TOPICS IN RANK-ORDER STATISTICS

RALPH A. BRADLEY
FLORIDA STATE UNIVERSITY

This paper is dedicated to the memory of Dr. Frank Wilcoxon.

1. Introduction

Frank Wilcoxon joined the faculty of the Florida State University in 1960. He brought with him a number of ideas for further research in rank-order statistics and proceeded to develop them in association with colleagues and graduate students. This paper is largely expository and reports research based on his suggestions.

Two major topics are presented: sequential, two-sample, rank tests and multivariate, two-sample, rank procedures.

2. Sequential two-sample rank tests

2.1 *Preliminary remarks.* The two-sample, rank-sum test was introduced by Wilcoxon [29], [30]. Two populations, X- and Y-populations, are given with distribution functions,

$$P(X \leq u) = F(u), \qquad P(Y \leq u) = G(u), \tag{2.1}$$

X and Y being the random variables associated with the two populations. The basic null hypothesis tested is that

$$H_0\colon G(u) \equiv F(u), \tag{2.2}$$

usually with the assumed alternative of location change, $G(u - \theta) \equiv F(u)$.

Samples of independent observations of sizes m and n from X- and Y-populations respectively are taken and ranked in joint array. The sum of ranks, T for the X-sample or S for the Y-sample, is taken as the test statistic, and departures of the statistic from its mean under H_0, $\frac{1}{2}m(m + n + 1)$ or $\frac{1}{2}n(m + n + 1)$, are judged for significance. In order that ties in ranks between X- and Y-observations occur with probability zero, one may restrict F and G to be continuous.

Small-sample tables are available as are approximate, large-sample distributions for the rank sum under H_0. A recent extensive set of tables was developed by Wilcoxon, Katti, and Wilcox [31]. This table is divided into four sections corresponding to four levels of significance, 0.05, 0.025, 0.01, and 0.005 for a

This paper is based in part on research supported by the Army, Navy, and Air Force under an Office of Naval Research Contract, NONR-988(08), NR 042-004. Reproduction in whole or in part is permitted for any purpose of the United States Government.

one-sided test, and exact probabilities are given to four decimal places for the rank totals which bracket these significance levels. The critical rank totals are tabulated for sample sizes, $m, n = 3(1)50$.

Much has been written on the asymptotic properties of ranking procedures for large sample sizes. Properties of these procedures for small samples are difficult to investigate and depend on the forms of distribution functions like F and G, although empirical methods have been the bases for a limited number of such studies. Lehmann [15] suggested a class of alternatives to H_0,

$$H_a\colon G(u) \equiv F^k(u), \qquad k > 0. \tag{2.3}$$

For this class of alternatives, the small-sample power of the rank-sum test may be evaluated with the power dependent upon k but free of the form of $F(u)$. The basic sequential methods of Wald [27] and the class of alternatives of Lehmann permitted the development of two sequential, two-sample, grouped rank tests.

2.2 *Basic procedures.* Define a basic group of observations to consist of m X-observations and n Y-observations with ranking effected within the group in joint array as for the original Wilcoxon method. The sequential aspect of the experimentation consists of deciding at the end of each group of observations whether to continue experimentation by taking an additional group of observations or to terminate experimentation with a decision to accept the null hypothesis or with a decision to accept an alternative hypothesis. In this paper we restrict attention to H_0 of (2.2) and take the specific, one-sided, alternative among those of (2.3) to be

$$H_1\colon G(u) \equiv F^{k_1}(u), \qquad k_1 > 1. \tag{2.4}$$

This is the basic sequential system in [32]; the two-sided procedure is developed briefly in [33]. More generally, it is possible to consider a null hypothesis $H_0^*\colon G(u) \equiv F^{k_0}(u)$ and an alternative hypothesis

$$H_1^*\colon G(u) \equiv F^{k_1}(u), \qquad k_1 > k_0 > 0. \tag{2.5}$$

Consider the γ-th group of observations. Let $r_{1,\gamma}, \cdots, r_{m,\gamma}$ and $s_{1,\gamma}, \cdots, s_{n,\gamma}$ be the ranks assigned to X- and Y-observations respectively. The relevant information is retained if we consider only the Y-ranks and take $s_{1,\gamma} < \cdots < s_{n,\gamma}$. From [15] or [32] the probability that the Y-sample is given the indicated ranks under (2.3) is

