SOME NONPARAMETRIC TESTS FOR TREATMENT EFFECTS IN PAIRED REPLICATIONS*

By H. S. Konijn

University of Sydney, Sydney, Australia

1. Summary

Several methods are considered for testing whether or not a certain treatment causes a shift in the distribution of some measured characteristic. Among the tests the distributions of which are studied in detail, the rank-sum test is most suitable for detecting small shifts if few *a priori* assumptions regarding the distribution of the measurements are justified. When members of each pair of objects have been drawn from a finite collection of objects, the sign test can be exceedingly inefficient in detecting small shifts; in other cases the asymptotic efficiency is more than one-half. In passing, tests for the two-sample location problem and some of their properties are reviewed in Sections 6–8.

2. Introduction

Consider n pairs of objects and suppose that one member of each pair has received a certain treatment, the other members serving as control. (Similarly, we could investigate two different treatments.) We shall suppose that sufficient care has been exercised so that, for each pair, the property to be observed would have the same distribution (F_i for each member of the i-th pair) if neither of the members had received the treatment, and would have the same distribution (G_i for each member of the i-th pair) if both members had received the treatment. This will be so not only if the objects are " perfectly matched " to begin with, but also if, for any given pair of objects, the assignment of treatments was random. [Hodges and Lehmann have commented on the desirability of perfect matching in (1954).]

We shall examine a few tests proposed for detecting whether the treatment has any effect, that is, whether $G_i = F_i$ for each i. Attention

* Presented by title to the April 1955 meeting of the American Mathematical Society, and in person to the Conference on Distribution–Free Methods of the Commonwealth Scientific and Industrial Research Organization in Melbourne during April 1956.

will be confined to the case in which the property of interest is a measured characteristic with F_i and G_i continuous, and to alternatives under which the treatment tends to either decrease or increase the value of this measurement. For simplicity we shall examine one-sided alternatives only—say, those under which the treatment tends to increase the measured characteristic.

Let $\mathbf{w}_i$ be the characteristic for the member which receives the treatment, $\mathbf{v}_i$ for the other member, and let $\mathbf{z}_i = \mathbf{w}_i - \mathbf{v}_i$ have the distribution Q_i. (We shall distinguish random variables from other quantities by printing them in bold face.)

Because of the continuity, the probability for any $\mathbf{z}_i$ to vanish is zero, and we shall suppose that there are no vanishing $\mathbf{z}$'s. In practice vanishing $\mathbf{z}$'s do occur and can usually be ascribed to limited fineness of measurement, that is, small discontinuities. Dixon and Massey (1951), in discussing the sign test, suggested ignoring the vanishing $\mathbf{z}$'s, which amounts to restricting oneself to the subpopulation in which the number of nonvanishing $\mathbf{z}$'s is given (and equal to the number observed). This method clearly preserves the validity of tests and is the best one for the discontinuous case (Putter, 1955). Similarly, in practice some of the $\mathbf{z}$'s sometimes turn out to be equal, for which case see Section 7.

We shall generally pay no attention in what follows to the fact to that, in any nonparametric test in which the decision to reject or not reject depends only on the observations, only a limited number of significance probabilities can be attained and, therefore, used to make comparisons of power.

3. Various Types of Experimental Conditions

When members of each pair of objects have been obtained at random (without replacement) from a finite population (for example, w_i and v_i may represent measurements on a randomly selected set of twins in the i-th litter), $\mathbf{v}_i$ and $\mathbf{w}_i$ are not independent. If, however, they can be regarded as random drawings from an indefinitely large population (for example, randomly selected individuals from a genetically homogeneous source), they are independent; the latter case will be referred to as independence *within pairs*. If the property to be observed for any one set of pairs is supposed to be distributed independently of that for any other set of pairs, we shall speak of independence *between pairs*.

The latter assumption is usually made. In a somewhat different context, Lehmann and Stein (1949) have pointed out that such an assumption is not always appropriate and have considered instead *symmetry* of the joint distribution function *in its arguments.* In the present case that would mean that if the joint distribution of the pairs $(v_i, w_i), \ldots, (v_n, w_n)$ is

$$Pr \cdot \{v_1 \leqq s_1, w_1 \leqq t_1; \ldots; v_n \leqq s_n, w_n \leqq t_n\} = A(s_1, t_1; \ldots; s_n, t_n)$$

and $(i_1, \ldots, i_n)$ is any permutation of $(1, \ldots, n)$, then

$$A(s_{i_1}, t_{i_1}; \ldots; s_{i_n}, t_{i_n}) = A(s_1, t_1; \ldots; s_n, t_n).$$

In particular, this would mean that each pair (v_i, w_i) would be supposed to have the same distribution. A more restricted and also more frequently applicable assumption would be for the z_i to have a joint distribution symmetric in its n arguments and so for each difference z_i to have the same distribution. Either assumption is fulfilled when the numbering of the pairs is random, in particular not only (obviously) when the pairs have themselves been drawn at random from some population (with replacement, if the population is finite), but also in the case of random partitions of the set of n pairs. If $i_1 < \ldots < i_m$ is a subset of $1, \ldots, n$ and $j_1 < \ldots < j_{n-m}$ is obtained from the set $1, \ldots, n$ by deleting $i_1, \ldots, i_m$, we understand by *the distribution* $B(t_1, \ldots, t_n)$ *of the random two-part partition* $z_{i1}, \ldots, z_{i_m}, z_{j_1}, \ldots, z_{j_{n-m}}$, the probability defined for any given numbers $t_1, \ldots, t_n$, that there exists a permutation $k_1, \ldots, k_m$ of $1, \ldots, m$ and a permutation $l_1, \ldots, l_{n-m}$ of $m + 1, \ldots, n$ such that

$$z_{i_1} \leqq t_{k1}, \ldots, z_{i_m} \leqq t_{k_m}, z_{j_1} \leqq t_{l_1}, \ldots, z_{j_{n-m}} \leqq t_{l_{n-m}}.$$

This probability is the same as the probability that

$$z_{h_1} \leqq t_1, \ldots, z_{h_m} \leqq t_m, z_{h_{m+1}} \leqq t_{m+1}, \ldots, z_{h_n} \leqq t_n,$$

for $h_1, \ldots, h_m$ a random permutation of $1, \ldots, n$.

When and only when the treatment has no effect, the z_i are, separately and jointly, distributed *symmetrically about zero.* This is so both in the case of randomization of treatments between members of a pair of objects and in the case of independence within the (perfectly matched) pairs. The distinction between the two cases is, however, reflected in the power of the tests, as we shall see.

4. Student's Test and Its Randomization Analog

For the case that the F_i are known to be normal distributions with the same variance σ^2, the z_i all have the same normal distribution with zero mean when there is no treatment effect, both in the case of independence within pairs and of randomization. For testing, in the case in which the G_i are also normal with variance σ^2, whether the treatment has any effect, Student's test performed on the z_i is very powerful, against the means of the G_i distributions exceeding those of the F_i distributions; in fact, most powerful against this excess being a constant.

Using the assumptions of Section 2, the same function of the observations can be compared in the population of 2^n *possible* assignment of treatments within pairs for $| w_i - v_i |$ given as observed, as noted by Fisher (1935), and Pitman (1937). Lehmann and Stein (1949) showed the test to be the most powerful nonparametric test against the above specified normal alternatives, and also (in the case of symmetry in the arguments of the distribution of the z_i when there is no treatment effect) against a wider class of alternatives. Note that

$$t' = n^{\frac{1}{2}} t \, (t^2 + n - 1)^{-\frac{1}{2}} = n \, \bar{z} \, (\Sigma \, z_i^2)^{-\frac{1}{2}}$$

is an increasing function of t. Since the test is performed by comparing t with its distribution for the $| z_i |$ given as observed, it is equivalent to compare $t'' = t' \, (\Sigma \, z_i^2)^{\frac{1}{2}} = n \, \bar{z}$ with its distribution for the $| z_i |$ given as observed, or $t''' = \frac{1}{2} \, (t'' + \Sigma \, | z_i |) = \Sigma_{(z_i > 0)} \, z_i$. For the case that the z_i have the same distribution with finite second moments and are independent, Hoeffding (1952) has shown that this test has asymptotically the same power as the corresponding Student test, whatever the common distribution of the z_i. However, the convergence may be quite slow if the z_i are not normal. For moderate size samples, this test becomes rather laborious to apply.

