

THE POWER OF RANK TESTS¹

BY E. L. LEHMANN

Stanford University and University of California, Berkeley

1. Summary. Simple nonparametric classes of alternatives are defined for various nonparametric hypotheses. The power of a number of such tests against these alternatives is obtained and illustrated with some numerical results. Optimum rank tests against certain types of alternatives are derived, and optimum properties of Wilcoxon's one- and two-sample tests and of the rank correlation test for independence are proved.

2. Introduction. The most pressing need in the theory and practice of nonparametric tests at this time seems to be for results concerning the power of such tests, particularly those based on ranks. This would provide a basis for comparing the many different tests proposed as well as for determining the sample sizes necessary to distinguish significant departures from a hypothesis with a reasonable degree of certainty.

The chief problem one is faced with when investigating the power of a nonparametric test is the choice of suitable alternatives. Even in the simplest problems the variety of alternatives is so great that it is clearly impossible to consider all of them. In the past, investigators have concentrated on alternatives postulating normal distributions for the random variables in question. These alternatives, which unfortunately are rather difficult to handle mathematically, must, of course, be studied if one wishes to find out how nonparametric methods compare with procedures based on normal theory. On the other hand, when comparing different rank tests, one is no longer tied to normal alternatives, but it would on the contrary seem rather desirable to make the comparisons in terms of nonparametric classes of alternatives.

As a specific example, consider the one-sided two-sample problem, and suppose that on the basis of samples $X_1, \dots, X_m; Y_1, \dots, Y_n$ from cumulative distribution functions F and G respectively we wish to test the hypothesis $H: F = G$ against the alternatives that $G(x) \leq F(x)$ for all x . If among these alternatives we look for some simple subclasses, parametric theory suggests

$$(2.1) \quad G(x) = F(x - a) \quad \text{for some } a > 0.$$

But under such alternatives, the distribution of the ranks will depend not only on a , but also on F , nor, in general, would a be a suitable measure of the difference

Received 8/7/52.

¹ This work was sponsored in part by the Office of Naval Research. Part of the computations were carried out in connection with a project on the evaluation of rain making undertaken by the Statistical Laboratory, University of California for the State Department of Public Works.

of F and G . The situation is similar to the corresponding one for normal distributions with different means and, say, common but unknown variance.

We shall in the present paper discuss mathematically "natural" nonparametric alternatives against which the distribution of the ranks is constant. Once these have been defined, it is relatively simple, on the basis of a theorem of Hoeffding, to obtain the power of any rank test and also to derive tests possessing various optimum properties.

The classes of alternatives with which we shall be dealing involve arbitrary functions for which one must make a definite choice in order to get specific power-results. This choice is here made solely on the grounds of simplicity for the resulting calculations. We do not, of course, claim that these are the alternatives that actually prevail when the hypothesis is not true. Rather, it seems that where nonparametric methods are appropriate, one usually does not have very precise knowledge of the alternatives. What is then required are alternatives representative of the principal types of deviation from the hypothesis, in terms of which one can study, at least in outline, the ability of various tests to detect such deviations. Such an approach is here presented, and the computations are carried through for a few examples. However, in order to get a valid comparison of such tests as the Wald-Wolfowitz run test and the Smirnov two-sample test, for example, much more systematic computation is required. Such computations seem entirely feasible and would seem to be a worthwhile undertaking.

I should like to express my gratitude to Miss E. L. Scott for her help in setting up and supervising the computations for Table 1 and to Mrs. M. Vasilewskis who carried out these computations, as well as to Mr. H. Wagner and Mr. J. Rosenbaum on whose computations Fig. 2 and 3 are based.

3. The hypothesis of randomness. While we shall be concerned mainly with the two-sample problem, it is convenient to present some preliminary considerations in the more general notation of the hypothesis of randomness. We shall here make the assumption, to hold throughout the paper, that all distribution functions that we consider are to be continuous.

Let f_i ($i = 1, \dots, N$) be continuous, nondecreasing functions defined over the interval $[0, 1]$ such that $f_i(0) = 0, f_i(1) = 1$. Let Z_1, Z_2, \dots, Z_N be independent random variables distributed according to cumulative distribution functions F_1, \dots, F_N . We shall denote by $\mathcal{F}(f_1, \dots, f_N)$ the family of all (F_1, \dots, F_N) such that $F_i = f_i(F)$ where F runs through all continuous cdf's. The classes $\mathcal{F}(f_1, \dots, f_N)$ for different choices of the functions f_1, \dots, f_N then define a partition of the family of all N -tuples (F_1, \dots, F_N) of the kind described. It should perhaps be pointed out that different N -tuples f_i do not necessarily generate different families of F 's. If f_1 is strictly increasing on $[0, 1]$, a natural normalization would be to take $f_1(x) = x, 0 \leq x \leq 1$. If (F_1, \dots, F_N) belongs to the class $\mathcal{F}(f_1, \dots, f_N)$ we shall write

$$(3.1) \quad F_1:F_2:\dots:F_N = f_1:f_2:\dots:f_N.$$

We shall now show that the distribution of the ranks of the Z 's is constant within each family $\mathfrak{F}(f_1, \dots, f_N)$.

LEMMA 3.1. *If F is a continuous cdf and if the cdf of Z is given by $P(\dot{Z} \leq z) = f(F(z))$ where f is nondecreasing on $[0, 1]$ with $f(0) = 0, f(1) = 1$, then the cdf of $F(Z)$ is f .*

PROOF. When $f(u) = u, 0 \leq u \leq 1$ this result is well known and implies in our case that $f(F(Z))$ is uniformly distributed over $[0, 1]$. Therefore

$$P(f(F(Z)) < f(u)) \leq P(F(Z) \leq u) \leq P(f(F(Z)) \leq f(u))$$

and the first and third member equal $f(u)$.

Let us denote the ranks of the N variables Z_1, \dots, Z_N by T_1, \dots, T_N . Then we have

LEMMA 3.2. *If Z_1, \dots, Z_N are independent, the distribution of T_1, \dots, T_N is constant within each family $\mathfrak{F}(f_1, \dots, f_N)$.*

PROOF. Clearly

$$P(F(Z_{i_1}) < \dots < F(Z_{i_N})) \leq P(T_{i_1} = 1, \dots, T_{i_N} = N) \\ \leq P(F(Z_{i_1}) \leq \dots \leq F(Z_{i_N})).$$

But the first and third members of this inequality are independent of F and equal since by Lemma 3.1 the distribution of the $F(\dot{Z}_i)$ is independent of F and continuous.

As an immediate consequence of this lemma we have

THEOREM 3.1. *Given any functions f_1^0, \dots, f_N^0 and any rank test of the hypothesis $H: (F_1, \dots, F_N) \in \mathfrak{F}(f_1^0, \dots, f_N^0)$, the power of this test depends only on $F_1: \dots: F_N$. That is, if $F_1: \dots: F_N = F'_1: \dots: F'_N$ so that (F_1, \dots, F_N) and (F'_1, \dots, F'_N) belong to the same class $\mathfrak{F}(f_1, \dots, f_N)$ the test has the same power against these two alternatives. Furthermore, given any class of alternatives $K: (F_1, \dots, F_N) \in \mathfrak{F}(f'_1, \dots, f'_N)$ there exists a uniformly most powerful rank test for testing H against K .*

PROOF. The first statement is just a specialization of Lemma 3.2. Since the distribution of the ranks is simple both under H and K , the second statement as well as a method of constructing the most powerful rank test follow from the Neyman-Pearson fundamental Lemma.