$$P(s_{1,\gamma}, \cdots, s_{n,\gamma}|m, n, k) = \frac{k^n}{\binom{m+n}{n}} \prod_{j=1}^{n} \frac{\Gamma(s_{j,\gamma} + jk - j)\Gamma(s_{j+1,\gamma})}{\Gamma(s_{j+1,\gamma} + jk - j)\Gamma(s_{j,\gamma})} \tag{2.6}$$

where $s_{n+1,\gamma}$ is taken as $(m + n + 1)$.

For a sequential rank test of H_0 versus H_1 based on the actual configuration of ranks, the probability ratio for the γ-th group is

$$r_\gamma(m, n, k_1, 1) = \frac{k_1^n(m+n)!}{(s_{1,\gamma} - 1)!} \prod_{j=1}^{n} \frac{\Gamma(s_{j,\gamma} + jk_1 - j)}{\Gamma(s_{j+1,\gamma} + jk_1 - j)}, \tag{2.7}$$

the ratio of two probabilities of the form (2.6) with $k = k_1$ in the numerator probability and $k = 1$ in the denominator probability. If experimentation has proceeded to the end of t groups, the probability ratio for the complete experiment to that stage is, from an assumption of independence of groups of observations,

$$p_{1t}/p_{0t} = \prod_{\gamma=1}^{t} r_\gamma(m, n, k_1, 1), \tag{2.8}$$

the statistic required for the Wald sequential analysis. The procedure based on (2.8) has been called the configural rank test.

For a sequential rank test of H_0 versus H_1 based on rank sums, let

$$S_\gamma = \sum_{j=1}^{n} s_{j,\gamma} \quad \text{and} \quad T_\gamma = \sum_{i=1}^{m} r_{i,\gamma}. \tag{2.9}$$

The probability of S_γ may be formally expressed as

$$P\left(\sum_{j=1}^{n} s_{j,\gamma} = S_\gamma | m, n, k\right) = \sum P(s_{1,\gamma}, \cdots, s_{n,\gamma} | m, n, k) \tag{2.10}$$

where the summation in (2.10) is over the limits,

$$1 \leq s_{1,\gamma} < \cdots < s_{n,\gamma} \leq m + n, \qquad \sum_{j=1}^{n} s_{j,\gamma} = S_\gamma. \tag{2.11}$$

The γ-th group probability ratio corresponding to (2.7) is

$$R_\gamma(m, n, k_1, 1) = P\left(\sum_{j=1}^{n} s_{j,\gamma} = S_\gamma | m, n, k_1\right) \Big/ P\left(\sum_{j=1}^{n} s_{j,\gamma} = S_\gamma | m, n, 1\right), \tag{2.12}$$

and that corresponding to (2.8) is

$$P_{1t}/P_{0t} = \prod_{\gamma=1}^{t} R_\gamma(m, n, k_1, 1). \tag{2.13}$$

The procedure based on (2.13) has been called the rank-sum test.

In sequential analysis, α and β, the probabilities of Type I and Type II errors respectively, are used to form the constants,

$$A = (1 - \beta)/\alpha \quad \text{and} \quad B = \beta/(1 - \alpha). \tag{2.14}$$

Let P_{1t}/P_{0t} be the generic probability ratio at stage t. Then the decision process in logarithmic form is as follows.

(i) If $\ln B < \ln (P_{1t}/P_{0t}) < \ln A$, take another observation (another group of observations).

(ii) If $\ln (P_{1t}/P_{0t}) \leq \ln B$, terminate experimentation and accept H_0.

(iii) If $\ln (P_{1t}/P_{0t}) \geq \ln A$, terminate experimentation and accept H_1.

Substitution of (2.8) or (2.13) for P_{1t}/P_{0t} is made for the sequential rank tests.