5. The Sign Test

Suppose one counts the number of positive z_i and rejects the hypothesis of no treatment effect if this is too large. For this well-known "sign" test Dixon and Mood (1946) discussed validity and [on the basis of previous work by Cochran (1937)] computed the local asymptotic efficiency with respect to Student's test against a sequence of alternatives under which the z_i are normal with the same variance and mean $\theta/\sqrt{n}$. In addition, they computed efficiencies for some finite values of n; see also Dixon (1953).

Let

$$k_i = Pr. \, \{w_i > v_i\} = 1 - Q_i \, (0), \tag{5.1}$$

where Q_i is the distribution of z_i. The hypothesis states that $k_i = \frac{1}{2}$ for all i and the alternatives imply that $k_i \geqq \frac{1}{2}$ for all i and $k_i > \frac{1}{2}$ for at least one i. Consider the alternatives for which $k_i = k > \frac{1}{2}$ for all i. Let Ω_d be the class of univariate distributions Q which are absolutely continuous with respect to a fixed σ-finite measure and for which $Q(0) = \frac{1}{2} + d$, and let Ω be the union of all Ω_d for $0 \leqq d < \frac{1}{2}$.

Consider any weight function which is nondecreasing in d and thus in $\theta/\sqrt{n}$. Then Hoeffding (1951) showed that among similar tests (that is, tests for which the probability of rejection is the same for all distributions satisfying the null hypothesis) of any given size, the sign test maximizes the greatest lower bound of the risk with respect to Q in $\Omega - \Omega_0$. Subsequently Fraser (1953) showed it to be most powerful. In our case we are restricted to the subspace of Ω for which the distributions are symmetric about some point; Ruist (1954) claims that the same result as Hoeffding's applies here, but his proof appears to be incorrect. We shall show below that the sign test has some rather unpleasant local properties, which would seem to imply that its use should not be generally encouraged.

6. The Two-sample Location Problem

Before considering further tests, let us examine a different problem. Consider n random variables $x_1, \ldots, x_m, y_1, \ldots y_{n-m}$ and the hypothesis that their joint distribution is symmetric in its arguments. One-sided alternatives may specify that this holds for $x_1', \ldots, x_m', y_1', \ldots, y_{n-m}'$, where the x_i are more likely to exceed the corresponding x_i' than not, and the y_i' equal the y_i or are more likely to exceed them than not.

There are several known tests for this so-called two-sample (location) problem. In the first place there is Student's test on the x's and y's appropriate in normal distributions, and there is the corresponding randomization test. The latter is equivalent with comparing the observed value of $\bar{x} - \bar{y}$ (or simply of $\Sigma^m_{i=1} x_i$) with its conditional distribution given a random permutation of the observed values of $x_1, \ldots, y_{n-m}$.

The rank-sum test and the median test will be discussed in the next two sections together with certain tests related to these two. Another test for this problem which uses the sample cumulative distribution function was proposed by Smirnov (1939); Marshall (1951) gave a version of it using a grouped sample distribution based on Doob's approach (1949) to Smirnov's results.

For a discussion of the power of certain of these tests, see Dixon (1954), Lehmann (1953), Van der Waerden (1952, 1953 *a* and 1953 *b*), Sundrum (1953), and Hodges and Lehmann (1956).

There are numerous tests designed for the general alternative that (if the x's have a common distribution H and the y's a common distribution K) $H \neq K$, and therefore not very powerful against the alternatives of the more narrowly defined location problem. We shall therefore refrain from discussing these.

7. The Rank-sum Test for the Two-sample Problem and Some Modifications of this Test

Let the ranks of the x's among the n random variables be $R_1, \ldots, R_m$. Festinger (1946) and Wilcoxon (1945 and 1947) proposed $\Sigma^m_{i=1} R_i$ as a test statistic. In our one-sided case, they would reject when this statistic would fall above a tabulated value. Let $U_{m,n}$ be the number of pairs (x_i, y_j) for which $x_i > y_j$, setting $U_{o,n} = U_{n,n} = 0$; this is Mann and Whitney's statistic (1947) which is related to $\Sigma^m_{i=1} R_i$ by

$$U_{m,n} = \Sigma^m_{i=1} R_i - \tfrac{1}{2} m (m + 1). \tag{7.1}$$

The rank-sum has mean $\frac{1}{2} m (n + 1)$ and variance $m (n - m) (n + 1)/12$ when the hypothesis holds, in which case Mann and Whitney established asymptotic normality. These results have been shown under the assumption of independence of the x's and y's but do not need this restriction. When some of the variables are equal, one may assign each of them the mean of the ranks they would be assigned if slightly different. This preserves the validity of the test and seems to be the best procedure in case of discontinuities (Putter, 1955).

Now assume independence of the x's and y's under alternative hypotheses, and let the x's have the same distribution H and the y's the same distribution K. Lehmann (1951) showed unbiasedness against alternatives for which $H(t) \leq K(t)$ for all t, and extended a method of Hoeffding (1948) to show that the test statistic is asymptotically normal [nondegenerate when $0 < Pr. (x_i > y_j) < 1$] under $H \neq K$ with m/n converging to a positive constant less than 1. Andrews (1954) proved asymptotic normality under a *sequence* of alternatives under which the distributions of $y_j + \theta/\sqrt{n}$ and x_i are the same, namely, H. For that purpose assume that H is absolutely continuous with

$$H(x) = \int_{-\infty}^{x} H'(t)\, dt, \text{ that } \int_{-\infty}^{\infty} H' dH < \infty \text{ and that}$$

$$\lim_{n \to \infty} \sqrt{n} \int_{-\infty}^{\infty} \left\{ H\left(x + \frac{\theta}{\sqrt{n}}\right) - H(x) \right\} dH(x) = \theta \int_{-\infty}^{\infty} H'\, dH.$$

The question of improving the power of this test, where it is known or believed that H is of a particular form, or close to it in an appropriate sense, but where it is still desired to use a non-parametric test, has been considered by Hoeffding (1951) and van der Waerden (1953 and 1953 *c*). Let $t_1, \ldots, t_n$ be a set of n independent observations on a random variable with distribution H, rearranged in ascending order of magnitude, and R_i^H the conditional expectation of t_R given R_i when the hypothesis is true. Hoeffding has shown that for $H = \Phi$, the standard normal distribution, $\Sigma^m_{i=1} R_i^H$ gives a test of the hypothesis discussed in this section, which among rank tests is locally most powerful against the x_i and the $y_j + \theta'$ having the same normal distribution. This test is also given in Tables 20 and 21 in Fisher and Yates (1938) and has been studied further by Terry (1952); extension to certain nonnormal H has been suggested by Dwass (1953). Van der Waerden suggested the numerically very similar statistic $\Sigma^m_{i=1}\Phi^{-1}\{R_i/(n+1)\}$. He pointed out that in the presence of " spurious outlayers " an otherwise normal population is still " close to normal " for his purpose.