In order to apply this theorem we require the distribution of (T_1, \dots, T_N) for the (f_1, \dots, f_N) of our choice. The relevant result was obtained by Hoeffding ([1], p. 88). Instead of stating it here we shall in the next section give its specialization for the two-sample problem.

4. The two-sample problem. Let X_1, \dots, X_m and Y_1, \dots, Y_n be independently distributed with cdf's F and G respectively. We wish to test the hypothesis $H: F = G$. The classes $\mathfrak{F}(f_1, \dots, f_N)$, in the present case, involve only two functions f and g and may be written as $\mathfrak{F}(f, g)$. To simplify our notation, we shall assume that f is strictly increasing. Then $\mathfrak{F}(f, g)$ may be represented by a single function g and is given by $\mathfrak{F}(g) = \{(F, g(F))\}$ where the domain of F is as before

the totality of continuous cdf's, and where g is a continuous nondecreasing function with $g(0) = 0, g(1) = 1$.

Let us denote the ordered X 's and Y 's by $X^{(1)} < X^{(2)} < \dots < X^{(m)}$ and $Y^{(1)} < \dots < Y^{(n)}$ and the ranks of the X 's and Y 's in the combined sample by $R_1 < \dots < R_m$ and $S_1 < \dots < S_n$ respectively. The complete set of the ranks is, of course, determined by the ranks of the Y 's alone. We shall assume here that the function g is differentiable on $[0, 1]$ with derivative g' . Specializing

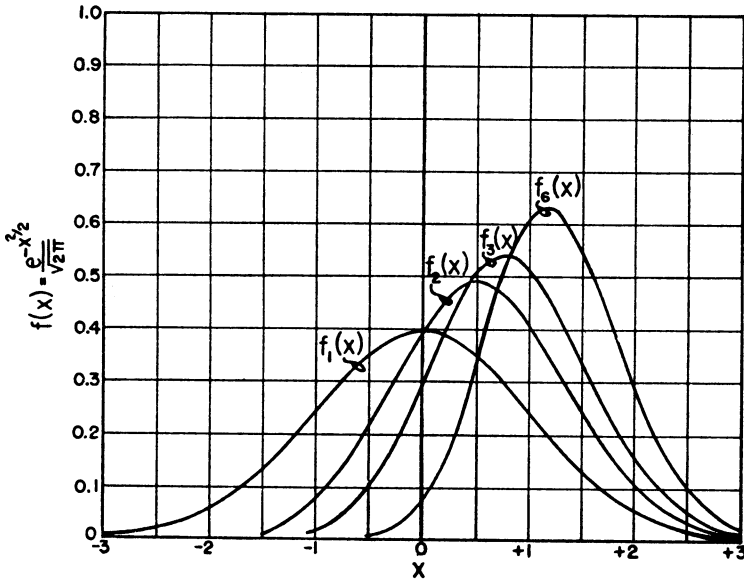


FIG. 1-A

a theorem of Hoeffding ([1], p. 88) to the present case we find that when F and $G = g(F)$ are the cdf's of the X 's and Y 's, then

$$P(S_1 = s_1, \dots, S_n = s_n) = \frac{1}{\binom{m+n}{m}} E_F [g'(F(Y^{(s_1)})) \cdot \dots \cdot g'(F(Y^{(s_n)}))]$$

where the expectation is computed under the assumption that F is the true distribution of both the X 's and Y 's. Since in this case $F(Y)$ is uniformly distributed over $[0, 1]$, we get

$$(4.1) \quad P(S_1 = s_1, \dots, S_n = s_n) = \frac{1}{\binom{m+n}{m}} E [g'(U^{(s_1)}) \cdot \dots \cdot g'(U^{(s_n)})]$$

where $U^{(s_1)}, \dots, U^{(s_n)}$ are the s_1^{st} to s_n^{th} order statistics in a sample of $m+n$ variables distributed uniformly over $[0, 1]$.

Since the probability distribution of the ranks can be expressed so simply in

terms of g it is seen that the difficulty in obtaining power results for a specific alternative is directly related to the complexity of the function g involved. This explains why the investigation for normal alternatives has proved so difficult. When F and G are two distinct normal cdf's, the function $g = G(F^{-1})$ is not particularly easy to handle.

Consider now the one-sided alternatives $G(x) \leq F(x)$. To this corresponds a function g such that $g(x) \leq x$, $0 \leq x \leq 1$. The simplest choice in view of (4.1) seems to be $g(x) = x^k$; $k > 1$. The associated problem is that of testing $H:G = F$ against the alternatives $K:G = F^k$. In addition to mathematical simplicity, this choice has the advantage of admitting a simple interpretation of the alternatives. Suppose that k is an integer. Then F^k is the distribution of the

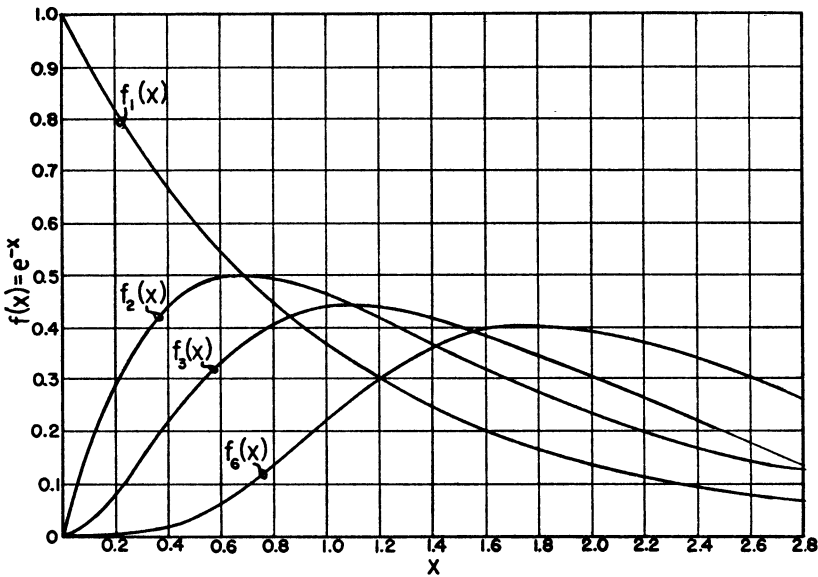


FIG. 1-B

maximum of k independent variables having distribution F . Thus under the alternative, the X 's have distribution F while the distribution of the Y 's is the same as that of the maximum of k X 's.

In order to give an idea of how much larger the Y 's are than the X 's, note that if $G = F^k$, $P(X < Y) = \int F dG = k/k + 1$. In Fig. 1, we have assumed that the distribution of X is given by the densities

$$f_1(x) = \frac{1}{\sqrt{2\pi}} e^{-x^2/2}; \quad f_1(x) = e^{-x}, \quad 0 \leq x; \quad f_1(x) = 1, \quad 0 \leq x \leq 1$$

respectively and show the density of f_k of Y when $G = F^k$ for $k = 2, 3$ and 6 . In terms of the present frame of reference the distance of the density f_k from f_1 is

the same in all three cases, since in each of them f_k is the density of the maximum of k observations from f_1 and since further in all three cases, every rank test has the same probability of detecting the hypothesis to be false when f_1 is the density of the X 's and f_k that of the Y 's.

It is clear that a similar interpretation of the alternatives F^k can be given when k instead of being an integer is any rational number. Altogether, we may think of the class of alternatives $G = F^k$, as a one parameter family of nonparametric classes of alternatives. The distribution of the ranks under these alternatives is now easily determined from (4.1).