Insight into the interpretation of the Lehmann model and the specification of k_1 may be obtained from the properties of the model. First, note that if X and Y are randomly selected from their respective populations,

$$p = P(X \leq Y) = k/(k + 1), \tag{2.15}$$

and conversely, $k = p/(1 - p)$. Second, when $G(u) \equiv F^k(u)$, $k > 1$, $G(u)$ is skewed to the right relative to $F(u)$, and if $F(u)$ is a standard normal distribution function, the mean of the Y-population is $\mu_y > 0$ and the variance is $\sigma_y^2 < 1$. Thus, associated with a value of k, we may consider also values of p and μ_y as defined. It is noted also that if k is an integer, $G(u)$ is the distribution function of the largest of k independent observations on X.

To facilitate use of the sequential, rank-sum test, tables [32] have been prepared for $m = n = 1(1)9$ giving values of T_γ, S_γ and corresponding values of

$$P_\gamma = P\left(\sum_{j=1}^{n} s_{j,\gamma} = S_\gamma | m, n, k_1\right) \tag{2.16}$$

defined in (2.10) and $\ln R_\gamma$ with R_γ defined in (2.12). The values of P_γ and $\ln R_\gamma$ are computed for $k_1 = 1.5$, 2.33, 4, and 9 with associated $p_1 = 0.6(0.1)0.9$ and $\mu_y = 0.282$, 0.658, 1.029, and 1.485. To facilitate use of the sequential, configural rank test, Wilcoxon has devised an ingenious algorithm for the computation of (2.6); its use is illustrated in [32]. The method consists of setting forth the ordered array of m X's and n Y's for the group and placing unity under each X and k_1 under each Y. Then cumulative totals from the left are obtained and the probability of (2.6) results from division of $k_1^n m!n!$ by the product of these cumulative totals.

2.3 *Examples.* The sequential rank tests were illustrated [32] for a screening experiment on chemical compounds for possible amelioration of the harmful effects of radiation. At each stage of the screening process ten laboratory animals were chosen, exposed to equal doses of radiation, divided randomly into equal size Control (X-sample) and Experimental (Y-sample) samples with the Experimental sample being subjected to injection of the chemical compound under study. The example is summarized in table II.1; although survival times are given in [32], only ranks representing orders of death are given here.

TABLE II.1

RANKS FOR CONTROL AND EXPERIMENTAL SAMPLES BY GROUPS IN A SCREENING EXPERIMENT TOGETHER WITH TEST STATISTICS

Group	X- and Y-ranks	T_γ	S_γ	$\ln r_\gamma$	$\ln (p_{1t}/p_{0t})$	$\ln R_\gamma$	$\ln (P_{1t}/P_{0t})$
1	1, 2, 3, 5, 8	19					
	4, 6, 7, 9, 10		36	1.511	1.511	1.403	1.403
2	1, 2, 5, 9, 10	27					
	3, 4, 6, 7, 8		28	0.182	1.693	−0.498	0.905
3	1, 2, 3, 7, 9	22					
	4, 5, 6, 8, 10		23	1.070	2.763*	0.669	1.574
4	1, 2, 3, 6, 7	19					
	4, 5, 8, 9, 10		36	—	—	1.403	2.977*

* The sequential process terminates with acceptance of H_1.

For this illustration, $m = n = 5$, $k_1 = 2.33$ ($p_1 = 0.7$), $\alpha = 0.15$, $\beta = 0.05$, $\ln A = 1.85$, and $\ln B = -2.83$. Values of $\ln r_\gamma$, $\ln (p_{1t}/p_{0t})$, $\ln R_\gamma$, and ln

(P_{1t}/P_{0t}) are shown in table II.1. The Wilcoxon algorithm was used to compute $\ln r_\gamma$ and thence $\ln (p_{1t}/p_{0t})$, and the tables noted were used to record $\ln R_\gamma$ and thence $\ln (P_{1t}/P_{0t})$. Both the sequential, configural rank test and the sequential, rank-sum test led to rejection of H_0 and acceptance of H_1, the former with one less group of observations than the latter. Termination is judged by comparison of $\ln (p_{1t}/p_{0t})$ or $\ln (P_{1t}/P_{0t})$ with $\ln A$ and $\ln B$ as explained in the preceding subsection.

Numerous other possible applications exist in the medical-biological area and, indeed, many of these necessitate the use of within-group ranking. A second example is given in [33] and is based on visual ordering of severity of ulceration in rats.

