathbf{a})$ in (3.7), for all values of $k > 1$ considered by Savage [3]. For these values of k, the ordering of the $P(\mathbf{a})$'s mentioned above is also the same. Hence in Table 1 the $P(\mathbf{a})$'s are reproduced from Savage's table for just one set of the values of k. From them we can construct the most powerful tests, defined by the critical regions $\mathbf{a}:P(\mathbf{a}) >$ const., as in (3.4), up to the significance level of 25 per cent. Each of these tests, as can be seen from Table 1, will be equal in power to the corresponding V-test. Further from Savage's table in [3] it appears that the performance of the V-test for sample sizes other than those considered in Table 1 is equally good.

The statistic that Savage [3] has proposed for the present problem is, in our notation,

$$T(\mathbf{a}) = \sum_{i=1}^{n+1} \sum_{j=0}^{a_i} \frac{a_1 + \cdots + a_{i-1} + j}{a_1 + \cdots + a_{i-1} + j + (i-1)}, \tag{5.5}$$

the corresponding critical region being

$$\mathbf{a}: T(\mathbf{a}) < \text{const.} \tag{5.6}$$

The 5 smallest values of $T(\mathbf{a})$ are reproduced in Table 1. It is difficult to see any connection between (5.4) and (5.6). Now Savage [3] has proved that for the cases $m = 2$, $n = 3$ and $m = 2$, $n = 4$, dealt with in Table 1, the simple ordering of the probabilities $P(\mathbf{a})$ in (3.7) when $\theta(F) = F^k$ does not depend on k for $k > 1$, and is given by the statistic $T(\mathbf{a})$ in (5.5).

TABLE 1

a	$P(\mathbf{a})$ for $G = F^k$ $k = 3.7769$	$V(\mathbf{a})$ as in (5.4)	$T(\mathbf{a})$ as in (5.5)
	$m = 2$ $n = 3$		
(2, 0, 0, 0)	.4394	8	1.4333
(1, 1, 0, 0)	.1840	4	1.9333
(1, 0, 1, 0)	.1242	2	2.2667
(1, 0, 0, 1)	.0963	1	2.5167
(0, 2, 0, 0)	.0487	0	2.9333

$m = 2$ $n = 4$

a	$P(\mathbf{a})$ for $G = F^k$ $k = 3.6173$	$V(\mathbf{a})$ as in (5.4)	$T(\mathbf{a})$ as in (5.5)
(2, 0, 0, 0, 0)	.3743	16	2.1000
(1, 1, 0, 0, 0)	.1621	8	2.6000
(1, 0, 1, 0, 0)	.1106	4	2.9333
(1, 0, 0, 1, 0)	.0862	2	3.1833
(1, 0, 0, 0, 1)	.0716	1	3.3833

$m = 3$ $n = 3$

a	$P(\mathbf{a})$ for $G = F^k$ $k = 3.0546$	$V(\mathbf{a})$ as in (5.4)	$T(\mathbf{a})$ as in (5.5)
(3, 0, 0, 0)	.2549	27	1.1500
(2, 1, 0, 0)	.1513	18	1.4833
(2, 0, 1, 0)	.1130	12	1.7333
(2, 0, 0, 1)	.0922	8	1.9333
(1, 2, 0, 0)	.0746	9	1.9833

Next we consider the alternative

$$G = \theta(F) = \lambda F^2 + (1 - \lambda)F, \qquad 0 < \lambda < 1. \tag{5.7}$$

As proved by Lehmann [2], the well-known Wilcoxon test is optimum against the alternative hypothesis (5.7) for very small values of λ. It is interesting to see that the V-statistic (4.5) for the present situation is again Wilcoxon's statistic. For from (5.7) we have

$$\theta'(F) = 2\lambda F + (1 - \lambda). \tag{5.8}$$

Now for very small values of λ, (5.8) can be written as

$$\theta'(F) = e^{2\lambda F}. \tag{5.9}$$

Hence the V-statistic is

$$V(\mathbf{a}) = \sum_{i=1}^{n} \frac{1 + a_1 + \cdots + a_i}{m + 2}, \tag{5.10}$$

the corresponding V-test being defined by the critical region

$$\mathbf{a} : V(\mathbf{a}) > \text{const.} \tag{5.11}$$

This is the usual one-sided Wilcoxon test.