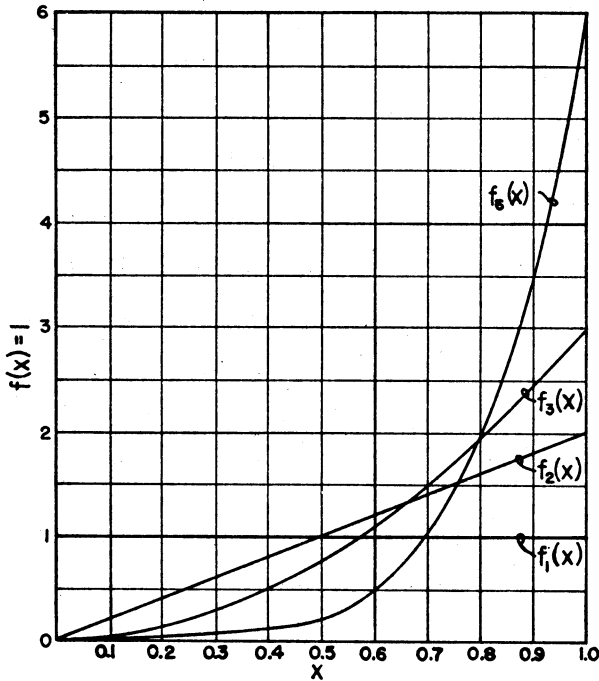


FIG. 1-C

For if we put $N = m + n$, $s_0 = 0$, $s_{n+1} = N + 1$, $u_0 = 0$, and $u_{n+1} = 1$, the joint density $p(u_1, \dots, u_n)$ of $U^{(s_1)}, \dots, U^{(s_n)}$ is given by

$$(4.2) \quad \frac{N!}{\prod_{i=0}^n (s_{i+1} - s_i - 1)!} \prod_{j=0}^n (u_{j+1} - u_j)^{s_{j+1} - s_j - 1}$$

over the region $0 = u_0 \leq u_1 \leq \dots \leq u_{n+1} = 1$. If here we make a transformation to new variables V_1, \dots, V_n defined by

$$(4.3) \quad u_i = v_i v_{i+1} \cdots v_n \quad (i = 1, \dots, n)$$

and put $v_0 = 0, v_{n+1} = 1$, it is seen that the joint density of the V 's is

$$(4.4) \quad \frac{N!}{\prod_{i=0}^n (s_{i+1} - s_i - 1)!} \prod_{j=1}^n v_j^{s_j-1} (1 - v_j)^{s_{j+1}-s_j-1}$$

over the region $0 \leq v_j \leq 1, j = 1, \dots, n$ so that the V 's are independently distributed according to Beta-distributions, that is, as are single order statistics from a uniform distribution.

TABLE 1

Test	$m = n = 4$		$m = n = 6$	
	$\beta(F^2)$	$\beta(F^3)$	$\beta(F^2)$	$\beta(F^3)$
T_1	.23	.33	.29	.45
T_2	.31	.47	.38	.59
T_3	.32	.49	.41	.64
T_4	.14	.20	.17	.29
T_5	.15	.22	.21	.36
T_6	.19	.32	.25	.44

Since $U^{(s_1)} \cdot \dots \cdot U^{(s_n)} = V_1 \cdot V_2^2 \cdot \dots \cdot V_n^n$, we have when $G = F^k$ and hence $g'(u) = ku^{k-1}$,

$$(4.5) \quad \begin{aligned} P(S_1 = s_1, \dots, S_n = s_n) &= \frac{k^n}{\binom{m+n}{m}} E[(U^{(s_1)} \cdot \dots \cdot U^{(s_n)})^{k-1}] \\ &= \frac{k^n}{\binom{m+n}{m}} \prod_{j=1}^n E(V_j^{jk-j}) = \frac{k^n}{\binom{m+n}{m}} \prod_{j=1}^n \frac{\Gamma(s_j + jk - j)}{\Gamma(s_j)} \\ &\quad \cdot \frac{\Gamma(s_{j+1})}{\Gamma(s_{j+1} + jk - j)}. \end{aligned}$$

In particular, when $k = 2$ so that $G = F^2$,

$$(4.6) \quad \begin{aligned} P(S_1 = s_1, \dots, S_n = s_n) &= \frac{2^n}{\binom{m+n}{m}} \\ &\quad \cdot \frac{s_1(s_2 + 1) \cdot \dots \cdot (s_n + n - 1)}{(m+n+1)(m+n+2) \cdot \dots \cdot (m+2n)}. \end{aligned}$$

Using (4.6) (or more generally (4.5) or (4.1)) one can now compute the power of various rank tests against the alternatives in question. One must list the sets (s_1, \dots, s_n) making up the critical region and then sum the right-hand side of (4.6) over these values of the ranks. In this manner Table 1 was com-

puted, which gives the power of six different rank tests $T_1 - T_6$ against the alternatives $G = F^2$ and $G = F^3$ at level of significance $\alpha = .1$. Since the computation of the exact power rapidly increases in difficulty with the sample size, these computations have been carried through only for the cases $m = n = 4$ and $m = n = 6$.

The above tests are defined as follows.

T_1 : One-sided median test. Rejects H when too many Y 's exceed the median of the combined sample. See Mood ([2], p. 394) and Westenberg [18].

T_2 : One-sided Wilcoxon test [3], [4]. Rejects when $S_1 + \dots + S_n$ is too large, or equivalently when there are too many pairs X_i, Y_j ($i = 1, \dots, m$; $j = 1, \dots, n$) with $X_i < Y_j$.

T_3 : This is the most powerful rank test for testing $G = F$ against $G = F^2$ (see Section 6). It rejects when $S_1(S_2 + 1) \cdot \dots \cdot (S_n + n - 1)$ is too large.

T_4 : Wald-Wolfowitz run test [5]. Rejects when the total number of runs of X 's and Y 's is too small.

T_5 : Two-sided median test ([2], p. 394). Rejects when either too many X 's or Y 's exceed the median of the combined sample.

T_6 : Two-sided Wilcoxon test [3], [4]. Rejects when

$$\left| S_1 + \dots + S_n - \binom{m+n+1}{2} \right|$$

is too large.

Although it is not shown in Table 1, we mention for later reference also

T_7 : Smirnov two-sample test [6]. Rejects when supt_t

$$| F_{x_1, \dots, x_m}(t) - G_{y_1, \dots, y_n}(t) |$$

is too large where $F_{x_1, \dots, x_m}(t)$ and $G_{y_1, \dots, y_n}(t)$ are the sample cumulative distribution functions of the X 's and Y 's respectively.

Of course, not all of these tests are directly comparable. While the first three are aimed only at alternatives under which the Y 's tend to be larger than the X 's, the fifth and sixth test are designed against two-sided alternatives, and the fourth and seventh against arbitrary deviations from the hypothesis.

5. Large sample power. For alternatives of the type $G = F^k$ it also becomes a relatively easy task to compute the approximate power of certain rank tests using large sample theory. Some results obtained in this way are shown in Fig. 2 and 3.

Fig. 2 gives the power of various tests against the alternative $g(F) = F^2$ for different sample sizes n . The lowest of the four curves (labeled β_4) corresponds to the run test (the subscripts refer to the numbering of the tests in the previous section) and is based on theory not yet completely verified. The next curve,

β_7 , gives a lower bound to the power of the Smirnov test, while the two upper curves show the power of the two-sided and one-sided Wilcoxon test. In Fig. 3 are shown the corresponding curves for $g(F) = F^3$, except that the run test has been omitted.