The sequential ranking methods may be particularly useful in research wherein measurement is difficult but subjective ordering within groups of limited size is possible. Sequential triangle and duo-trio tests for the selection of expert taste panels have been discussed [2]; Kramer [13], [14] has extended the matching process to more samples. Sequential rank methods could be used for judge selection for tests where X- and Y-samples differ in and are to be ordered by basic taste characteristics.

Other applications may be noted in life testing. A specific case noted by the author involved a testing machine with a rotating cam which flexed six small springs under test. If two sets of three springs each, the sets differing in metal composition, were randomly allocated to test-machine positions, an appropriate sequential experiment could be devised and depend on order of failure of the springs.

2.4 *Properties.* Wald has provided means of evaluation of the average sample number (ASN) function and the power function of sequential tests at special values of the parameter under test. These formulas have been applied to the sequential rank tests [32] but do not adequately characterize the functions.

Monte Carlo studies were undertaken [3] to further evaluate properties of the sequential tests. These studies included empirical calculation of points on the ASN and power functions for the sequential rank-sum test with $\alpha = \beta = .05$, the calculations for each set of design parameters being based on 500 simulated sequential experiments. The studies were made for data generated in accordance with the Lehmann model and also for data generated from normal populations with unit variance and with means differing by μ_y corresponding to the parallel value of k in the Lehmann model.

It was found that the power function $\Phi(k, \mu_y)$ could be adequately represented by the probit model,

$$\Phi(k, \mu_y) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{a+b\mu_y-5} e^{-w^2/2}\, dw. \tag{2.17}$$

Estimates of a and b for the various designs are given in table II.2; k and μ_y are related as noted above and as tabled in [32]. In general, it was found tha α and β were somewhat less than the nominal values of the designs. Note that $\hat{a}$ tends to

TABLE II.2

ESTIMATED PARAMETER VALUES FOR FITTED POWER FUNCTIONS FOR THE SEQUENTIAL RANK-SUM TEST

(Main entries are for the Lehmann model; values in parentheses are for the normal model.)

Design		$m = n$ 2	3	4	5
$k_1 = 2.33$	$\hat{a}$	3.89 (3.68)	3.44 (3.31)	3.16 (3.99)	3.51 (3.86)
$\alpha = \beta = .05$	$\hat{b}$	4.34 (4.25)	5.01 (4.80)	5.59 (3.79)	4.93 (3.84)
$k_1 = 4$	$\hat{a}$	2.59 (2.98)	3.16 (3.10)	2.63 (2.81)	2.71 (3.08)
$\alpha = \beta = .05$	$\hat{b}$	4.30 (3.43)	3.60 (3.43)	4.27 (3.70)	4.17 (3.40)
$k_1 = 9$	$\hat{a}$	2.83 (3.05)	2.56 (2.65)	2.40 (2.69)	2.42 (2.47)
$\alpha = \beta = .05$	$\hat{b}$	2.87 (2.30)	3.11 (2.69)	3.42 (2.75)	3.36 (2.97)

be larger for the normal model than for the Lehmann model, and $\hat{b}$ tends to be smaller indicating a larger Type I error and lower power for positive μ_y or k in excess of unity; $\hat{a}$ and $\hat{b}$ are estimates of a and b respectively in (2.17).

In table II.3, some typical values of the ASN functions are given for the sequential, rank-sum test with data generated from the Lehmann model and from the normal model. This table also contains information on the ASN function for a modified configural rank test to be discussed in the next subsection. Note that ASN values tend to be slightly higher for the studies on the normal model but are not excessively so.

The Monte Carlo studies suggest an element of robustness for the sequential rank-sum test and give confidence in its use when the Lehmann model may not be entirely appropriate.

Additional information from the Monte Carlo studies is given in [3]. Effects of truncation were considered as were the distributions of termination numbers.

2.5 *Modified sequential rank tests.* It appears intuitively that better sequential rank tests might be obtained when it is feasible to effect complete re-ranking of the totality of X- and Y-observations at each stage of the sequential process. Such a procedure has at least theoretical interest.