8. The Median Test for the Two-Sample Problem and Some Related Tests

Another test of this hypothesis is the median test, due to Westenberg (1948, 1950 and 1952), in which we count the number $L_{m,n}$ of x's which exceed the sample median of all the n random variables. When the hypothesis holds, $L_{m,n}$ has a hypergeometric distribution. For $L_{m,n}$ is the number of observations above the median in a sample of m out of n, there being $n - [(n+1)/2]$ observations exceeding the median. Consequently, under the hypothesis, $L_{m,n}$ has mean $\frac{1}{2}m$ and variance $\frac{1}{4}m\,(n-m)/(n-1)$ when n is even, and mean $\frac{1}{2}m\,(1-1/n)$ and variance $\frac{1}{4}m\,(n-m)\,(1-1/n^2)/(n-1)$ when n is odd, and has an asymptotic normal distribution when m/n converges to a constant. Hemelrijk (1950 *a* and 1950 *b*) has noted that in case of equal observations the test can still validly be applied as a conditional one given the number of z's exceeding the median and has briefly considered the use of a quantile different from the median or of more than one quantile.

Now assume independence of the x's and y's under the alternative hypotheses, and let the x's have the same distribution H and the y's the same distribution K. Again, it follows from Lehmann (1951) that the test is unbiased against alternatives for which $H(t) \leq K(t)$ for all t. Mood (1954) has sketched a proof of asymptotic normality under any alternative with H and K possessing densities H' and K'

differing from zero at a median c of the mixed distribution $(m/n) H + \{(n - m)/n\} K$ and m/n tending to a positive constant k less than 1:

$$Pr. \left[\sqrt{m} \left\{ m^{-1} L_{m,n} - 1 + H(c) \right\} \times \left\{ \frac{1}{H(c)\{1 - H(c)\}} + \frac{m}{n-m} \frac{1}{K(c)\{1 - K(c)\}} \right\}^{\frac{1}{2}} \leq t \right] \quad (8.1)$$

converges to $\Phi(t)$. Andrews (1954) proved asymptotic normality under a *sequence* of alternatives under which the distributions of $y + \theta/\sqrt{n}$ and x_i are both H.

A method which utilizes the number of x's exceeding the median of the y's rather than the median of all n observations was given, with tabulations, by Mathisen (1943) and was already discussed in more general terms by Thompson (1938). Behrens (1933, Chapter 3 *b*) discussed and tabulated (incorrectly) a somewhat similar test (he gave a two-sided test only) which is most easily described for m and $n - m$ odd, say equal to $2h - 1$ and $2k - 1$. Let the x's arranged in ascending order of magnitude be denoted by $x^{(1)}, \ldots, x^{(2h-1)}$, the y's so arranged by $y^{(1)}, \ldots, y^{(2k-1)}$. He would reject if $x^{(h)} \geq y^{(k)}$ and if S, the largest nonnegative integer satisfying $x^{(h-S)} > y^{(k+S)}$, would exceed a tabulated value. (Behrens actually uses S *plus* the fraction

$$\frac{\{x^{(h-S)} - y^{(k+S)}\}}{\{x^{(h-S)} - x^{(h-S-1)} + y^{(k+S+1)} - y^{(k+S)}\}},$$

but this fraction evidently cannot affect the test.) Finally we mention an interesting test applied as early as 1876 by Galton for Darwin (1876, p. 17): reject where for too many i

$$x^{(i)} > y^{(i)} \ (i = 1, \ldots, m = n - m).$$

See Hodges (1955). For none of these tests have we studied the power.

9. Application of Two-sample Tests to the Testing of the Hypothesis of Symmetry about Zero

Hemelrijk (1950 *a* and *b*) has noted that any two-sample test symmetric in the observations can be used to construct tests for the hypothesis of symmetry about zero. Under that hypothesis the number m of positive z's is a binomial random variable with parameters $\frac{1}{2}$ and n, and, moreover, the division of the set of measurement differences into positive and negative ones constitutes a random partition of the $|z|$'s.

Therefore, by Section 3, if, in random order, $x_1, \ldots, x_m$ constitute the positive and $y_1, \ldots, y_{n-m}$ the absolute values of the negative z's, we have under the hypothesis that the joint distribution of

$$x_1, \ldots, x_m, y_1, \ldots, y_{n-m}$$

is symmetric in its n arguments.

Let us consider the class of similar, level α tests of our hypothesis which are symmetric in the observations $z_1, \ldots, z_n$. (Incidentally, for such tests it is really immaterial whether the order of listing the observations is fixed or random.) We are indebted to Lehmann for the remark that this class constitutes the totality of similar regions for testing our hypothesis, due to the fact that the set of $|z|$'s constitute of sufficient complete statistic for this hypothesis, and have Neyman structure with respect to this set (which means that the conditional probability of rejection equals α for any given set of values for the $|z|$'s). [For the notions involved in this remark, *see* Lehmann and Scheffé (1950).]

The simplest of the tests of this class is the sign test, which does not utilize the $|z|$'s at all. The randomization test mentioned in Section 4 also belongs to this class. It is merely one of the class of tests obtained by applying any two-sample test statistic symmetric in the observations with m as observed and comparing it with its (unconditional) null distribution for random m and a given set of $|z|$'s.

The rank-sum test was applied to this problem by Wilcoxon (1945); in 1947 he gave a table to significance points for $n \leqq 20$, beyond which number the normal approximation fits well. In 1949 and elsewhere, Walsh examined a class of tests which was noted by Tukey [*see* Lehmann (1953, p. 36)] to include the rank-sum test. Walsh examined in more detail the most efficient tests in a subclass of his class of tests for which significance levels are easily computed; we shall not discuss these here.

The variance of the rank-sum statistic equals the mean of the conditional variance of the rank-sum given m, *plus* the variance of the conditional mean of the rank-sum, which under the hypothesis is

$$\frac{E\,m\,(n-m)\,(n+1)}{12} + \operatorname{var}\{\tfrac{1}{2}\,m\,(n+1)\} = \frac{n\,(n-1)\,(n+1)}{48}$$

$$+ \frac{n\,(n+1)^2}{16} = \frac{n\,(n+1)\,(2n+1)}{24}, \qquad (9\cdot1)$$

which nearly equals the expression given by Wilcoxon (1947). The mean under the hypothesis is

$$E\,\tfrac{1}{2}m\,(n+1) = \tfrac{1}{4}n\,(n+1). \tag{9.2}$$

For the Mann-Whitney test applied to this problem one gets under the null hypothesis a mean of

$$E\,\tfrac{1}{2}m\,(n-m) = \frac{n\,(n-1)}{8} \tag{9.4}$$

and a variance of

$$\frac{E\,m\,(n-m)\,(n+1)}{12} + \operatorname{var}\left\{\tfrac{1}{2}m\,(n-m)\right\} = \frac{n\,(n-1)\,(n+1)}{48}$$

$$+\frac{n\,(n-1)}{32} = \frac{n\,(n-1)\,(2n+5)}{96}. \tag{9.3}$$

The median test applied to this problem has a mean of

$$E\,\tfrac{1}{2}m = \tfrac{1}{2}n \tag{9.5}$$

for n even and of

$$E\,\tfrac{1}{2}m\left(1-\frac{1}{n}\right) = \tfrac{1}{4}n\left(1-\frac{1}{n}\right) \tag{9.5 a}$$

for n odd, and a variance of

$$\frac{E\,\tfrac{1}{4}m\,(n-m)}{(n-1)} + \operatorname{var}\left(\tfrac{1}{2}m\right) = \frac{n}{16}+\frac{n}{16} = \frac{n}{8} \tag{9.6}$$

for n even,

$$E\,\tfrac{1}{4}m\,(n-m)\frac{\left(1-\frac{1}{n^2}\right)}{(n-1)} + \operatorname{var}\left\{\tfrac{1}{2}m\left(1-\frac{1}{n}\right)\right\}$$

$$= \frac{(n^2-1)}{(16n)} + \frac{(n-1)^2}{(16n)} = \frac{(n-1)}{8} \tag{9.6 a}$$

for n odd, when the hypothesis holds.