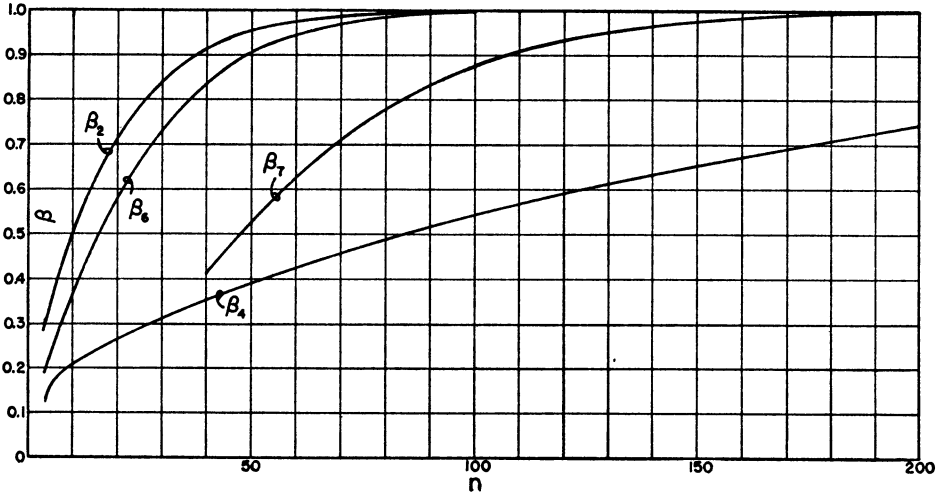


FIG. 2. $\alpha = 0.1 \quad g(F) = F^2$

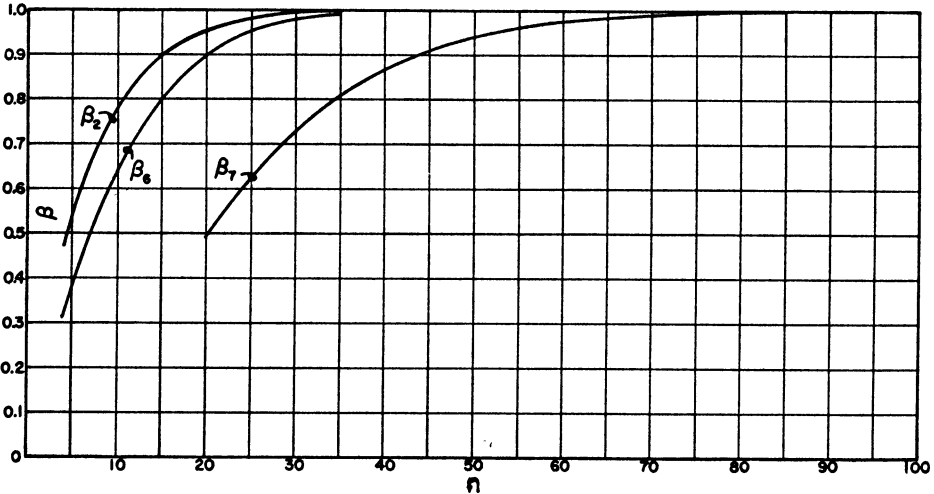


FIG. 3. $\alpha = 0.1 \quad g(F) = F^3$

The remainder of this section is devoted to a discussion of the formulae from which Fig. 2 and 3 were computed.

For the Wilcoxon statistic U which counts the number of pairs X_i, Y_j with $X_i < Y_j$, it was proved by Mann and Whitney [4] that for large samples

$$\frac{\frac{U}{mn} - E\left(\frac{U}{mn}\right)}{\sigma\left(\frac{U}{mn}\right)}$$

is approximately normally distributed when $F = G$, and the proof of the corresponding fact when $F \neq G$ was given in [7]. Mann and Whitney also gave the first two moments of U as

$$(5.1) \quad \begin{aligned} E\left(\frac{U}{mn}\right) &= \int F dG \\ mn\sigma^2\left(\frac{U}{mn}\right) &= \left[\frac{m+n+1}{12} + (m-1)(\lambda - \epsilon_1) + (n-1)(\lambda - \epsilon_2) \right. \\ &\quad \left. - \lambda^2(m+n-1) \right] \end{aligned}$$

where

$$\begin{aligned} \lambda &= \frac{1}{2} - \int F dG \\ \epsilon_1 &= \frac{1}{3} - \int F^2 dG, \quad \epsilon_2 = \frac{1}{3} - \int (1-G)^2 dF. \end{aligned}$$

(Note that the notation used here differs from that in [4].)

If $G = F^k$ we have

$$\int F dG = \frac{k}{k+1}, \quad \int F^2 dG = \frac{k}{k+2}, \quad \int G^2 dF = \frac{1}{2k+1},$$

and hence on substituting in (5.1)

$$(5.4) \quad \begin{aligned} E\left(\frac{U}{mn}\right) &= \frac{k}{k+1} \\ mn\sigma^2\left(\frac{U}{mn}\right) &= \frac{k}{(k+1)^2} \left[\frac{m-1}{k+2} + \frac{k(n-1)}{2k+1} + 1 \right]. \end{aligned}$$

The theory of the run test was developed by Wald and Wolfowitz [5] and certain extensions were given in [8]. If W denotes the total number of runs of X 's and Y 's and if $m/n = \gamma$ it was shown in [5] that when $F = G$, $(W/m - E(W/m))/\sigma(W/m)$ is asymptotically normally distributed. It was also proved that when $G = g(F)$ where the derivative g' of g is continuous and positive on $0 < x < 1$, then

$$(5.3) \quad E\left(\frac{W}{m}\right) \rightarrow 2 \int_0^1 \frac{g'(x)}{\gamma + g'(x)} dx \quad \text{as } m \rightarrow \infty.$$

In [8] Wolfowitz stated that the distribution of W is approximately normal even when $g(x) \equiv x$, and he derived the asymptotic formula

$$(5.4) \quad \frac{1}{4}\sigma^2 \left(\frac{W}{\sqrt{m}} \right) \rightarrow \int_0^1 \frac{\gamma g'^2}{(\gamma + g')^3} dx + \int_0^1 \frac{g'(\gamma^3 + g'^3)}{(\gamma + g')^4} dx - \left[\int_0^1 \frac{g'^2}{(\gamma + g')^2} dx \right]^2 - \gamma^3 \left[\int_0^1 \frac{g'}{(\gamma + g')^2} dx \right]^2.$$

When $G = F^k$, then $g(x) = x^k$ and $g'(x) = kx^{k-1}$, and the integrals on the righthand side of (5.3) and (5.4) can be evaluated without much difficulty. The power of the run test against $g(F) = F^2$ shown in Fig. 2 was computed in this manner. Since then it has been pointed out to me by R. Savage that when the limit result for

$$\left(\frac{W}{m} - E \left(\frac{W}{m} \right) \right) / \sigma \left(\frac{W}{m} \right)$$

we replace

$$E(W/m)$$

by

$$2 \int_0^1 g'(x)/(\gamma + g'(x)) dx,$$

the error is of the order

$$\sqrt{m} \left[E(W/m) - 2 \int_0^1 g'(x)/(\gamma + g'(x)) dx \right],$$

as is seen from (5.4). Thus (5.3) is not enough to guarantee the validity of this substitution. However, the numerical results obtained seemed sufficiently interesting to leave them in, in the hope that a proof of their validity will soon be forthcoming.