Suppose that X- and Y-observations are still taken in groups of m and n and that no group or block effects are present. Then, at the t-th stage of such a process, mt X-observations and nt Y-observations are ranked in joint array. If the Lehmann model is used again, a modified, configural rank test might be based on the probability ratio,

$$p^*_{1t}/p^*_{0t} = r(mt, nt, k_1, 1) \tag{2.18}$$

with r defined in (2.7). Similarly, a modified, rank-sum test might be based on

$$P^*_{1t}/P^*_{0t} = R(mt, nt, k_1, 1) \tag{2.19}$$

with R as in (2.12). Use of Wald bounds (2.14) may be tried. The case with

TABLE II.3

AVERAGE SAMPLE NUMBERS OF THE MODIFIED CONFIGURAL RANK TEST IN COMPARISON WITH THE GROUPED RANK-SUM TESTS

(Main entries for grouped rank-sum tests are for the Lehmann model; values in parentheses are for the normal model.)

Design	k	μ_y	ASN*: Modified Configural Rank Test	ASN*: Grouped Rank-Sum Tests $m = n = 2$	$m = n = 3$	$m = n = 4$	$m = n = 5$
$k_1 = 1.5$	1	0	54.93**	116.20	110.00	110.22	112.52
$\alpha = \beta = .05$	3	0.846	21.93	31.96 (35.31)	30.62 (35.21)	30.30 (34.78)	29.83 (33.05)
	5	1.163	16.66	22.84 (25.45)	21.61 (25.46)	21.84 (24.87)	21.66 (24.82)
$k_1 = 2.33$	1	0	19.33	29.16	30.47	29.39	28.61
$\alpha = \beta = .05$	3	0.846	14.78	21.36 (26.20)	20.68 (26.03)	21.15 (23.73)	20.71 (24.93)
	5	1.163	10.12	13.71 (16.86)	14.27 (16.82)	13.60 —	14.20 (16.87)
$k_1 = 4$	1	0	8.78	12.46	13.62	14.23	13.75
$\alpha = \beta = .05$	3	0.846	13.09	18.89 (20.74)	18.44 (20.78)	19.90 (23.93)	19.66 (23.24)
	5	1.163	8.64	11.68 (13.97)	11.37 (13.70)	12.58 (14.22)	12.24 (15.20)
	7	1.352	6.98	—	9.55 —	10.18 (12.46)	10.02 (12.82)
$k_1 = 9$	1	0	4.53	6.65	6.82	7.65	7.89
$\alpha = \beta = .05$	3	0.846	9.07	12.65 (12.68)	13.84 (15.20)	14.85 (15.92)	16.08 (15.41)
	5	1.163	7.29	9.54 (11.22)	11.60 (13.09)	10.81 (13.20)	12.07 (14.99)
	7	1.352	6.05	7.95 (9.22)	9.31 (11.54)	9.40 (10.78)	10.36 (12.89)
	9	1.485	5.65	7.05 (8.70)	8.45 (10.15)	7.70 (10.11)	8.61 (11.13)

* Average Sample Numbers are average numbers of observations from each population.
** Based on 30 simulated experiments only; all remaining entries based on 500 simulated experiments.

$m = n = 1$ assumes more interest now and is the case considered in the Monte Carlo studies noted below.

Savage and Savage [22] have demonstrated that the bounds (2.14) are appropriate for the modified, configural rank test but have not so demonstrated for the modified, rank-sum test. In the former case, they have shown the properties of finite termination and finite expected termination under H_0 and H_1. Savage and Sethuraman [23] are developing stronger termination results following somewhat the approach of Jackson and Bradley [12]. Hall, Wijsman, and Ghosh [11] discuss the problem as does Berk [1]. Further theoretical work is required to obtain information on ASN functions and power functions when the basic assumption of Wald's methods are not met. Much more extensive tables of $\ln R$ would be required to facilitate use of a modified, rank-sum test.

Limited Monte Carlo studies have been conducted under the Lehmann model for the modified, configural rank test, and values of the ASN functions are shown in table II.3. Values of the power function are very close to those for the grouped, sequential rank tests. The comparison in table II.3 is somewhat confounded in that the modified, configural rank test with $m = n = 1$ is compared with the grouped, rank-sum test. Appreciable reductions in ASN values are shown.

More complete discussion of the modified, sequential, rank tests is given in [4].