One can similarly apply the other tests mentioned in Sections 6–8; Smirnov did this in 1947 for his two-sample test.

Hemelrijk's proposal in 1950 *a* and *b* was to combine the sign test with such two-sample tests in the following manner. Select a number ζ less than the level of significance α and find the smallest number m_0, such that $Pr.\,\{m > m_0\} \leq \zeta$ under the hypothesis. Reject the hypothesis if $m > m_0$ or if $m \leq m_0$ but the two-sample test (with m as observed, that is, applied as a conditional test) rejects at level ϵ, where ϵ is the largest number for which the rejection probability for the whole procedure does not exceed α. As we shall see, for the shift alternatives considered below, the sign test is often of relatively low power, so that for these alternatives Hemelrijk's tests are not most suitable. Nevertheless, his method of constructing tests leads to tests which are evidently more powerful against certain interesting alternatives mentioned by him under the heading of " Type T_2 region ", which type of region differs somewhat from the one mentioned above which is called " Type T_1 region ".

10. Power against Shift Alternatives Assuming Independence between Pairs; the Randomisation Test and the Sign Test

In the remainder of this paper we shall be concerned with the power of the randomization and sign tests, the rank-sum and the Mann-Whitney tests, and the median test for the hypothesis of absence of treatment effect against alternatives under which the treatment shifts the distribution of values of the characteristic of the treated member of the i-th pair of objects by $\theta'(i)$. In that case the distribution S_i of $z_i - \theta'(i)$ is symmetric about 0. We shall assume independence between pairs in all discussions of power. We shall use Pitman's methods of studying local asymptotic power; for definitions and methods the reader might consult Konijn (1956).

Whenever we shall write the following symbols

$$S_n = \frac{\Sigma^n{}_{i=1} S_i}{n}, \qquad \bar{\theta'}_n = \frac{\Sigma^n{}_{i=1} \theta'(i)}{n},$$

$$\sigma_i^2 = \int_{-\infty}^{\infty} \{t - \int_{-\infty}^{\infty} s dF_i(s)\}^2 dF_i(t), \qquad \bar{\sigma}_n^2 = \frac{\Sigma^n{}_{i=1} \sigma_i^2}{n},$$

$$\tau_i^2 = \int_{-\infty}^{\infty} t^2 dS_i(t), \qquad \bar{\tau}_n^2 = \frac{\Sigma^n{}_{i=1} \tau_i^2}{n},$$

we shall suppose that the right hand sides exist and as $n \to \infty$ have limits which are well-defined and finite.

The randomization test is asymptotically equivalent to Student's test or the corresponding test with known variances τ_i^2, which at $\bar{\theta}_n'$ near zero has asymptotic power

$$1 - \Phi\left\{\delta - \frac{\bar{\theta}_n' \sqrt{n}}{\bar{\tau}_n}\right\} \tag{10.1}$$

since under the alternative $\sqrt{n}\,\bar{z}/\bar{\tau}_n$ has mean $\bar{\theta}_n'\sqrt{n}/\bar{\tau}_n$ and variance 1, and $Pr.\ \{\sqrt{n}\,(\bar{z} - \bar{\theta}_n')/\bar{\tau}_n \leqq t\}$ is independent of the $\theta'(i)$ and converges to $\Phi(t)$. Since under independence within pairs $\bar{\tau}_n^2 = 2\bar{\sigma}_n^2$ we have then the special result

$$1 - \Phi\left\{\delta - \frac{\bar{\theta}_n'\sqrt{n}}{(\bar{\sigma}_n\sqrt{2})}\right\} \qquad (10.1\,a)$$

The asymptotic distribution of $\boldsymbol{m}$ *minus* its mean and divided by its standard deviation is easily seen to be normal under a sequence of alternatives with shifts $\theta/\sqrt{n}$ for each z_i. In fact $\boldsymbol{m}/n$ is a U-statistic in the sense of Theorem 8.2 of Hoeffding (1948) with $\Phi(z_i) = k_i$, $\Psi_{1(i)}(z_i) = \boldsymbol{m}_i - k_i$, denoting by $\boldsymbol{m}_i$ the random variable with mean k_i which is unity of $z_i > 0$ and 0 otherwise. Also, (8.15) and (8.16) of that paper are fulfilled uniformly in θ near 0 provided the k_i are bounded away from 0 and 1, while (8.2) and (8.3) are evidently fulfilled uniformly in θ.

Since $E\,\boldsymbol{m}_i = 1 - Q_i(0)$, and var $\{\boldsymbol{m}_i\} = Q_i(0)\,\{1 - Q_i(0)\}$, we have under a shift by $\theta/\sqrt{n}$

$$\sqrt{n}\,\frac{d}{d\theta}\,E\,\boldsymbol{m}\Big]_{\theta=0} = \Sigma^n_{i=1}\,S_i'(0),$$

while the variance converges to $\frac{1}{4}n$. If $\bar{S}_n'$ exists and is continuous near the origin, the asymptotic power of the sign test for a fixed θ' near 0 is, therefore

$$1 - \Phi\,\{\delta - 2\theta'\sqrt{n}\,\bar{S}'_n(0)\}, \qquad (10.2)$$

and the asymptotic efficiency with respect to the randomization test is the square of

$$2\,\bar{\tau}_n\,\bar{S}'_n(0), \qquad (10.3)$$

which is $2/\pi$ in the normal case. In general the efficiency is highly dependent on local properties of the density of the distribution and has a greatest lower bound of zero, so that the sign test is not in general to be recommended as a symmetry test against shift alternatives. Under conditions of independence within pairs, however, the local asymptotic power is (when the F_i are all equal to F and the conditions stated in Section 7 on H hold for F)

$$1 - \Phi\,\{\delta - 2\theta'\sqrt{n}\int_{-\infty}^{\infty}F'dF\}, \qquad (10.2\,a)$$

and the local asymptotic efficiency becomes the square of

$$2\sqrt{2}\,\bar{\sigma}\int_{-\infty}^{\infty} F'\,dF, \qquad (10.3\,a)$$

which, as follows from Hodges and Lehmann (1956), has a minimum of (2/3) (0·864) = 0·576. (It is again $2/\pi$ in the normal case.) The same results hold if independence between pairs is not fulfilled, but there exist independent variates u_i, v_i', w_i' such that

$$v_i = u_i + v_i', \qquad w_i' = u_i + w_i'. \qquad (10.4)$$

In either case the local asymptotic power of the sign test is easily seen not to be affected by a scale change (the same for each i) accompanying the shift when the F_i are symmetric.

It is evident that for $\boldsymbol{m} = m$ the joint distribution of $x_1, \ldots, x_m$ is symmetric in its arguments, and that the same holds for the distribution of $y_1, \ldots, y_{n-m}$. In the following we shall take $\theta'(i)$ constant and the z's independent. In that case the joint distribution of the x's and y's for $\boldsymbol{m} = m$ is the average over the $n!$ permutations $(h_1, \ldots, h_n)$ of

$$\Pi^m_{i=1}\left\{\frac{S_{h_i}(t_i-\theta') - S_{h_i}(-\theta')}{1 - S_{h_i}(-\theta')}\right\}\Pi^n_{j=m+1}\left\{\frac{S_{h_j}(-\theta') - S_{h_j}(-t_j-\theta')}{S_{h_j}(-\theta')}\right\}, \qquad (10.5)$$

for $t_1, \ldots, t_n$ nonnegative, so that any x_i has distribution

$$Pr.\,(x_i \leqq t) = \frac{1}{n}\Sigma^n_{k=1}\left\{\frac{S_k(t-\theta') - S_k(-\theta')}{1 - S_k(-\theta')}\right\} \quad (t \geqq 0), \qquad (10.6)$$

and any y_j

$$Pr.\,(y_j \leqq t) = \frac{1}{n}\Sigma^n_{k=1}\left\{\frac{S_k(-\theta') - S_k(-t-\theta')}{S_k(-\theta')}\right\} \quad (t \geqq 0). \qquad (10.7)$$

Note that in general the x's and y's will not be independent, unless $S_k = S$ for all k.