The large sample distribution of the Smirnov statistic has not yet been investigated when $F \neq G$. However, it was pointed out by Massey [9] that a lower bound to the power can be obtained simply by the inequality

$$(5.5) \quad P(\sup_t | F_{x_1, \dots, x_m}(t) - G_{y_1, \dots, y_n}(t) | \geq C) \geq P(| F_{x_1, \dots, x_m}(t_0) - G_{y_1, \dots, y_n}(t_0) | \geq C)$$

where t_0 is any particular value of t . If $F(x) = x$ ($0 \leq x \leq 1$), $G(x) = F^k(x) = x^k$ and we take for t_0 the point of maximum difference between F and G , we get $t_0 = 1/k - 1/\sqrt{k}$. Now $F_{x_1, \dots, x_m}(t_0)$ and $G_{y_1, \dots, y_n}(t_0)$ are the proportion of successes in m and n binomial trials with probability of success equal to $F(t_0) = t_0$ and $G(t_0) = t_0^k$ respectively. Thus for moderately large sample sizes

$$F_{x_1, \dots, x_m}(t_0) - F_{y_1, \dots, y_n}(t_0)$$

is approximately normally distributed with mean $t_0 - t_0^k$ and variance

$$(t_0(1 - t_0)/m) + (t_0^k(1 - t_0^k)/n).$$

The constant C of (5.5) can be obtained from [10] or [11].

6. Optimum rank tests for the two-sample problem. We next consider the problem of determining the optimum rank test of $H:G = F$ against $K:G = g(F)$.

Under the hypothesis all $\binom{m+n}{m}$ possible combinations of s_1, \dots, s_n are equally likely while their probabilities under K are given by (4.1). Thus the problem, in terms of ranks, reduces to that of testing a simple hypothesis against a simple alternative, and its solution is given by the fundamental lemma of Neyman and Pearson. The most powerful test rejects when the ratios of the probabilities is too large. Since the denominator of this ratio is constant (independent of s_1, \dots, s_n), this is equivalent to rejecting when (4.1) is too large.

If we take, for example, $g(F) = F^2$ the most powerful rank test rejects when

$$s_1(s_2 + 1) \cdots (s_n + n - 1) > C.$$

The power of this test is shown against $G = F^2$ and $G = F^3$ in Table 1.

Since usually one does not have any precise alternatives in mind, it is perhaps more interesting to turn the problem around and to investigate what optimum properties (if any) are possessed by some of the standard rank tests. This gives an indication of the type of deviation from the hypothesis that the test under consideration is particularly suited to detect and therefore of the circumstances in which the application of the test is appropriate. As one such example, we shall discuss here the Wilcoxon test.

Consider to this end the one-parameter family of nonparametric alternatives given by

$$(6.1) \quad g_p(F) = qF + pF^2, \quad 0 \leq p \leq 1, \quad p + q = 1.$$

If $\beta(p)$ denotes the power of a test against $g_p(F)$ we shall show that the one-sided Wilcoxon test among all rank test maximizes $\beta'(0)$, the slope of the power function at the hypothetical point. It is thus "locally most powerful" just against the type of alternative we have been considering.

To prove this result we must consider $\beta(p)$. Since $d/du (g_p(u)) = q + 2pu$ it follows from (4.1) that

$$P(S_1 = s_1, \dots, S_n = s_n | p) = \frac{1}{\binom{m+n}{m}} E[(q + 2pU^{(s_1)}) \cdots (q + 2pU^{(s_n)})]$$

and hence that

$$\begin{aligned} \left. \frac{d}{dp} P(S_1 = s_1, \dots, S_n = s_n \mid p) \right|_{p=0} &= \frac{1}{\binom{m+n}{m}} E \left[\sum_{i=1}^n (2U^{(s_i)} - 1) \right] \\ &= \frac{1}{\binom{m+n}{m}} \left[\frac{2}{m+n+1} \sum_{i=1}^n s_i - n \right]. \end{aligned}$$

Therefore

$$(6.2) \quad \beta'(0) = \sum \frac{1}{\binom{m+n}{m}} \left[\frac{2}{m+n+1} \sum_{i=1}^n s_i - n \right],$$

where the summation extends over the sets (s_1, \dots, s_n) that form the critical region. It follows as before from the fundamental lemma of Neyman and Pearson that we maximize $\beta'(0)$ at a fixed level of significance by rejecting H when the right-hand side of (6.2), and hence $\sum_{i=1}^n s_i$, is too large. This is the desired result.

The above property of the Wilcoxon two-sample test can be generalized in various directions, which we shall sketch only briefly. First, an analogous property holds for any test whose region of rejection is of the form

$$(6.3) \quad h(s_1) + h(s_2) + \dots + h(s_n) \geq C.$$

In any such case one can find a function h^* for which the test given by (6.3) maximizes the slope of the power function $\beta(p)$ against the alternatives $g_p(F) = qF + ph^*(F)$ at $p = 0$. A particular example of (6.3) is Mood's median test T_1 , which rejects when the number of s_i exceeding a given constant is too large. However, in most cases, and this seems to include the one under consideration, the function h^* is too complicated to be very enlightening and to warrant the tedious computations necessary to obtain it. The existence of h^* follows from the fact that the s_i can only take on a finite number of values so that without loss of generality, the function h in (6.3) may be taken to be a polynomial. Furthermore, as in the case of the Wilcoxon statistic the test maximizing $\beta'(0)$ against the alternatives $qF + ph^*(F)$ is given by the rejection region

$$E[h^*(U^{(s_1)}) + \dots + h^*(U^{(s_n)})] > C.$$

To complete the proof it is enough to show that there exists a polynomial h^* and constants $a > 0$, b such that $h^*(0) = 0$, $h^*(1) = 1$, $h^*(u) \geq 0$ for $0 \leq u \leq 1$ and $E[h^*(U^{(s)})] = a[h(s) + b]$. Now from the fact (see (4.5)) that

$$E[(U^{(s)})^k] = \frac{s(s+1) \cdots (s+k-1)}{(m+n+1) \cdots (m+n+k)}$$

it is seen that there exists a polynomial P for which $E[P'(U^{(s)})] = h(s)$. Putting $h^*(s) = a[P(s) + bs] + c$ we need to show only that given any polynomial P

there exist constants $a > 0$, b , c such that $aP(0) + c = 0$, $a[P(1) + b] + c = 1$ and $a[P'(s) + b] \geq 0$ for $0 \leq s \leq 1$, and this is easily verified.

Another extension of the above result concerns a problem different from but closely related to the two-sample problem. (In this connection see Hemelrijk [19]). Let Z_1, \dots, Z_N be identically and independently distributed with cdf M . The hypothesis to be tested is that M is symmetric about the origin, that is, that for all z , $M(z) + M(-z) = 1$. If we assume M to be continuous, put $1 - M(0) = \rho$ and denote by F and G the conditional distributions of Z given that $Z > 0$ and of $-Z$ given that $Z < 0$, the hypothesis is equivalent to the two statements $\rho = \frac{1}{2}$, $F = G$. Let m and n be the number of positive and negative Z 's respectively, and denote by X_1, \dots, X_m and Y_1, \dots, Y_n the positive Z 's and the absolute values of the negative Z 's respectively, in their original order of subscripts. Consider now the probability of any particular set of ranks of the Y 's under some alternative. Given m and n , this is independent of ρ and is given by (4.1) when $G = g(F)$. In addition, n is a binomial variable with probability ρ of success. Thus we get

$$(6.4) \quad \begin{aligned} P(\text{The number of } Z\text{'s} > 0 \text{ is } n \text{ and } S_1 = s_1, \dots, S_n = s_n) \\ = \rho^n (1 - \rho)^{N-n} E[g'(U^{(s_1)}) \cdot \dots \cdot g'(U^{(s_n)})]. \end{aligned}$$

If in particular, one considers alternatives with $\rho = \frac{1}{2}$, the right-hand side of (6.4) becomes $2^{-N} E[g'(U^{(s_1)}) \cdot \dots \cdot g'(U^{(s_n)})]$, which formally differs from (4.1) only by a multiplicative constant. Thus any optimum test of the two-sample problem derived on the basis of (4.1) gives rise to a dual one for the hypothesis of symmetry. As an example, the Wilcoxon two-sample test which rejects when $s_1 + \dots + s_n > C$, under translation in the above manner becomes a test of the hypothesis of symmetry also proposed by Wilcoxon [3] and recently shown by Tukey [12] to be equivalent to a test proposed independently by Walsh [13]. This test is now seen to maximize $\beta'(0)$ against the alternatives according to which $M(0) = \frac{1}{2}$, the conditional distribution of Z given $Z > 0$ is F and that of $-Z$ given $Z > 0$ is $qF + pF^2$.