2.6 *Other research.* Milton [17] has computed probabilities of possible rank configurations for $1 \leq n \leq m \leq 7$ and $n = 1$, $m = 8(1)12$ for $F(u)$ and $G(u)$ normal with unit variances and differences in means, $\mu_y = 0(0.2)1, 1.5, 2, 3$. He has used these tables for various nonparametric power and efficiency computations and comparisons [18] and also to develop grouped, rank-sum and configural rank tests of the normal shift hypothesis [19].

For the sequential rank tests, Milton has tabled values of the group probability ratios comparable to (2.7) and (2.12). He has used Wald's formulas to evaluate ASN and power functions at the selected points for which such formulas are available. It is difficult to compare his results with the Monte Carlo studies [3], [4] because his values of μ_y under the alternative hypothesis do not match those corresponding with our values of k_1 or p_1 very well. Limited comparisons suggest that his ASN values are slightly lower than those of our Monte Carlo studies, but it appears that the Wald formulas may underestimate ASN values, at least for the Lehmann model (see [3] table 4).

Parent [20] has defined *sequential ranks* and applied the concept to the two-sample problem discussed above and also to the paired-sample problem leading to a type of sequential, signed rank procedure. The sequential rank of an observation X_t relative to the set $X_1, \cdots, X_t$ is k if X_t is the k-th smallest observation of the set; $X_1, \cdots, X_{t-1}$ are not given new ranks at stage t, and several of these X's could have received rank k also as they were observed in sequence. It is pointed out that the sequential ranks for $X_1, \cdots, X_t$ are uniquely determined, and moreover, that they uniquely determine the ordering of $X_1, \cdots, X_t$.

The use of sequential ranks provides some simplifications in modified sequential rank tests discussed above. However, sequential ranks are not independent for the Lehmann model unless $k = 1$, and they do not avoid the requirement for stronger basic theory of sequential analysis than that provided by Wald.

Signed sequential ranks are discussed by Parent also. Independence of the signed sequential ranks is demonstrated when the observations giving rise to them are independent and equally distributed from a population with cdf $F(u)$ satisfying the "symmetry" relationship,

$$F(-u) = F(0)[1 - F(u) + F(-u)], \qquad u > 0. \tag{2.20}$$

This condition is met by distributions of positive, negative, and symmetric-about-zero random variables, but for random variables taking both positive and negative values with median different from zero, the condition is rather restrictive, ruling out many common distributions. A sequential, signed rank test analogous to Wilcoxon's procedure is not developed; rather, a procedure to detect a change in distribution from $F(u)$ to some $G(u)$ at some stage in a sequence of observations is developed and applied to process control.

3. Multivariate two-sample rank tests

3.1 *The multivariate problem.* Consider two, p-variate populations with associated, column-vector variates, X and Y. Let $F(u)$ and $G(u)$ be the distribution functions as in (2.1), X and u now being vectors. The null hypothesis is expressed again as in (2.2), and alternatives specifying location change only are usually the ones of interest. Samples, $x_1, \cdots, x_m$ and $y_1, \cdots, y_n$ of independent, column-vector observations from X- and Y-populations respectively are taken.

The problem considered is basically the two-sample form of Hotelling's problem with the generalized Student ratio when $F(u)$ and $G(u)$ are multivariate normal with common dispersion matrix Σ and, under H_0, identical mean vectors. The well-known statistic used then is

$$T^2 = \frac{mn}{m+n} (\bar{x} - \bar{y})' S^{-1} (\bar{x} - \bar{y}) \tag{3.1}$$

where

$$\bar{x} = \frac{1}{m} \sum_{\alpha=1}^{m} x_\alpha, \qquad \bar{y} = \frac{1}{n} \sum_{\beta=1}^{n} y_\beta \tag{3.2}$$

and

$$S = \left[\sum_{\alpha=1}^{m} x_\alpha x'_\alpha - m\bar{x}\bar{x}' \sum_{\beta=1}^{n} y_\beta y'_\beta - n\bar{y}\bar{y}' \right] / (m + n - 2). \tag{3.3}$$

Given the multivariate normal populations, it is known that $(m + n - p - 1)T^2/(m + n - 2)p$ has the variance-ratio distribution with p and $m + n - p - 1$ degrees of freedom while, asymptotically with m and n as they become large in constant ratio, T^2 has the chi-square distribution with p degrees of freedom. These distributions are central under H_0 but noncentral under the location change alternative with noncentrality parameter,

$$\lambda^2 = \frac{mn}{m+n}\mu'\Sigma^{-1}\mu \tag{3.4}$$

where $\mu = \mu_x - \mu_y$, μ_x and μ_y, the mean vectors of X- and Y-populations. When the multivariate normal assumption is removed, little is known about the small-sample distribution of T^2 and nonparametric methods may be needed.