In Section 11 we shall state and prove a general proposition which will be used to derive asymptotic distributions of the remaining tests under shift alternatives in Sections 12 and 13.

11. A Proposition Concerning the Limit Distribution of a Class of Random Variables for which the Limit of the Conditional Distribution is known

In the present context and many other situations we meet with a problem of the following kind. A given statistic $\boldsymbol{B}_{m,n}$ (such as $\boldsymbol{U}_{m,n}$

or $\boldsymbol{L}_{m,\,n}$) depends on n and also on another nonnegative integer m. Moreover, we know that for m and n increasing indefinitely with $m/n = m\,(n)/n$ approaching any positive k, the statistic is asymptotically normally distributed with asymptotic variance $f_1^2\,(k)$:

$$\boldsymbol{b}_{m(n),\,n}/f_1\,(k) \to \boldsymbol{B} \text{ in law } (\boldsymbol{b}_{m,\,n} = \boldsymbol{B}_{m,\,n} - E\,\boldsymbol{B}_{m,\,n}), \qquad (11.1)$$

where $\boldsymbol{B}$ is a standard normal variate: $Pr.\,(\boldsymbol{B} \leq t) = \Phi\,(t)$.

Suppose $m = m\,(n)/n$ itself is determined by a chance process $\boldsymbol{m}\,(n)$, which has the property that $\boldsymbol{m}\,(n)/n$ converges in probability to k:

$$\boldsymbol{m}\,(n)/n \to k \text{ in probability}; \qquad (11.2)$$

and consider the statistic

$$\boldsymbol{A}_n = \boldsymbol{B}_{\boldsymbol{m}\,(n),\,n} \qquad (11.3)$$

whose conditional asymptotic distribution for $\boldsymbol{m}\,(n) = m\,(n)$ is already known to be normal. Let the conditional expectation of $\boldsymbol{B}_{m,\,n}$ given $\boldsymbol{m}$ be denoted by

$$C\,(\boldsymbol{m}, n). \qquad (11.4)$$

Our aim is to find the asymptotic marginal (unconditional) distribution of $\boldsymbol{A}_n$ or of

$$\boldsymbol{a}_n = \boldsymbol{A}_n - E\,\boldsymbol{A}_n. \qquad (11.5)$$

It is well known that the variance of $\boldsymbol{a}_n$ if it exists equals the sum of the expectation of the variance of $\boldsymbol{B}_{m,\,n}$ and the variance of $C\,(\boldsymbol{m}, n)$. We shall now prove that if the asymptotic distribution of $C\,\{\boldsymbol{m}\,(n), n\}$ is normal with asymptotic variance $f_2^2\,(k)$:

$$c\,\{\boldsymbol{m}\,(n), n\}/f_2\,(k) \to \boldsymbol{B} \text{ in law } \{c\,(m, n) = C\,(m, n) - EC\,(\boldsymbol{m}, n)\}, \qquad (11.6)$$

then $\boldsymbol{a}_n/f\,(k)$ converges in law to $\boldsymbol{B}$, where

$$f^2\,(k) = f_2^2\,(k) + f_2^2\,(k). \qquad (11.7)$$

(Of course, the notation $\boldsymbol{B}$ is meant to be generic, so that the different $\boldsymbol{B}$'s are not necessarily the *same* random variables.)

For our purposes we need a somewhat more general result, since we are interested in the limit distribution under a sequence of parameters which generally depends on n and in our case converges to 0: $\theta' = \theta/\sqrt{n}$. We shall suppose that $\boldsymbol{m}\,(n)/n$ converges in probability to a positive number k_0 less than 1 depending only on the limit of that parameter sequence (in our applications $k_0 = \frac{1}{2}$), but that for any sufficiently small but fixed θ', $[\boldsymbol{m}\,(n)/n] - k$ converges to zero uniformly (here k depends on n through $\theta' = \theta\sqrt{n}$). In our applications this condition is fulfilled as pointed out in the discussion of the asymptotic distribution of $\boldsymbol{m}$ in Section 10. Similarly, we suppose that the asymptotic variance of $C\,\{\boldsymbol{m}\,(n), n\}$ is $f_2^{\,2}\,(k_o)$. We are indebted to Dr. J. Putter (who in turn acknowledges the idea of using sequence spaces to Dr. L. Le Cam) for assistance in crucial points of the demonstration that constitutes the rest of this section.

To prove the proposition, note that

$$\boldsymbol{a}_n = \boldsymbol{b}_{\boldsymbol{m}\,(n),\,n} + c\,\{\boldsymbol{m}\,(n), n\},$$

so that, if

$$\phi_n\,(t) = E \text{ exp. } (it\,\boldsymbol{a}_n),$$

we have

$$\phi_n\,(t) = E\,[\text{exp. }\{it\,c\,\big(\boldsymbol{m}\,(n),\, n\big)\}\,E\,\{\text{exp. }(it\,\boldsymbol{b}_{\boldsymbol{m}\,(n),\,n})\,|\,\boldsymbol{m}\,(n)\}].$$

For any increasing sequence ν', ν'',.... of natural numbers we define the space $E_{\nu'}$, $E_{\nu''}$, ... of finite increasing sequences $S_{\nu'}$, $S_{\nu''}$, ... of natural numbers defined by

$s_{\nu'}$ is the one-member sequence $m\,(\nu')$,

$s_{\nu''}$ is the two-member sequence $m\,(\nu')$, $m\,(\nu'')$,

.................................;

and on them the measures $P_{\nu'}$, $P_{\nu''}$.... generated by the distributions of $\boldsymbol{m}\,(\nu')$ under $\theta/\sqrt{\nu'}$, $\boldsymbol{m}\,(\nu')$ and $\boldsymbol{m}\,(\nu'')$ under $\theta/\sqrt{\nu''}$,..... (Or one can use measures on the corresponding cylinder sets.)

Now let n', n'',.... be a subsequence of the sequence of natural numbers. Given a decreasing sequence ϵ_1, ϵ_2,.... of positive numbers with finite positive sum $\epsilon < 1$, it follows from the convergence condition on $[\boldsymbol{m}\,(n)/n] - k$, by methods identical to those used in the proof of the Borel-Cantelli lemma, that there exists a further subsequence ν', ν'', ... and a collection of subsets $e_{\nu'}$, of $E_{\nu'}$, $e_{\nu''}$ of $E_{\nu''}$, ... ($E_{\nu'}$, $E_{\nu_i''}$.... as defined above) such that for all i

$$\left|\frac{m\,(\nu')}{\nu'} - k_o\right| < \epsilon_1, \;\ldots,\; \left|\frac{m\,(\nu^i)}{\nu^i} - k_o\right| < \epsilon_i$$

for $\{m\,(\nu'), \ldots, m\,(\nu^i)\}$ in e_{ν_i} and

$$P_{\nu_i}\,(e_{\nu_i}) > 1 - \Sigma^i_{\,j=1}\,\epsilon_j$$

which exceeds $1 - \epsilon$. Let e_ϵ be the intersection of $e_{\nu'}, e_{\nu''}, \ldots,$ then e_ϵ is not empty, for $\lim_{\nu \to \infty} P_\nu (e_\nu) > 1 - \epsilon$; it is the set on which $m(\nu)/\nu$ converges to k_o.