As another application of this approach, let us once more consider the two-sample problem, but this time with a two-sided class of alternatives. For simplicity, we take $m = n$, and we assume that either the X 's are distributed according to F and the Y 's according to $qF + pF^2$ or vice versa. Let $\beta(p)$ denote the power of a rank test against the first of these alternatives and $\beta^*(p)$ that against the second. We shall then maximize the average power $\frac{1}{2}[\beta(p) + \beta^*(p)]$ at $p = 0$. Since it turns out that $\beta'(0) + \beta^{*'}(0) = 0$ this is equivalent to maximizing $\beta''(0) + \beta^{*''}(0)$.

From (4.1) we see that the sum of the probabilities of $R_1 = r_1, \dots, R_n = r_n$, $S_1 = s_1, \dots, S_n = s_n$ under the two alternatives is

$$\begin{aligned} E[(q + 2pV^{(s_1)}) \cdot \dots \cdot (q + 2pV^{(s_n)})] + E[(q + 2pU^{(r_1)}) \cdot \dots \cdot (q + 2pU^{(r_n)})] \\ = 2 + pE \left[\sum_{i=1}^n (2U^{(s_i)} - 1) + \sum_{i=1}^n (2U^{(r_i)} - 1) \right] \\ + p^2 E \left[\sum_{i < j} (2V^{(s_i)} - 1)(2V^{(s_j)} - 1) + 2 \sum_{i < j} (U^{(r_i)} - 1)(U^{(r_j)} - 1) \right] + o(p^2). \end{aligned}$$

Now

$$\sum_{i=1}^n [E(U^{(s_i)}) + E(V^{(r_i)})] = \frac{\sum r_i + \sum s_i}{2n + 1} = n$$

so that the coefficient of p is zero. The coefficient of p^2 is, except for a constant, given by

$$\begin{aligned} 4 \sum_{i < j} [E(V^{(s_i)} V^{(s_j)}) + E(U^{(r_i)} U^{(r_j)})] - 2 \sum_{i < j} E[V^{(s_i)} + V^{(s_j)} + U^{(r_i)} + U^{(r_j)}] \\ = 4 \sum_{i < j} \frac{s_i(s_j + 1) + r_i(r_j + 1)}{(2n + 1)(2n + 2)} - 4 \sum_{i \neq j} \frac{s_i + s_j + r_i + r_j}{(2n + 1)}. \end{aligned}$$

Using the fact that $\sum_{i < j} s_i s_j = \sum (n - i) s_i$ and that $\sum (r_i + s_i)$ as well as $\sum [r_i^2 + s_i^2]$ are constants, the coefficient of p^2 is, except for a constant,

$$(6.5) \quad \frac{2}{(2n + 1)(2n + 2)} [(\sum s_i)^2 + (\sum r_i)^2 + \sum (s_i - i)^2 + \sum (r_i - i)^2].$$

Thus we maximize the average power at $p = 0$ by maximizing (6.5) or equivalently

$$(6.6) \quad \left[\sum s_i - \frac{n(2n + 1)}{2} \right]^2 + \left[\sum r_i - \frac{n(2n + 1)}{2} \right]^2 + \sum (s_i - i)^2 + \sum (r_i - i)^2.$$

Rather surprisingly this is not the two-sided Wilcoxon test which is given just by the first two terms of (6.6).

This result can be given a slightly different form. Let us write the alternative $qF + pF^2$ in the form

$$(6.7) \quad g_\theta(F) = \frac{1}{1 + \theta} (F + \theta F^2)$$

where we shall be interested only in values of θ close to zero, and where we may then consider also negative values of θ . If $\beta(\theta)$ denotes the power of some rank test against θ we may in analogy to the type A tests of Neyman and Pearson [14] maximize $\beta''(0)$ subject to $\beta(0) = \alpha$, $\beta'(0) = 0$. This will clearly again lead to (6.6).

Such a parametric approach can be carried further. Consider, for example, samples from k populations F_1, \dots, F_k and the hypothesis $H: F_1 = \dots = F_k$. If we then consider alternatives of the form $F_i = (1/1 + \theta_i)(F + \theta_i F^2)$ with $\sum \theta_i = 0$, we can, for example, maximize the average power over the sphere $\sum \theta_i^2 = \delta^2$ for small δ . This is analogous to a formulation given by Wald [15] for the normal case, and leads to an extension of (6.6).

7. The hypothesis of independence. To illustrate the general approach of this paper with another example, consider a sample $(X_1, Y_1), \dots, (X_n, Y_n)$ from a bivariate distribution. The hypothesis to be tested is that X and Y are in-

dependent. Nonparametric alternatives to this will be defined by means of a function h of two variables such that h is a continuous cdf over the unit square $0 \leq x, y \leq 1$ (so that $h(0, 0) = 0, h(1, 1) = 1$). A nonparametric class $\mathcal{H}(h)$ of bivariate distributions is then formed by the totality of distributions $h(F, G)$ where F, G are arbitrary continuous univariate cdf's. Suppose now that the X 's are ordered and that in this ordering the rank of X_i is R_i ($i = 1, \dots, n$). Similarly, we shall denote by S_i the rank of Y_i among the Y 's. We then have, analogously to the corresponding result in Section 3,

THEOREM 7.1. *For the distributions of a class $\mathcal{H}(h)$ the distribution of the R 's and S 's is constant, that is, independent of F and G .*

This follows from

LEMMA 7.1. *If h, F and G are continuous and if $P(X \leq x, Y \leq y) = h(F(x), G(y))$ then*

$$(7.1) \quad P(F(X) \leq u, G(Y) \leq v) = h(u, v).$$

PROOF. Let x, y be such that $F(x) = u, G(y) = v$. Then

$$P(X < x, Y < y) \leq P(F(X) \leq u, G(Y) \leq v) \leq P(X \leq x, Y \leq y).$$

But

$$P(X < x, Y < y) = P(X \leq x, Y \leq y) = h(F(x), G(y)) = h(u, v).$$

Again we can write down the distribution of the R 's and S 's using Hoeffding's theorem. In fact if $h'(u, v) = (\partial^2/\partial u \partial v)h(u, v)$ we have

$$(7.2) \quad P(R_1 = r_1, \dots, R_n = r_n; S_1 = s_1, \dots, S_n = s_n) \\ = E[h'(U^{(r_1)}, V^{(s_1)}) \dots h'(U^{(r_n)}, V^{(s_n)})]$$

where $U_1, \dots, U_n; V_1, \dots, V_n$ are two independent samples from a uniform distribution on $[0, 1]$ and $U^{(r_i)}, V^{(s_i)}$ are the associated order statistics. It should be noted that in (7.2) it is not assumed that either the r_i or the s_i are arranged in natural order. Alternatively we may take $r_i = i$ and define s_i as the rank of the Y that is associated with the i th smallest X . In this notation only the S_i remain random and instead of (7.2) we get

$$(7.3) \quad P(S_1 = s_1, \dots, S_n = s_n) = \frac{1}{n!} E[h'(U^{(1)}, V^{(s_1)}) \dots h'(U^{(n)}, V^{(s_n)})].$$