6. Normal alternatives. Let F and G be the following distributions:

$$F(x) = (2\pi)^{-\frac{1}{2}} \int_{-\infty}^{x} \exp\left(-\tfrac{1}{2}h^2\right) dh = \psi(x) \tag{6.1}$$

and

$$\begin{aligned} G(x) &= (2\pi)^{-\frac{1}{2}} \int_{-\infty}^{x} \exp\left(-\tfrac{1}{2}(h + \delta)^2\right) dh \\ &= (2\pi)^{-\frac{1}{2}} \int_{-\infty}^{x+\delta} \exp\left(-\tfrac{1}{2}h^2\right) dh = \psi(x + \delta). \end{aligned} \tag{6.2}$$

Now though for convenience of notation it has been assumed above that both F and G have variances equal to 1, the following arguments are valid for any unknown common variance σ^2. Write

$$G(x) = \theta(F(x)). \tag{6.3}$$

Then it is seen from (6.1) and (6.2) that

$$\begin{aligned} \theta'(F) &= \left(\frac{\partial G}{\partial x}\right) \Big/ \left(\frac{\partial F}{\partial x}\right) = \exp\left(-\delta^2 - 2\delta\, x\right) \\ &= \exp\left(-\delta^2 - 2\delta\psi^{-1}(F)\right). \end{aligned} \tag{6.4}$$

Next since $\theta'(F)$ in (6.4) depends on δ it follows from (3.7) that the most powerful rank order test of H_0 against H_1 may in general depend on δ, though it has been proved to be independent of δ for all sufficiently small values of δ. In fact, it is then Hoeffding's C_1 criterion [1]. Furthermore, some empirical sampling investigations by Teichroew [4] suggest that the most powerful rank order test may exist uniquely for all $\delta > 0$. The situation for $\delta < 0$ is similar. However, no theoretical result is available in that direction.

It is interesting to see that in the present case the V-test obtained from (6.4) above does not depend upon δ. It follows from (6.4) and (4.6) that the V-test is defined by the critical region

$$\mathbf{a} : V(\mathbf{a}) > \text{const.}, \tag{6.5}$$

where the V-statistic is defined by

$$V(\mathbf{a}) = -\sum_{i=1}^{n} \psi^{-1}\left(\frac{1 + a_1 + \cdots + a_i}{m + 2}\right). \tag{6.6}$$

In the following illustrations the relative frequencies are reproduced from Teichroew's experiments [4] for some specified δ. However, the ordering of the $\mathbf{a}$'s by their relative frequencies is more or less the same for the other values of δ considered by Teichroew. It can be seen from Table 2 that the ordering of $\mathbf{a}$ by $V(\mathbf{a})$ in (6.6) is nearly the same as that by the relative frequencies. The performance of $V(\mathbf{a})$ is as good for the other sample sizes of Teichroew [4] as for those considered in Table 2. It must, however, be said that for all these illustrative cases Hoeffding's C_1-test [1] or van der Waerden's X-test [5] has comparably good performance.

TABLE 2

$m = 3 \qquad n = 2$

$\mathbf{a}$	$V(\mathbf{a})$ as in (6.6)	$\sigma = 0.75$[3] Relative Frequency	$X(\mathbf{a})$ as in (6.7)	Hoeffding's C_1
(3, 0, 0)	−1.68	2.25	−1.40	−1.66
(2, 1, 0)	−1.09	3.45	−0.97	−1.16
(1, 2, 0)	−0.59	4.45	−0.54	−0.66
(2, 0, 1)	−0.50	5.45	−0.45	−0.50
(1, 1, 1)	−0.00	7.40	−0.00	−0.00
(0, 3, 0)	0.00	8.10	0.00	0.00
(1, 0, 2)	0.50	11.15	0.45	0.50
(0, 2, 1)	0.59	12.00	0.54	0.66
(0, 1, 2)	1.09	18.45	0.97	1.16
(0, 0, 3)	1.68	27.30	1.40	1.66

Now van der Waerden's X-statistic, in our notation is defined as follows.

(6.7)
$$X(\mathbf{a}) = -\sum_{i=1}^{n} \psi^{-1}\left(\frac{a_1 + \cdots + a_i + i}{m + n + 1}\right).$$

When m is large enough compared to n, $X(\mathbf{a})$ is nearly equal to $V(\mathbf{a})$ in (6.6). For the case considered in Table 2, we see that even for $m = 3$ and $n = 2$, the critical regions given by the two statistics are identical. It may also be of interest to note that the X-test has been shown to be asymptotically equivalent to the C_1-test.