3.2 *Nonparametric procedures.* Wald and Wolfowitz [28] seem to have been first to consider multivariate, two-sample, randomization tests. (They considered the univariate case in some detail and indicated that extensions to the multivariate case were straightforward.) A modified statistic, proportional to

$$T_M^2 = \frac{mn}{m+n}(\bar{x}-\bar{y})'S_M^{-1}(\bar{x}-\bar{y}), \tag{3.5}$$

was used, where

(3.6)

$$S_M = \left[\sum_{\alpha=1}^{m} x_\alpha x_\alpha' + \sum_{\beta=1}^{n} y_\beta y_\beta' - (m+n)^{-1}(m\bar{x}+n\bar{y})(m\bar{x}+n\bar{y})'\right]/(m+n-1).$$

The statistic T_M^2 is monotonically related to T^2 through the relationship,

$$T_M^2 = (m+n-1)T^2/[(m+n-2)+T^2]. \tag{3.7}$$

The randomization test of H_0 is conditional on the numerical values of the observation vectors. Let $z_1, \cdots, z_{m+n}$ constitute the complete set of vectors, $x_1, \cdots, x_m, y_1, \cdots, y_n$. Under H_0, the designation of a z-vector as an X- or Y-vector is taken as a matter of random labeling; each of the $\binom{m+n}{m}$ possible distinct assignments of m X-labels and n Y-labels is taken to be equally likely. For each labeling, T_M^2 is evaluated, and one of these values is the observed one, say, $T^2_{M,\text{obs.}}$. Let a test with significance level α be desired. Let η be the number of values of $T_M^2 \geq T^2_{M,\text{obs.}}$. If $\eta \leq \alpha\binom{m+n}{m}$, the observed value $T^2_{M,\text{obs.}}$ is taken to be significant and H_0 is rejected. Since S_M is invariant under the random labeling and S is not, T_M^2 is used for the test because it is considerably easier to compute or study than T^2. The test based on T_M^2 is equivalent to the similar test based on T^2 because of (3.7). It is seen that the randomization test in the multivariate case follows the same principles as in the univariate case.

Minor simplifications may be made. The constant multiplier in (3.5) may be dropped, and we may replace $(\bar{x}-\bar{y})$ by $(\sum_{\alpha=1}^{m} x_\alpha - m\bar{z})$ where $\bar{z} = \sum_{\gamma=1}^{m+n} z_\gamma/(m+n)$ and obtain a new statistic monotonically related to T^2 and T_M^2. Since $\bar{z}$ is fixed for given observation vectors, the latter substitution yields a statistic for which only the vector $\sum_{\alpha=1}^{m} x_\alpha$ changes from one labeling to the next. Even so, the randomization test is numerically difficult and tabling is not possible.

Wald and Wolfowitz show that the limiting distribution of T_M^2 is the central chi-square distribution with p degrees of freedom for the randomization test subject to mild restrictions on the sequence of vectors of real constants, $z_1, \cdots,$

z_{m+n}. It follows directly that T^2 has the same limiting distribution under H_0 for the randomization test. Thus, for moderate sizes of m and n, one might assume that use of the limiting distribution is adequate for applications as an approximation to the randomization test.