Perhaps the simplest choice for h would seem to be

$$(7.4) \quad h(u, v) = u + v; \quad h(u, v) = \frac{1}{2}(u^2 + v^2).$$

This corresponds to the family of cdf's $H(x, y) = \frac{1}{2}[F(x)G^2(y) + F^2(x)G(y)]$ and can be interpreted similarly to the alternatives discussed in previous sections. The observations X, Y are drawn with probability $\frac{1}{2}$ each from two bivariate populations with independent components. According to one of these X is an observation from F while Y is the larger of two observations from G ; according

to the other the situation is just reversed. A general family of alternatives could be obtained in this way, given by

$$H(x, y) = \sum_{i=1}^l p_i F^{a_i}(x) G^{b_i}(y), \quad p_i \geq 0, \quad \sum p_i = 1.$$

The distribution of the ranks under such alternatives can be written down on the basis of (4.5) and (7.3). However, even in the simplest cases such as (7.4) the resulting expressions are quite complex except for very small values of n .

It seems of interest that one of the best known tests for independence, that based on the rank correlation coefficient, possesses an optimum property similar to the ones derived in Section 6 for the Wilcoxon and related tests. For let

$$(7.4) \quad h_p(u, v) = quw + pu^2v^2.$$

Using (7.2) we see that the probability of $R_i = r_i, S_j = s_j$ ($i, j = 1, \dots, n$) is

$$E[(q + 4pU^{(r_1)}V^{(s_1)}) \cdot \dots \cdot (q + 4pU^{(r_n)}V^{(s_n)})].$$

Differentiating this with respect to p and setting $p = 0$ we get

$$E \left[\sum_{i=1}^n (-1 + 4U^{(r_i)}V^{(s_i)}) \right] = -n + \frac{4}{(n+1)^2} \sum_{i=1}^n r_i s_i.$$

Thus the test that maximizes the slope of the power function against the alternatives (7.4) at $p = 0$ rejects when $\sum r_i s_i$ is too large, and hence when the rank correlation coefficient is too large. More generally, if $h_p(u, v) = quw + pg_1(u)g_2(v)$, the test that maximizes the slope of the power function rejects when $g_1(U^{(r_i)})g_2(V^{(s_i)})$ is too large. In this manner we obtain a generalization of the result connected with (6.3).

8. Invariance. The definitions of nonparametric classes of alternatives given in the earlier sections for various problems may appear somewhat arbitrary. The only apparent justification is their success in permitting the derivation of results concerning the power of nonparametric tests. Also, it is not clear at this point to what problems, in addition to those discussed here, the method is applicable and what would be the appropriate classes of alternatives for such additional problems. We shall now show that the notion of invariance provides an approach to a general class of problems, which in the special cases treated in Sections 4 and 7, yields the definitions given there.

The general concept of invariance due to Hunt and Stein, and presented in [16] is not sufficiently broad for our purpose, so that we shall first indicate a slight extension of this notion. As usual, we are concerned with a sample space \mathfrak{X} on which an additive class \mathfrak{A} has been defined as the class of measurable sets. Let \mathfrak{F} be a class of possible distributions F of a random variable X over \mathfrak{A} . We are also given measurable transformations τ of \mathfrak{X} onto itself (see [17] Ch. VIII). These transformations are such that when X is distributed according to $F \in \mathfrak{F}$,

the distribution $\bar{\tau}F$ of τX is again in \mathfrak{F} . The definition of $\bar{\tau}F$ may be expressed by the transformation equation

$$(8.1) \quad P_F(\tau X \in A) = P_{\bar{\tau}F}(X \in A) \quad \text{for all } A \in \mathcal{G}.$$

Since we are not assuming that the transformations τ or $\bar{\tau}$ are 1:1, they do not necessarily have inverses and cannot therefore be assumed to form a group. Instead, we shall assume that the classes \mathcal{S} and $\bar{\mathcal{S}}$ of transformations τ and $\bar{\tau}$ include the identity transformation and are closed under multiplication so that $\tau_1, \tau_2 \in \mathcal{S} \rightarrow \tau_1\tau_2 \in \mathcal{S}$; that is, \mathcal{S} and $\bar{\mathcal{S}}$ are semi-groups with an identity.

We shall say that a function φ defined on \mathfrak{X} is invariant with respect to \mathcal{S} if

$$(8.2) \quad \varphi(\tau x) = \varphi(x) \quad \text{for all } \tau \in \mathcal{S}, x \in \mathfrak{X}.$$

Furthermore, we shall define as *maximal invariant*, the partition of \mathfrak{X} generated by the relation R :

$$(8.3) \quad x_1 R x_2$$

when there exists $\tau \in \mathcal{S}$ such that $x_2 = \tau x_1$, that is, the partition generated by the smallest equivalence relation \sim closed with respect to (8.3). We note that the existence of τ_1, τ_2 such that

$$(8.4) \quad \tau_1 x_1 = \tau_2 x_2$$

implies $x_1 \sim x_2$. For then $x_1 R \tau_1 x_1, x_2 R \tau_2 x_2$ and hence $x_1 \sim \tau_1 x_1 = \tau_2 x_2 \sim x_2$.

LEMMA 8.1. *A function φ is invariant if and only if*

$$(8.5) \quad x_1 \sim x_2 \text{ implies } \varphi(x_1) = \varphi(x_2).$$

PROOF. Suppose that (8.5) holds and that $x' = \tau x$. Then $x' \sim x$ and hence $\varphi(x') = \varphi(x)$ so that φ is invariant. Conversely, suppose that φ is invariant and consider the equivalence relation according to which two points x_1, x_2 are equivalent if and only if $\varphi(x_1) = \varphi(x_2)$. This equivalence relation is closed with respect to R and hence contains the relation \sim so that $\varphi(x_1) = \varphi(x_2)$ implies $x_1 \sim x_2$.

THEOREM 8.1. *If φ is invariant with respect to \mathcal{S} then the distribution of $\varphi(X)$ is constant over the equivalence classes mod $\bar{\mathcal{S}}$ that is, if $F \sim F'$ mod $\bar{\mathcal{S}}$ then*

$$P_F(\varphi(X) \in A) = P_{F'}(\varphi(X) \in A).$$

PROOF. Consider the equivalence in the space of distributions according to which two distributions F and F' are equivalent if and only if

$$P_F(\varphi(X) \in A) = P_{F'}(\varphi(X) \in A) \quad \text{for all } A \in \mathcal{G}.$$

We need only show that this relation is closed with respect to the relation FRF' if there exists $\bar{\tau}$ such that $F' = \bar{\tau}F$. But let $F' = \bar{\tau}F$. Then

$$P_{F'}(\varphi(X) \in A) = P_F(\varphi(\tau X) \in A) = P_{\bar{\tau}F}(\varphi(X) \in A) = P_{F'}(\varphi(X) \in A),$$

so that $F'RF$ implies $F' \sim F$ as was to be proved.

Consider now a nonparametric problem of the kind studied in the previous sections. For a suitable class \mathfrak{S} of transformations, the invariant tests (that is, those given by a critical function φ) are precisely those depending only on the ranks of the observations. It further turns out that the maximal invariant partition in the space of distributions coincides with our partition of the alternatives into nonparametric classes. Thus in particular, Theorems 3.1 and 7.1 are special cases of Theorem 8.1.