Next we consider two normal populations with the same mean but different variances,

(6.8)
$$\begin{aligned} F(x) &= (2\pi\sigma_1^2)^{-\frac{1}{2}} \int_{-\infty}^{x} \exp\left(-\tfrac{1}{2}h^2/\sigma_1^2\right) dh \\ &= (2\pi)^{-\frac{1}{2}} \int_{-\infty}^{x/\sigma_1} \exp\left(-\tfrac{1}{2}h^2\right) dh = \psi(x/\sigma_1) \end{aligned}$$

[3] These are Monte Carlo results; see [4].

and

$$\begin{aligned} G(x) &= (2\pi\sigma_2^2)^{-\frac{1}{2}} \int_{-\infty}^{x} \exp\,(-\tfrac{1}{2}h^2/\sigma_2^2)\, dh \\ &= (2\pi)^{-\frac{1}{2}} \int_{-\infty}^{x/\sigma_2} \exp\,(-\tfrac{1}{2}h^2)\, dh = \psi(x/\sigma_2). \end{aligned} \tag{6.9}$$

Consider the problem of testing $H_0 : \sigma_1^2 = \sigma_2^2$ against $\sigma_1^2 < \sigma_2^2$.

If we put $G(x) = \theta(F(x))$ from (6.8) and (6.9) we have

$$\begin{aligned} \theta'(F) &= \left(\frac{\partial G}{\partial x}\right) \Big/ \left(\frac{\partial F}{\partial x}\right) = \exp \tfrac{1}{2}(x^2/\sigma_1^2 - x^2/\sigma_2^2) \\ &= \exp \{\tfrac{1}{2}(\psi^{-1}(F))^2 - \tfrac{1}{2}(\sigma_1^2/\sigma_2^2)(\psi^{-1}(F))^2\} \\ &= \exp \tfrac{1}{2}(1 - k^2)(\psi^{-1}(F))^2 \end{aligned} \tag{6.10}$$

where $\sigma_1^2/\sigma_2^2 = k^2$. Hence, to test $\sigma_1^2 < \sigma_2^2$, the critical region of the V-test would be, from (6.10) and (4.6),

$$\mathbf{a} : V(\mathbf{a}) > \text{const.} \tag{6.11}$$

Here the V-statistic is given by

$$V(\mathbf{a}) = \sum_1^n \left(\psi^{-1}\left(\frac{1 + a_1 + \cdots + a_i}{m + 2}\right)\right)^2. \tag{6.12}$$

Unfortunately we do not have the necessary Monte Carlo frequencies to judge empirically for small samples the performance of $V(\mathbf{a})$ in (6.12). Later on, we shall have an interesting comparison of $V(\mathbf{a})$ in (6.12) with some other statistics. Furthermore a different application of the V-method, suggested by the theorem in the next section, gives another test for the same problem of testing the variance ratio of two normal populations with the same mean. This test is discussed at the end of the next section.

7. A theorem about the V-statistic. Consider a case where $G = \theta(F)$ and

$$\frac{\partial^2 \theta(F)}{\partial F^2} \geqq 0 \tag{7.1}$$

for all F, which implies that

$$\frac{d}{dx} \frac{\partial G/\partial x}{\partial F/\partial x} \geqq 0. \tag{7.2}$$

Now (7.2) is precisely the monotone likelihood ratio condition of Savage [3]. Thus it follows from Theorem 6.1 of Savage [3] that if

$$\begin{aligned} \mathbf{a} &= (\cdots, a_i, \cdots, a_j, \cdots) \\ \mathbf{a}' &= (\cdots, a_i + 1, \cdots, a_j - 1, \cdots), \end{aligned} \tag{7.3}$$

the other components of $\mathbf{a}$ and $\mathbf{a}'$ being identical and $i < j$, then

$$P(\mathbf{a}') \geqq P(\mathbf{a}). \tag{7.4}$$

Further, from (4.5) and (7.1) we have

$$V(\mathbf{a}') \geqq V(\mathbf{a}). \tag{7.5}$$

Now the V-statistic has been defined for a simple alternative $G = \theta(F)$, and as such we can order simply the vectors $\mathbf{a}$, according to the values $V(\mathbf{a})$. Similarly the vectors $\mathbf{a}$ can be ordered simply according to the values of $P(\mathbf{a})$ in (3.7). Let $\mathbf{S}$ be a set of vectors which could be arranged in such a way that for any pair of successive vectors, $\mathbf{a}_k$ and $\mathbf{a}_{k+1}$ say, $\mathbf{a}_k$ is related to $\mathbf{a}_{k+1}$ as $\mathbf{a}$ to $\mathbf{a}'$ in (7.3) for some $i, j (i < j)$, where i, j can vary with k. Then from (7.3), (7.4) and (7.5) we have

THEOREM 7.[4] *For any function θ satisfying* (7.1), *a simple ordering of* $\mathbf{S}$, *according to $P(\mathbf{a})$ is identical with a simple ordering given by $V(\mathbf{a})$.*

There is a similar theorem if instead of (7.1) we have

$$\frac{\partial^2 \theta(F)}{\partial F^2} \leqq 0.$$

Now as already noted by Savage [3], for both Lehmann's alternative in (5.1) and the normal alternative in (6.1)–(6.2), the monotone likelihood ratio condition (7.1) is fulfilled. This can also be checked from (5.2) and (6.4). Hence Theorem 7 above is applicable in both cases.