We have that $g_\nu \{t, m(\nu)\} = E \exp. (it\, \boldsymbol{b}_{m(\nu),\, \nu})$ converges to $\exp. \{-\frac{1}{2} f_1^2 (k_o)\, t^2\}$ on e_ϵ. Therefore the convergence of $E \exp. [it\, c\, \{m(\nu), v\}]$ to $\exp. \{-\frac{1}{2} f_2^2 (k_o)\, t^2\}$ implies that given $\epsilon > 0$ there exists ν_o such that for all elements in the ν-sequence exceeding ν_o

$$\Big| \int_{E_\nu} \exp. [it\, c\, \{m(\nu), \nu\}].\ g_\nu \{t, m(\nu)\}\, dP_\nu (e) - \exp. (-\tfrac{1}{2} f^2 (k_o)\, t^2) \Big|$$

$$\leqq \int_{e_\nu} | \exp. [it\, c\, \{m(\nu), \nu\}] | \cdot | g_\nu \{t, m(v)\}$$

$$- \exp. \{-\tfrac{1}{2} f_1^2 (k_o)\, t^2\} |\, dP_\nu (e)$$

$$+ \exp. \{-\tfrac{1}{2} f_1^2 (k_o)\, t^2\} \int_{e_\nu} | \exp. [it\, c\, \{m(\nu), \nu\}]$$

$$- \exp. (-\tfrac{1}{2} f_2^2 (k_o)\, t^2) |\, dP_\nu (e) + \epsilon$$

$$\leqq \epsilon \int_{E_\nu} | \exp. [it\, c\, \{m(\nu), \nu\}] |\, dP_\nu (e)$$

$$+ \exp. (-\tfrac{1}{2} f_1^2 (k_o)\, t^2) \int_{E_\nu} | \exp. [it\, c\, \{m(\nu), \nu\}]$$

$$- \tfrac{1}{2} f_2^2 (k_o)\, t^2 |\, dP_\nu (e) + 2\, \epsilon$$

$$\leqq \epsilon + \epsilon \exp. [-\tfrac{1}{2} f_1^2 (k_o)\, t^2] + 2\, \epsilon \leqq 4\epsilon.$$

So every subsequence of the natural numbers contains a further (ν-) subsequence for which $\log \phi_\nu (t)$ converges to $-\frac{1}{2} f^2 (k_o)\, t^2$. Therefore, $\log \phi_n (t)$ converges to $-\frac{1}{2} f^2 (k_o)\, t^2$, and the proposition is proved.

12. Asymptotic Distribution of the Mann-Whitney Statistic and the Rank-sum, and the Power of the Corresponding Tests for Symmetry about Zero against Shift Alternatives

We apply the proposition of the previous section to $\boldsymbol{U}_{m,\, n}$ and the rank-sum by writing $\boldsymbol{B}_{m,\, n}$ for $2\sqrt{3}\, \boldsymbol{U}_{m,\, n}/\{n\sqrt{(n+1)}\}$ and $2\sqrt{3}\, \Sigma^m_{i=1} \boldsymbol{R}_i/\{n\sqrt{(n+1)}\}$, respectively. For shift alternatives $\theta/\sqrt{n}$ we have by Section 7 (using the equidistribution of the $\boldsymbol{x}$'s and of the $\boldsymbol{y}$'s show in Section 10) that in both cases $\boldsymbol{b}_{m,\, n}$ converges to $\frac{1}{2} \boldsymbol{B}$ if m/n converges to $\frac{1}{2}$. Now, if the Q_i are the same, $c\, (\boldsymbol{m}, n)$ comes out to be

$$\frac{2\sqrt{3}}{k(1-k)} \int_0^\infty \{Q(0) - Q(-t)\}\, dQ(t) \left\{\frac{m - kn}{\sqrt{(n+1)}} \left(1 - \frac{m}{n} - k\right) + \frac{k(1-k)}{\sqrt{(n+1)}}\right\},$$

and

$$2\sqrt{3}\frac{m-kn}{\sqrt{(n+1)}}\left[\frac{m+1}{2n}+\frac{k}{2}+\frac{1-\frac{m}{n}-k}{k(1-k)}\int_0^\infty \{Q(0)-Q(-t)\}\,dQ(t)\right]$$

$$-\sqrt{3}\frac{1}{\sqrt{(n+1)}}\left[k(1-k)-2\int_0^\infty \{Q(0)-Q(-t)\}\,dQ(t)\right]$$

respectively. Since under $\theta/\sqrt{n}$

$$p\lim_{n\to\infty}\frac{m}{n}=\lim_{n\to\infty}k=4\lim_{n\to\infty}\int_0^\infty\{Q(0)-Q(-t)\}\,dQ(t)=\tfrac{1}{2},$$

$$1-\frac{m}{n}-k \quad\text{and}\quad k(1-k)-2\int_0^\infty\{Q(0)-Q(-t)\}\,dQ(t)$$

converge to zero, and $c(\boldsymbol{m}, n)$ converges in law to 0 and $\frac{1}{2}\sqrt{3}\,\boldsymbol{B}$ respectively; and the same result holds if the S_i differ and satisfy the conditions on H of Section 7. The proposition then yields that $\boldsymbol{a}_n=\boldsymbol{B}_{m,n}-E\boldsymbol{B}_{m,n}$ converges in law to $\boldsymbol{B}/2$ and $\boldsymbol{B}$ respectively.

Under the abovementioned conditions on the S_i together with continuity of the S_i' at the origin we also easily establish that the variance and the derivative of the means of these statistics are continuous in θ and that at $\theta=0$

$$n^{-3/2}\frac{n}{n-1}\frac{d}{d\theta}E\boldsymbol{U}_{m,n}=\overline{I}_n-\tfrac{1}{2}\overline{S}_n'(0),$$

$$n^{-3/2}\frac{n}{n-1}\frac{d}{d\theta}E\Sigma^{m}{}_{i=1}\boldsymbol{R}_i=\overline{I}_n+\frac{1}{n-1}\overline{S}_n'(0),$$

where

$$\overline{I}_n=n^{-2}\Sigma^n{}_{i=1}\Sigma^n{}_{k=1}\int_{-\infty}^{\infty}S_i'(t)\,S_k'(t)\,dt,\quad \overline{S}_n'(0)=n^{-1}\Sigma^n{}_{i=1}S_i'(0). \qquad (12.1)$$

Therefore, when the S_i satisfy the conditions on H in Section 7 and the S_i' are continuous at 0 (or, sufficient in the case of the rank-sum test, $\overline{S}_n'$ is bounded near 0) and $\overline{I}_n$ and $\overline{S}_n'(0)$ have limits, the asymptotic power of the Mann-Whitney test for a fixed θ' near 0 is

$$1-\Phi\,[\delta-4\sqrt{3}\,\theta'\sqrt{n}\,\{\overline{I}_n-\tfrac{1}{2}\overline{S}_n'(0)\}] \qquad (12.2)$$

and of the rank-sum test

$$1-\Phi\,\{\delta-2\sqrt{3}\,\theta'\sqrt{n}\,I_n\}. \qquad (12.3)$$

Thus in the normal case, we obtain a local asymptotic efficiency with respect to the randomization analog of Student's test of $(3/\pi)\cdot 4\,(1\frac{1}{2} - \sqrt{2}) = (\cdot 343)\cdot 3/\pi = \cdot 32$ and $3/\pi$ respectively. The minimum of $12\,\bar{I}_n^2\,\bar{\tau}_n^2$ is $\cdot 864$. Evidently, the Mann-Whitney test for symmetry about zero will often be less powerful against shift alternatives than the rank-sum test.