We shall now prove the correspondence just indicated in the case of the two-sample problem. The analogous result for the hypothesis of independence can be proved in much the same way. The sample space \mathfrak{X} is the $m + n$ -dimensional Euclidean space of points $(x_1, \dots, x_m, y_1, \dots, y_n)$. The family \mathfrak{F} is the family of all pairs (F, G) of continuous distribution functions. To define the class of transformations, consider the totality of real-valued (not necessarily continuous) functions τ of a real variable which are strictly increasing and such that $\tau(\pm\infty) = \pm\infty$. The transformations of \mathfrak{S} are obtained by applying the same function τ to all coordinates x_i and y_j . Since each τ is strictly increasing the inverse functions τ^{-1} are uniquely defined and their domains may be extended to be the entire real line through the condition that τ^{-1} should be everywhere nondecreasing. This will insure also that τ^{-1} is continuous. With these conventions, if a random variable X has continuous cdf F , the variable τX again has a continuous cdf, which is given by

$$(8.5) \quad \bar{\tau}F = F(\tau^{-1}).$$

It is clear that the maximal invariant with respect to \mathfrak{S} is given by the ranks of $X_1, \dots, X_m, Y_1, \dots, Y_n$, that is, by the equivalence

$$(x_1, \dots, x_m, y_1, \dots, y_n) \sim (x'_1, \dots, x'_m, y'_1, \dots, y'_n)$$

if and only if the two sets of numbers are in the same order. This relationship certainly is invariant, since with respect to a strictly increasing function the ordinates are in the same order relation as the abscissae. Conversely, given any two such sets of numbers there exists a strictly increasing function taking the first into the second.

What we shall now prove is that the maximal invariant classes in the space of distributions coincide with the classes $(f(F), g(F))$ of Section 4. In fact, we shall prove the more precise statement. The relation

$$(8.6) \quad \begin{aligned} F_1(\tau_1^{-1}) &= F_2(\tau_2^{-1}), \\ G_1(\tau_1^{-1}) &= G_2(\tau_2^{-1}) \end{aligned}$$

is equivalent to

$$(8.7) \quad \begin{aligned} F_1 &= f(F), & G_1 &= g(F) \\ F_2 &= f(F'), & G_2 &= g(F') \end{aligned}$$

where all the symbols have the significance given to them earlier. We first state the obvious

LEMMA 8.2.

(a) Given any two continuous cdf's F, F' there exist continuous, nondecreasing functions τ_1^{-1}, τ_2^{-1} such that

$$(8.8) \quad F(\tau_1^{-1}) = F'(\tau_2^{-1}).$$

(b) Given any two continuous cdf's F_1, G_1 there exist f, g, F such that

$$(8.9) \quad F_1 = f(F), \quad G_1 = g(F)$$

and such that $F(x) = F(y)$ whenever $F_1(x) = F_1(y)$ and $G_1(x) = G_1(y)$.

Suppose now that (8.7) holds and let τ_1, τ_2 be defined so as to satisfy (8.8). Then

$$F_1(\tau_1^{-1}) = f(F(\tau_1^{-1})) = f(F'(\tau_2^{-1})) = F_2(\tau_2^{-1})$$

and analogously for the G 's.

To prove the converse, assume that (8.6) holds and that f, g, F are defined so as to satisfy (8.9) as well as the side condition imposed there. If we put

$$(8.10) \quad F' = F(\tau_1^{-1}(\tau_2))$$

we have that $f(F') = F_1(\tau_1^{-1}(\tau_2)) = F_2(\tau_2^{-1}(\tau_2)) = F_2$ and similarly that $g(F') = G_2$. Thus our result is established provided we can show that F' as defined by (8.10) is continuous.

Since F and τ_1^{-1} are continuous by assumption, discontinuities in F' can result only from discontinuities in τ_2 . Suppose therefore that $\tau_2(x-) = a, \tau_2(x+) = b$. Then $\tau_2^{-1}(a) = \tau_2^{-1}(b)$ and hence because of (8.6) $F_1(\tau_1^{-1}(a)) = F_1(\tau_1^{-1}(b)), G_1(\tau_1^{-1}(a)) = G_1(\tau_1^{-1}(b))$. The assumption following (8.9) then implies $F(\tau_1^{-1}(a)) = F(\tau_1^{-1}(b))$ and therefore $F'(x-) = F'(x+)$.

REFERENCES

- [1] W. HOEFFDING, "Optimum nonparametric tests," *Proceedings of the Second Berkeley Symposium on Mathematical Statistics and Probability*, University of California Press, (1951), pp. 83-92.
- [2] A. M. MOOD, *Introduction to the theory of statistics*, McGraw-Hill Book Co., 1950.
- [3] F. WILCOXON, "Individual comparison by ranking methods," *Biometrics*, Vol. 1 (1945), pp. 80-83.
- [4] 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.
- [5] A. WALD AND J. WOLFOWITZ, "On a test whether two samples are from the same population," *Ann. Math. Stat.*, Vol. 11 (1940), pp. 147-162.
- [6] N. V. SMIRNOV, "On the estimation of the discrepancy between empirical curves of distribution for two independent samples," *Bull. Math. Univ. Moscow, Serie Int.*, Vol. 2 (1939), pp. 3-16.
- [7] E. L. LEHMANN, "Consistency and unbiasedness of certain nonparametric tests," *Ann. Math. Stat.*, Vol. 22 (1951), pp. 165-179.
- [8] J. WOLFOWITZ, "Nonparametric statistical inference," *Proceedings of the Berkeley Symposium on Mathematical Statistics and Probability*, University of California Press, (1949), pp. 93-113.

- [9] F. MASSEY, "A note on the power of a nonparametric test," *Ann. Math. Stat.*, Vol. 21 (1950), pp. 440-443.
- [10] N. SMIRNOV, "Table for estimating the goodness of fit of empirical distributions," *Ann. Math. Stat.*, Vol. 19 (1948), pp. 279-281.
- [11] F. MASSEY, "The distribution of the maximum deviation between two sample cumulative step functions," *Ann. Math. Stat.*, Vol. 22 (1951), pp. 125-130.
- [12] J. TUKEY, "The simplest signed-rank tests," Stat. Res. Group, Princeton University, *Memorandum Report 17*, 1949, unpublished.
- [13] J. WALSH, "Some significance tests for the median which are valid under very general conditions," *Ann. Math. Stat.*, Vol. 20 (1949), pp. 64-81.
- [14] J. NEYMAN AND E. S. PEARSON, "Contributions to the theory of testing statistical hypotheses. I. Unbiased critical regions of type A and type A_1 ," *Stat. Res. Memoirs*, Vol. 1 (1936), pp. 1-37.
- [15] A. WALD, "On the power function of the analysis of variance test," *Ann. Math. Stat.*, Vol. 13 (1942), pp. 434-439.
- [16] E. L. LEHMANN, "Some principles of the theory of testing hypotheses," *Ann. Math. Stat.*, Vol. 21 (1950), pp. 1-26.
- [17] P. R. HALMOS, *Measure Theory*, D. Van Nostrand, 1950.
- [18] J. WESTENBERG, "Significance test for median and interquartile range in samples from continuous populations of any form," *Koninklijke Nederlandse Akademie van Wetenschappen Proc.*, Vol. 51 (1948), pp. 252-261.
- [19] J. HEMELRIJK, "A family of parameter-free tests for symmetry with respect to a given point," *Nederl. Akad. Wetensch., Proc.*, Vol. 12 (1950), Part I, pp. 340-350, Part II, pp. 419-431.