On the other hand, for the alternative in (6.8)–(6.9), the condition (7.1) is not fulfilled. This can be seen from (6.10). We therefore substitute

$$\tfrac{1}{2}x^2 = z. \tag{7.6}$$

(This transformation is due to the referee.) Now the c.d.f.'s of z, viz. $F(z)$ and $G(z)$, corresponding to (6.8) and (6.9), are

$$F(z) = (\pi)^{-\frac{1}{2}} \int_0^{z/\sigma_1^2} e^{-h}\, h^{-\frac{1}{2}}\, dh \tag{7.7}$$

and

$$G(z) = (\pi)^{-\frac{1}{2}} \int_0^{z/\sigma_2^2} e^{-h}\, h^{-\frac{1}{2}}\, dh. \tag{7.8}$$

Now putting $G = \theta(F)$, we have

$$\begin{aligned} \theta'(F) &= \text{const.} \exp\left[(1 - \sigma_1^2/\sigma_2^2) z/\sigma_1^2\right] \\ &= \text{const.} \exp\left[(1 - \sigma_1^2/\sigma_2^2)\sqrt{2} I^{-1}(F, -\tfrac{1}{2})\right] \end{aligned} \tag{7.9}$$

[4] The author is indebted to the referee for an important clarification in this theorem.

where I^{-1} is the inverse of I in (7.14). It is clear from (7.7), (7.8) and (7.9) that the monotone likelihood ratio condition of (7.1) is now satisfied. Next referring to the notation of Section 2, it can be seen that a point

$$\mathbf{x} = (x_1, \cdots, x_{m+n})$$

in $\mathfrak{X}$ is transformed by the substitution (7.6), i.e., $\frac{1}{2}x_i^2 = z_i$, $i = 1, \cdots, m + n$, into a point

$$\mathbf{z} = (z_1, \cdots, z_{m+n}) \tag{7.10}$$

in $\mathcal{Z}$, say. Further, exactly analogous to $\mathbf{a} = (a_1(\mathbf{x}), \cdots, a_{n+1}(\mathbf{x}))$ in (2.3), we can define on $\mathcal{Z}$ a vector-valued function

$$\mathbf{c}(\mathbf{z}) = (c_1(\mathbf{z}), \cdots, c_{n+1}(\mathbf{z})). \tag{7.11}$$

Now from (4.5), (7.9), and (7.11) the V-test for testing the null hypothesis against the alternative in (7.7)–(7.8) is defined by the critical region

$$\mathbf{c}: V(\mathbf{c}) > \text{const.} \tag{7.12}$$

where the V-statistic is given by

$$V(\mathbf{c}) = \sum_{i=1}^{n+1} I^{-1}\left(\frac{c_1 + \cdots + c_i + 1}{m + 2}, -\frac{1}{2}\right), \tag{7.13}$$

I^{-1} as before being the inverse of

$$I(u, -\tfrac{1}{2}) = (\pi)^{-\frac{1}{2}} \int_0^{u\sqrt{\frac{1}{2}}} e^{-h}\, h^{-\frac{1}{2}}\, dh. \tag{7.14}$$

(The values of (7.14) for different u's are tabulated in *Tables of the Incomplete Γ-Function*, Cambridge University Press, 1946.)

Two suggestions for comparing the statistics (7.13) and (6.12) are as follows: (i) It will be remembered from (6.8) and (6.9) that for convenience of notation we assumed the common mean of the two populations to be zero. Actually, if μ is the common mean, the transformation (7.6) would be $\frac{1}{2}(x - \mu)^2 = z$, which means the vector c in (7.11) depends on μ. That is, contrary to the test (6.11), the test (7.12) cannot be worked out unless μ is known. (ii) Theorem 7 above is valid for (7.13) while it is not valid for (6.12).

Now Theorem 7 implies *some* justification for using any of the V-statistics, such as (5.4) or (6.6), where θ satisfies (7.1), for testing against a wider alternative hypothesis $G = \theta(F)$, which does not specify anything about θ, excepting that it satisfies the monotone likelihood ratio condition (7.1).