13. Asymptotic Distribution of the Median Test Statistic, and the Power of the Median Test for Symmetry about Zero against Shift Alternatives

For c_o a point satisfying

$$m\,\{\bar{S}_n(c_o) - \tfrac{1}{2}\} + (n-m)\,\{\tfrac{1}{2} - \bar{S}_n(-c_o)\} = \tfrac{1}{2}\,n$$

with $\bar{S}_n = \Sigma^n_{i=1} S_i/n$, we have by the symmetry of the S_i that $\bar{S}_n(c_o) = \frac{3}{4}$, $\bar{S}_n(-c_o) = \frac{1}{4}$. Let $\bar{S}_n$ have a continuous derivative throughout a neighbourhood of c_o with $\bar{S}_n'(c_o) \neq 0$; then, as is easy to see, the equation

$$\frac{m}{n}\,\Sigma^n_{i=1}\frac{S_i(c-\theta')-S_i(-\theta')}{1-S_i(-\theta')} + \frac{n-m}{n}\,\Sigma^n_{i=1}\frac{S_i(-\theta')-S_i(-c-\theta')}{S_i(-\theta')} = \tfrac{1}{2}\,n$$

or

$$\frac{m}{n}\left\{\frac{1}{n}\,\Sigma^n_{i=1}\frac{S_i(-c+\theta')}{S_i(\theta')} - \frac{1}{n}\,\Sigma^n_{i=1}\frac{S_i(-c-\theta')}{S_i(-\theta')}\right\} = \tfrac{1}{2} - \frac{1}{n}\,\Sigma^n_{i=1}\frac{S_i(-c-\theta')}{S_i(-\theta')}$$

has a unique solution for c near c_o if m/n and θ' are given and θ' is small. If the bracket vanishes, c is a one-to-one continuous function of θ' for small θ'. Suppose for small $\theta' \neq 0$ the bracket does not vanish and let $\mu(c\,/\,\theta')$ be the ratio of the right hand side by the bracket. For fixed small $\theta' \neq 0$ this is a one-to-one function of c, so that for $\mu(c\,/\,\theta') = \bar{\mu}$

$$c = \mu^{-1}(\bar{\mu}\,/\,\theta'),$$

and μ and μ^{-1} are continuous. One easily shows that for $\theta' = \theta/\sqrt{n}$ and m/n convergent,

$$\lim_{n\to\infty} \mu^{-1}\left(\frac{m}{n}\Big/\theta'\right) = c_o.$$

Therefore, if $m/n \to k$ and $\overline{S}_n'$ is continuous also at 0, we have under the shift alternatives, by a Taylor expansion

$$E\,L_{m,\,n} = \frac{m}{2} - 4\,\theta\sqrt{n}\,\frac{m}{n}\left(1 - \frac{m}{n}\right)\Big\{ - \overline{S}_n'(c_0) +$$

$$\frac{2}{n}\sum_{i=1}^{n} S_i'(0)\,\{1 - S_i(c_0)\}\Big\}\, f\left(\frac{m}{n}, \frac{\theta}{\sqrt{n}}\right),$$

with $f\left(\frac{m}{n}, \frac{\theta}{\sqrt{n}}\right) \to 1$. Write $B_{m,\,n}$ for $\dfrac{2L_{m,\,n}}{\sqrt{n}}$. (13·1)

By Section 8, we have that for shift alternatives $\theta/\sqrt{n}$, $b_{m,\,n}$ converges to $\frac{1}{2}B$ if m/n coverges to $\frac{1}{2}$ (using the fact of equidistribution of the x's and of the y's noted in Section 10). Also $2\,c\,(m, n)$ converges in law to B, since under $\theta/\sqrt{n}$

$$p\lim_{n\to\infty}\frac{m}{n} = \lim_{n\to\infty} k = \tfrac{1}{2}, \quad p\lim_{n\to\infty} f\left(\frac{m}{n}, \frac{\theta}{\sqrt{n}}\right) = 1,$$

and $2\,m/\sqrt{n}$ *minus* its mean converges in law to B.

The asymptotic variance of $b_{m,\,n}$ evidently converges to the asymptotic variance under the null hypothesis, $\frac{1}{4}$. The asymptotic variance of $c\,(m, n)$ converges to $\frac{1}{4}$, and so, by Section 11, $a_n = B_{m,\,n} - E\,B_{m,\,n}$ converges to $B/\sqrt{2}$ in law, and so does

$$a_n' = \frac{2L_{m,\,n}}{\sqrt{n}} - \Big[\sqrt{n}\,\overline{S}_n\left(\frac{\theta}{\sqrt{n}}\right) + 2\,\theta\,\Big\{\overline{S}_n'(c_0)$$

$$- \frac{2}{n}\sum_{i=1}^{n} S_i'(0)\,\{1 - S_i(c_0)\}\Big\}\Big]. \qquad (13.2)$$

Since the derivative of the square bracket is continuous in θ and at $\theta = 0$ equals

$$\overline{S}_n'(0) - 4\,n^{-1}\,\Sigma^n_{i=1}\,S_i'(0)\,\{1 - \overline{S}_i(c_0) + 2\,S_n'(c_c)\},$$

we have that if $\overline{S}_n$ has a continuous nonvanishing derivative throughout a neighbourhood of c_0 and if it exists at a point c_0 satisfying $\overline{S}_n(c_0) = \frac{3}{4}$ and if the limits of $\overline{S}_n'(0)$, $\overline{S}_n'(c_0)$ and

$$\overline{J}_n = n^{-1}\,\Sigma^n_{i=1}\,S_i'(0)\,\{1 - S_i(c_0)\} \qquad (13.3)$$

exist, the asymptotic power of the median test for a fixed θ' near 0 is

$$1 - \Phi\,[\delta - 2\sqrt{2}\,\theta'\sqrt{n}\,\{\tfrac{1}{2}\,\bar{S}_n'(0) - 2\,\bar{J}_n + \bar{S}_n'(c_a)\}]. \qquad (13\cdot4)$$

If the S_i are all equal to each other and to S, $\bar{J}_n = \frac{1}{4}\,S'(0)$ and the asymptotic power is

$$1 - \Phi\,\{\delta - 2\sqrt{2}\,\theta'\,\sqrt{n}\,S'(c_a)\} \qquad (13\cdot4\,a)$$

In the normal case this gives a local asymptotic efficiency of ·81 with respect to the randomization analog of Student's test. In general, however, the efficiency of the median test for symmetry against shift hypotheses has the undesirable property of depending strongly on local properties of the density.

14. Acknowledgements

Besides the acknowledgement expressed in Section 11, the author wishes to record his debt to the Staff of the Mathematical Center of Amsterdam for calling attention to several inaccuracies in the original version of this paper.

References

Andrews, F. C. .. "Asymptotic Behaviour of some Rank Tests for Analysis of Variance," *Annals of Math. Stat.*, 1954, **25**, 724–36.

Behrens, W. U. .. *Mathematische Methoden für Versuchsansteller auf den Gebieten der Naturwissenschaften, Landwirtschaft und Medizin*, Verlag E. Ulmer, Stuttgart, 1933, pp. 137.

Cochran, W. G. .. "The Efficiencies of the Binomial Series Tests of Significance of a Mean and of a Correlation Coefficient," *Journ. Roy. Stat. Soc.*, 1937, **100**, 69–73.

van Dantzig, D. .. "On the Consistency and the Power of Wilcoxons' Two-Sample Test," *Nederl. Akad. Wetensch., Proc.*, Ser. A, 1951, **54**, 3–10, or *Indagationes Math.*, 1951, **13**, 1–8.

Darwin, C. .. *The Effects of Cross and Self Fertilisation in the Vegetable Kingdom*, John Murray, London, 1876, pp. viii and 482.