In the next section we develop a test of the null hypothesis $H_0: F = G$ against the general alternative $H_1: F \neq G$.

8. ϕ-test. Substituting in (3.7)

$$\phi(\mathbf{a}/\mathbf{p}) = \frac{m!}{\prod_1^{n+1} a_i\,!} \prod_1^{n+1} (p_i)^{a_i}, \tag{8.1}$$

we have

$$(8.2)\qquad P(\mathbf{a}) = \int \underset{D}{\cdots} \int \phi(\mathbf{a}/\mathbf{p}) n! \prod_1^n \theta'(p_1 + \cdots + p_i) \prod_1^n dp_i .$$

Now in (8.1), a_i/m is the maximum likelihood estimate of p_i, $i = 1, \cdots, n$. Therefore if

$$(8.3)\qquad \phi(\mathbf{a}) = \frac{m!}{\prod_1^{n+1} a_i!} \prod_1^{n+1} (a_i/m)^{a_i},$$

it follows that

$$(8.4)\qquad \phi(\mathbf{a}) \geqq \phi(\mathbf{a}/\mathbf{p}).$$

Thus from (8.1), (8.2) and (8.4) we have

$$(8.5)\qquad P(\mathbf{a}) < \phi(\mathbf{a}) \int \underset{D}{\cdots} \int n! \prod_1^n \theta'(p_1 + \cdots + p_i) \prod_1^n dp_i .$$

Now using the transformation $p_1 + \cdots + p_i = q_i$, $i = 1, \cdots, n$ we have

$$(8.6)\qquad \int \underset{D}{\cdots} \int \theta'(p_1 + \cdots + p_i) \prod_1^n dp_i = \int \underset{D'}{\cdots} \int \prod_1^n \theta'(q_i) \prod_1^n dq_i ,$$

where

$$D' = \{q_1, \cdots, q_n : 0 \leqq q_i \leqq 1, q_{i-1} \leqq q_i, i = 1, \cdots, n\},$$

q_0 denoting 0. Integrating the right hand side of (8.6) term by term and noting that $\theta(0) = 0$ and $\theta(1) = 1$, we have

$$(8.7)\qquad \int \underset{D'}{\cdots} \int \prod_1^n \theta'(q_i) \prod_1^n dq_i = 1/n!.$$

Hence from (8.5), (8.6) and (8.7) it follows that

$$(8.8)\qquad P(\mathbf{a}) \leqq \phi(\mathbf{a}).$$

Next we recollect the definition of

$$(8.9)\qquad \mathbf{b} = (b_1, \cdots, b_{m+1})$$

in (2.4) of Section 2. It has also been noted that **a** defines **b** uniquely and conversely. Therefore we have

$$(8.10)\qquad P(\mathbf{a}) = P(\mathbf{b}).$$

Further it follows from (8.3) and (8.8) that if

$$(8.11)\qquad \phi(\mathbf{b}) = \frac{n!}{\prod_1^{m+1} b_i!} \prod_1^{m+1} (b_i/n)^{b_i},$$

then

$$(8.12)\qquad P(\mathbf{b}) \leqq \phi(\mathbf{b}).$$

We can write (8.8), (8.10) and (8.12) together as

$$P(\mathbf{a}) = P(\mathbf{b}) \leqq \min (\phi(\mathbf{a}), \phi(\mathbf{b})). \tag{8.13}$$

Now define the ϕ-statistic as the minimum of $\phi(\mathbf{a})$ and $\phi(\mathbf{b})$ i.e.,

$$\phi = \min (\phi(\mathbf{a}), \phi(\mathbf{b})) = \phi(\mathbf{a}, \mathbf{b}). \tag{8.14}$$

Then the ϕ-test for testing $H_0:F = G$ against $H_1:F \neq G$ is defined by the critical region

$$\phi > \text{const.} \tag{8.15}$$

The motivation for this test lies in the inequality (8.13) and the fact that in a degenerate case when there exist two numbers u and v, $v > u$, such that $F(u) = 1$ and $G(v) = 0$, then

$$P(\mathbf{a}) = P(\mathbf{b}) = \phi = 1 \tag{8.16}$$

with probability equal to unity.

Wolfowitz [6] proposed a test statistic equivalent to

$$W = \phi(\mathbf{a}) \cdot \phi(\mathbf{b}) \tag{8.17}$$

for testing $H_0:F = G$ against $F \neq G$. In the numerical illustrations in section 10, the ϕ-statistic defined in (8.14) appears to be better than Wolfowitz' statistic in (8.17), though possibly quite a few statements made hereafter in case of the ϕ-statistic may also hold for W in (8.17). A simple method for computing ϕ-statistic could be obtained from one suggested by Wolfowitz [6] for his statistic W above.

9. Some properties of the ϕ-statistic. Let

$$\mathbf{a} = (\cdots, a_i, \cdots, a_j, \cdots) \tag{9.1}$$

$$\mathbf{a}' = (\cdots, a_i + 1, \cdots, a_j - 1, \cdots) \tag{9.2}$$

be two vectors with their ith and jth components ($i < j$) as shown ($a_i, a_j \geqq 1$), other components of $\mathbf{a}$ being equal to the corresponding ones of $\mathbf{a}'$. Now from (8.3) we have

$$\frac{\phi(\mathbf{a}')}{\phi(\mathbf{a})} = \frac{(1 + 1/a_i)^{a_i}}{(1 + 1/(a_j - 1))^{a_j - 1}}. \tag{9.3}$$

Thus from the monotonicity of function $(1 + 1/z)^z$, if in (9.1) and (9.2)

$$a_i \geqq a_j, \tag{9.4}$$

we have

$$\phi(\mathbf{a}') > \phi(\mathbf{a}). \tag{9.5}$$

Further, it follows from the above argument and the symmetry of $\phi(\mathbf{a})$ in

$a_1, \cdots, a_{n+1}$ that if

$$\mathbf{a}'' = (\cdots, a_i + a_j, \cdots, 0, \cdots) \tag{9.6}$$

with its ith and jth components as shown, other components being equal to the corresponding ones of $\mathbf{a}$ in (9.1), then

$$\phi(\mathbf{a}'') > \phi(\mathbf{a}) \tag{9.7}$$

regardless of the condition (9.4) above. Again as in (8.9) define $\mathbf{b}$ and $\mathbf{b}'$ from $\mathbf{a}$ and $\mathbf{a}'$ in (9.1) and (9.2). Then it follows from arguments similar to the above that

$$\phi(\mathbf{b}') \geqq \phi(\mathbf{b}). \tag{9.8}$$

Thus from (9.5) and (9.8) we have for $\mathbf{a}$, $\mathbf{b}$ and $\mathbf{a}'$, $\mathbf{b}'$ defined by (9.1) and (9.2) above, *provided* (9.4) holds,

$$\phi(\mathbf{a}', \mathbf{b}') \geqq \phi(\mathbf{a}, \mathbf{b}), \tag{9.9}$$

ϕ being given by (8.14). Similarly from (9.6) and (9.7) we have, *regardless* of (9.4),

$$\phi(\mathbf{a}'', \mathbf{b}'') \geqq \phi(\mathbf{a}, \mathbf{b}). \tag{9.10}$$

Now suppose F and G are such $G(x) = \theta(F(x))$ and

$$\frac{\partial^2\theta(F)}{\partial F^2} \geqq 0 \tag{9.11}$$

for all F. Then as said in Section 7, (9.11) implies

$$\frac{d}{dx}\frac{\partial G/\partial x}{\partial F/\partial x} \geqq 0 \tag{9.12}$$

for all x. Now (9.12) is the monotone likelihood condition in terms of Savage [3]. Thus it follows from Theorem 6.1 of Savage [3] that if the condition (9.12) or (9.11) above is satisfied,

$$P(\mathbf{a}') \geqq P(\mathbf{a}) \tag{9.13}$$

for $\mathbf{a}$ and $\mathbf{a}'$ in (9.1) and (9.2) assuming $i < j$.

Now since $\mathbf{b}$ in (8.9) is uniquely determined by $\mathbf{a}$, we may consider the statistic $\phi(\mathbf{a}, \mathbf{b})$ in (8.14) as a function of $\mathbf{a}$ alone, ignoring $\mathbf{b}$. Thus we can get a simple ordering of the vectors $\mathbf{a}$, according to the values of $\phi(\mathbf{a}, \mathbf{b})$. Similarly for a simple alternative $G = \theta(F)$ we can have a simple ordering of vectors $\mathbf{a}$, according to the statistic $P(\mathbf{a})$ in (3.7). Consider a *fixed* vector

$$\mathbf{a} = (\cdots, a_i, \cdots, a_j, \cdots)$$

having $a_i = a_j$, $(i < j)$ and let $\mathbf{a}'$ be the vector obtained from $\mathbf{a}$ by replacing the ith and jth coordinates of $\mathbf{a}$ by $a_i + k$ and $a_j - k$ respectively. For the same i and j, allowing k to take values $1, 2, \cdots, a_j$, we get a set of vectors $\mathbf{a}'$ which we denote by $\mathbf{S}'$. Then we have from (9.1), (9.2), (9.4), (9.9), (9.11) and (9.13)

THEOREM 9. *For any function θ satisfying* (9.11), *a simple ordering of* $\mathbf{S}'$ *by the statistic* $P(\mathbf{a}')$ *in* (3.7) *is identical with a simple ordering given by* $\phi(\mathbf{a}', \mathbf{b}')$ *in* (8.14).