Dixon, W. J. .. "Power Functions of the Sign Test and Power Efficiency for Normal Alternatives," *Annals of Math. Stat.*, 1953, **24**, 467–73.

——— .. "Power Under Normality of Several Nonparametric Tests," *Ibid.*, 1954, **25**, 610–14.

Dixon, W. J. and Massey, F. J., Jr. *An Introduction to Statistical Analysis*, McGraw-Hill Book Co., New York, 1951, x and 370 pp.

—— and Mood, A. M. "The Statistical Sign Test," *Journ. Am. Stat. Assn.*, 1946, **41**, 557–66.

Doob, J. L. .. "Heuristic Approach to the Kolmogorov-Smirnov Theorems," *Annals of Math. Stat.*, 1949, **20**, 393–403.

Dwass, M. .. "On the Asymptotic Normality of Certain Rank Order Statistics," *Ibid.*, 1953, **24**, 303–306.

Festinger, L. .. "The Significance of Difference Between Means Without Reference to the Frequency Distribution Function," *Psychometrika*, 1946, **11**, 97–105.

Fisher, R. A. .. *The Design of Experiments*, Oliver and Boyd, Edinburgh, 1935, xii and 236 pp.

—— .. "Coefficient of Racial Likeness and the Future of Craniometry," *J. R. Anthrop. Inst.*, 1936, **66**, 57–63.

Fisher, R. A. and Yates, F. *Statistical Tables for Biological, Agricultural and Medical Research*, Oliver and Boyd, Edinburgh, 1938, viii and 90 pp.

Fraser, D. A. S. .. "Non-Parametric Theory: Scale and Location Parameters," *Can. J. Math.*, 1953, **6**, 46–68.

Hemelrijk, J. .. "A Family of Parameter-Free Tests for Symmetry with Respect to a Given Point," *Nederl. Akad. Wetensch., Proc.*, Ser. A, 1950 *a*, **53**, 945–55, 1186–1198; or *Indagationes Math.*, **11**, 340–50, 419–31.

—— .. *Symmetrietoetsen en Andere Tœpassingen van de Theorie van Neyman en Pearson*, diss. Amsterdam, Excelsior, The Hague, 1950 *b*, 91 and 4 pp.

Hodges, J., Jr. .. "Galton's Rank-Order Test," *Biometrika*, 1955, **42**, 261–62.

—— and Lehmann, E. L. "Matching in Paired Comparisons," *Annals of Math. Stat.*, 1954, **25**, 787–91.

—— .. "The Efficiency of Some Nonparametric Competitors of the *t*-Test," *Ibid.*, 1956, **27**, 324–35.

Hoeffding, W. .. "A Class of Statistics with Asymptotically Normal Distributions," *Ibid.*, 1948, **19**, 293–325.

—— .. "'Optimum' Nonparametric Tests," *Proc. Second Berkeley Symposium on Mathematical Statistics and Probability*, Univ. of California Press, Berkeley and Los Angeles, 1951, pp. 83–92.

—— .. "The Large-Sample Power of Tests Based on Permutations of Observations," *Annals of Math. Stat.*, 1952, **23**, 169–92.

Konijn, H. S. .. "On the Power of Certain Tests for Independnece in Bivariate Populations," *Ibid.*, 1956, **27**, 300–23.

Lehmann, E. L. .. "Consistency and Unbiasedness of Certain Nonparametric Tests," *Ibid.*, 1951, **22**, 165–79.

——— .. "The Power of Rank Tests," *Ibid.*, 1953, **24**, 23–43.

—— and Scheffé .. "Completeness, Similar Regions, and Unbiased Estimation," Part I, *Sankhyā*, 1950, **10**, 305–40.

—— and Stein, C. .. "On the Theory of Some Nonparametric Hypotheses," *Annals of Math. Stat.*, 1949, **20**, 28–45.

Mann, H. B. and Whitney, D. R. .. "On a Test of Whether One of Two Random Variables is Stochastically Larger than the Other," *Ibid.*, 1947, **18**, 50–60.

Marshall, A. W. .. "A Large-Sample Test of the Hypothesis that One of Two Random Variables is Stochastically Larger than the Other," *Journ. Am. Stat. Assn.*, 1951, **46**, 366–74.

Mathisen, H. C. .. "A Method of Testing the Hypothesis that Two Samples are from the Same Population," *Annals of Math. Stat.*, 1943, **14**, 188–94.

Mood, A. M. .. "On the Asymptotic Efficiency of Certain Nonparametric Two-Sample Tests," *Ibid.*, 1954, **25**, 514–22.

Pitman, E. J. G. .. "Significance Tests Which May Be Applied to Samples from Any Populations," *Journ. Roy. Stat. Soc. Suppl.*, 1937, **4**, 117–30.

Putter, J. .. "The Treatment of Ties in Some Nonparametric Tests," *Annals of Math. Stat.*, 1955, **26**, 368–86.

Ruist, E. .. "Comparison of Tests for Nonparametric Hypotheses," *Arkiv f. Matematik*, 1954, **3**, 133–63.

Smirnov, N. V. .. "On the Estimation of the Discrepancy between Empirical Curves of Distribution for Two Independent Samples," *Bull. Math. Univ. Moscou*, Série Int., 1939, **2**, No. 2, 16 pp.

——— .. "Sur un critère de symétrie de la loi de distribution d'une variable aléatoire," *C.R. (Doklady), Acad. Sci., URSS (N.S.)*, 1947, **56**, 11–14.

Sundrum, R. M. .. "The Power of Wilcoxon's Two-Sample Test," *Jour. Roy. Stat. Soc.*, Ser. B, 1953, **15**, 246–52.

Terry, M. E. .. "Some Rank Order Tests which are most Powerful against Specific Parametric Alternatives," *Annals of Math. Stat.*, 1952, **23**, 346–66.

Thompson, W. R. .. "Biological Application of Normal Range and Associated Significance Tests in Ignorance of Original Distribution Forms," *Ibid.*, 1938, **9**, 281–87.

van der Waerden, B. L. .. "Order Tests for the Two-Sample Problem and Their Power," *Nederl. Akad. Wetensch., Proc.*, Ser. A, 1952, **55**, 453–58, or *Indagationes Math.*, **14**.

van der Waerden, B. L. "Order Tests for the Two-Sample Problem (Second Communication)," *Nederl. Akad. Wetensch. Proc.*, Ser. A, 1953 *a*, **56**, 303–10, or *Indagationes, Math.*, **15**.

——— .. "Order Tests for the Two-Sample Problem (Third Communication)," *Ibid.*, 1953 *b*, **56**, 311–16, or *Ibid.*, **15**.

——— .. "Ein neuer Test für das Problem der zwei Stichproben," *Math. Annalen*, 1953 *c*, **126**, 93–107.

Walsh, J. E. .. "Some Significance Tests for the Median which are Valid Under Very General Conditions," *Annals of Math. Stat.*, 1949, **20**, 64–81.

Westenberg, J. .. "Significance Tests for Median and Interquartile Range in Samples from Continuous Populations of any Form," *Nederl. Akad. Wetensch.*, *Proc.*, Ser A, 1948, **51**, 252–61.

——— .. "A Tabulation of the Median Test for Unequal Samples," *Ibid.*, 1950, **53**, 77–82, or *Indagationes Math.*, **12**, 8–13.

——— .. "A Tabulation of the Median Test with Comments and Corrections to Previous Papers," *Ibid.*, 1952, **55**, or *Ibid.*, **14**, 10–15.

Wilcoxon, F. .. "Individual Comparisons by Ranking Methods," *Biometrics Bull.*, 1945, **1**, 80–83.

——— .. "Probability Tables for Individual Comparisons by Ranking Methods," *Biometrics*, 1947, **3**, 119–22.