We shall have a similar theorem, if instead of (9.11) we have

$$\partial^2 G/\partial F^2 \leqq 0. \tag{9.14}$$

Theorem 9 above also establishes a relation between the ϕ-statistic in (8.14) and the V-statistic defined in (4.5) for which Theorem 7 was true.

10. Numerical illustration. In Table 3 we have given values of different statistics, for comparison. As already noted in Section 5, the ranking of the

TABLE 3

$m = 3 \quad n = 2$

1	2	3	4	5	6	7	8	9	10
				F, G Normal					
a	$G = F^{1/k}$ $k = 7.1663$ $P(\mathbf{a})$ (3.7)	X-stat. (6.7)	C_1	equal var. $\delta =$ 0.75 Rel. freq. (Monte Carlo)	equal var. $\delta > 0$ V-stat. (6.6)	equal means $\sigma_1 < \sigma_2$ V-stat. (6.12)	ϕ-stat. (8.14)	W-stat. (8.17)	Smirnov stat.
(3, 0, 0)	.0038	−1.40	−1.66	2.25	−1.68	1.41	1.00	1.00	1.00
(2, 1, 0)	.0054	−0.97	−1.16	3.45	−1.09	0.77	0.44	0.22	0.66
(1, 2, 0)	.0093	−0.54	−0.66	4.45	−0.59	0.77	0.44	0.22	0.50
(2, 0, 1)	.0073	−0.43	−0.50	5.45	−0.50	0.12	0.44	0.44	0.66
(1, 1, 1)	.0128	−0.00	−0.00	7.40	−0.00	0.12	0.22	0.11	0.33
(0, 3, 0)	.0667	0.00	0.00	8.10	0.00	1.41	0.50	0.50	0.50
(1, 0, 2)	.0214	0.43	0.50	11.15	0.50	0.12	0.44	0.44	0.66
(0, 2, 1)	.0919	0.54	0.66	12.00	0.59	0.77	0.44	0.22	0.50
(0, 1, 2)	.1537	0.97	1.16	18.45	1.09	0.77	0.44	0.22	0.66
(0, 0, 3)	.6277	1.40	1.66	27.30	1.68	1.41	1.00	1.00	1.00

vectors $\mathbf{a}$ according to the probabilities $P(\mathbf{a})$ in column 2 remains the same for all values of $k > 1$ in Savage's Table [3]. Column 5 gives the relative frequencies in Teichroew's experiments [4]. Again the ranking of $\mathbf{a}$'s according to the relative frequencies, for all values of δ considered in [4], remains nearly the same.

The ranking of vectors $\mathbf{a}$ in column 1, by the V-statistic (for testing the variance ratio) in column 7 agrees better with the ranking by the ϕ statistic in column 8 than with the rankings by the statistics in columns 9 and 10 respectively. It should be noted that the statistics X, C_1, and V in columns 3, 4 and 6 respectively are meant to test one-sided alternatives. We can however construct, *intuitively*, two sided tests based on them, having the corresponding critical regions $| X | >$ const., $| C_1 | >$ const. and $| V | >$ const. Now in Table 3, the ranking of the vectors $\mathbf{a}$ in column 1 by any of the statistics $| X |$, $| C_1 |$, and $| V |$ agrees better with the ranking by the ϕ statistic in column 8 than with the rankings by the statistics in columns 9 or 10 respectively. Of course more empirical investigation is necessary to arrive at practically usable conclusions.

11. Some possibilities for the asymptotic behavior of the V and ϕ-statistics. This section consists of a few *conjectures* or *guesses*. From (4.5) we have,

$$\log V = \sum_1^n \log \theta' \left(\frac{1 + a_1 + \cdots + a_i}{m + 2} \right). \tag{11.1}$$

Now fixing the second sample $y_1, \cdots, y_n$, let the size m of the first sample go to ∞. Then in view of (4.3), for both null and alternative hypotheses, with probability 1,

$$\log V = \sum_1^n \log \theta'(F(y_i)). \tag{11.2}$$

Now the asymptotic normality of $\log V$ in (11.1) as $n \to \infty$ could possibly be derived from the fact that the $F(y)$'s in (11.2) are distributed identically and independently on both the null and alternative hypotheses. On the null hypothesis the $F(y)$'s are distributed rectangularly, $0 \leqq F \leqq 1$. On the alternative hypothesis, the frequency function of $F(y)$ is $\theta'(F)$, $0 \leqq F \leqq 1$. This also suggests that the mean values of the asymptotic distribution of $\log V$ in (11.1) might be

$$n \int_0^1 (\log \theta'(F))\, dF, \tag{11.3}$$

$$n \int_0^1 (\log \theta'(F))\theta'(F)\, dF \tag{11.4}$$

on the null and alternative hypotheses respectively. Similarly the variances, on the null and alternative hypotheses, could possibly be expressed as follows.

$$n \left\{ \int_0^1 (\log \theta'(F))^2\, dF - \left[\int_0^1 \log \theta'(F)\, dF \right]^2 \right\} \tag{11.5}$$

$$n \left\{ \int_0^1 (\log \theta'(F))^2 \theta'(F)\, dF - \left[\int_0^1 (\log \theta'(F))\theta'(F)\, dF \right]^2 \right\}. \tag{11.6}$$

The above integrals could be evaluated by means of numerical integration.

Next it seems from (4.3) that when $m \to \infty$, n being fixed, the power of the V-test is equal to that of the corresponding optimum parametric test. In particularly van der Waerden's X-test (6.7), which is equivalent to the corresponding V-test (6.6) as $m \to \infty$, has been proved in [5] to be asymptotically as powerful as the t-test.

The asymptotic distribution of the ϕ-statistic in (8.14) is difficult to guess. However $\phi(\mathbf{a})$ in (8.3) seems to be relatively easy to handle. From (8.3) we have

$$(11.7)\quad \log \phi(\mathbf{a}) = \log m! - m \log m + \sum_{1}^{n+1} a_i \log a_i - \sum_{1}^{n+1} \log a_i\,!.$$

Now suppose $\mathbf{a} = (a_1, \cdots, a_i, \cdots, a_{n+1})$ in (11.7) is such that all the a_i's are large enough so that Stirling's approximation can be applied to $a_i\,!$, $i = 1, \cdots, n + 1$. Then from (11.7) we have

$$(11.8)\qquad \log \phi(\mathbf{a}) = \text{const.} - \tfrac{1}{2}\sum_{1}^{n+1} \log a_i\,.$$

Further (except for degenerate alternatives) all the a_i will be large enough for Stirling's approximation, with as large a probability as we may wish, if, fixing the second sample $y_1, \cdots, y_n$, we increase the size m of the first sample sufficiently. The asymptotic normality of the expressions in (11.7) and (11.8), ignoring constants, follows from a theorem due to Wolfowitz [6], under the condition $m = n + 1$. Otherwise the asymptotic normality of (11.8) is obtainable from arguments similar to those in the preceding paragraph.

Acknowledgment. The author acknowledges with pleasure several helpful comments on the present work, by G. A. Barnard, Allan Birnbaum, E. L. Lehmann, J. L. Hodges, A. R. Kamat, S. S. Shrikhande, S. K. Mitra, C. M. Deo, the referee, and a reader. The author also thanks Mr. Modak for doing the necessary typing.

REFERENCES

[1] WASSILY HOEFFDING, "Optimum nonparametric tests," pp. 88-92 of *Proceedings of the Second Berkeley Symposium on Mathematical Statistics and Probability*, University of California Press, Berkeley, 1951.

[2] ERICH L. LEHMANN, "The power of rank tests," *Ann. Math. Stat.*, Vol. 24 (1953), pp. 22–43.

[3] I. RICHARD SAVAGE, "Contribution to the theory of rank order statistics: Two sample case," *Ann. Math. Stat.*, Vol. 27 (1956), pp. 590–615.

[4] D. TEICHROEW, "Empirical power functions for non-parametric two sample tests for small samples," *Ann. Math. Stat.*, Vol. 26 (1955), pp. 340–344.

[5] H. R. VAN DER WAERDEN, "Order tests for the two sample problem," *Koninkl. Nederl. Akademie van Wetenschappen*, Ser. A, Vol. 56, No. 4 (1954), pp. 303–316.

[6] J. WOLFOWITZ, "Additive partition functions and a class of statistical hypotheses," *Ann. Math. Stat.*, Vol. 13 (1942), pp. 247–279.