Bradley and Patel [5] in work in progress have considered moments of T^2_M over the randomization distribution. We note only the first two moments here:

$$E(T^2_M) = p,$$

$$E(T^4_M) = \left(\frac{c_1}{c_0^2}\right) p(p+2) + \left(\frac{m+n}{mn}\right)^2 (c_0 - 6c_1) \sum_{\gamma=1}^{m+n} \lambda^2_{\gamma\gamma} \tag{3.8}$$

$$\approx p(p+2) \qquad \text{for } m, n \text{ large,}$$

where

$$c_0 = mn/(m+n)(m+n-1),$$

$$c_1 = m(m-1)n(n-1)/(m+n)(m+n-1)(m+n-2)(m+n-3), \tag{3.9}$$

$$\lambda_{\gamma\gamma} = (z_\gamma - \bar{z})' S_M^{-1} (z_\gamma - \bar{z}).$$

A basis for an approximation to the randomization distribution of T^2_M is to fit a continuous density function of appropriate type to it. The statistic $T^2/(m+n-1)$ has the beta distribution on $(0, 1)$ with parameters $\frac{1}{2}p$ and $\frac{1}{2}(m+n-p-1)$ under normal theory. Suppose a beta distribution is the appropriate type and determine its unknown parameters $\frac{1}{2}\nu_1$ and $\frac{1}{2}\nu_2$ by the method of moments, two moments of $T^2_M/(m+n-1)$ being available from (3.8). Then

$$\nu_1 = \phi p, \qquad \nu_2 = \phi(m+n-p-1) \tag{3.10}$$

with

$$\phi = \frac{2}{m+n-1}\left[\left\{\frac{p(m+n-p-1)}{v}\right\} - 1\right] \tag{3.11}$$

where

$$v = \operatorname{var}(T^2_M) = E(T^4_M) - [E(T^2_M)]^2 \tag{3.12}$$

from (3.8). The corresponding approximation to the randomization distribution of $(m+n-p-1)T^2/(m+n-2)p$ is the variance-ratio distribution with ν_1 and ν_2 degrees of freedom as computed from (3.10). Note that ϕ approaches unity with large m, n since $v \approx 2p$.

In the univariate problem, it is possible to replace the original observations with ranks. Then the randomization distribution depends only on m and n, and tables are available [31] as we have seen. Use of ranks in the multivariate problem leads to only slight simplifications. Suppose that observations on each variate are ranked separately as in the univariate problem, and let $r_1, \cdots, r_m, s_1, \cdots, s_n$ be the resulting vectors of ranks. The computation of S_M is simplified as all diagonal elements are known and equal, but the nondiagonal elements are proportional to the various rank correlations of the data. In different problems, for given m and n, different arrays of rank correlations will arise and tabling of the distribution of T_M^2 is not feasible. Bradley and Patel have considered the use of ranks as well as normal scores and have developed some large-sample theory associated with them.

Following the paper by Wald and Wolfowitz, much new theory of nonparametric and rank-order statistics was developed, but there was an interval wherein little more was done on the multivariate problem, an exception being the work of Lynch and Freund [16]. Recently, there has been more activity. In addition to [5], Chatterjee and Sen [6] discuss the bivariate problem with use of ranks and give some consideration to the nonnull distribution of T_M^2. Sen [24], [25], Sen and Govindarajulu [26], Govindarajulu [10], and Chatterjee and Sen [7] provide more general results including two-sample, multivariate problems, C-sample, multivariate problems, and limit theory. Robson [21] proposes a distance method with application to ecology.

Wilcoxon had a long-term interest in the multivariate generalization of the rank-sum test. He was seeking a procedure of relative simplicity and proposed two bivariate methods with that characteristic. Neither of these methods has an adequate theoretical base, but both have intuitive appeal. They are presented here in order to record his ideas and perhaps to stimulate further consideration of them.

3.3 *Wilcoxon's first bivariate method.* Consider m bivariate X-observations and n bivariate Y-observations with corresponding rank vectors $(r_{11}, r_{21}), \cdots, (r_{1m}, r_{2m}), (s_{11}, s_{21}), \cdots, (s_{1n}, s_{2n})$. Let the sample mean vectors be $(\bar{r}_1, \bar{r}_2)$ and $(\bar{s}_1, \bar{s}_2)$, and let r be the pooled correlation coefficient calculated from the ranks,

(3.13)
$$r = \frac{\left[\sum_{\alpha=1}^{m} (r_{1\alpha} - \bar{r}_1)(r_{2\alpha} - \bar{r}_2) + \sum_{\beta=1}^{n} (s_{1\beta} - \bar{s}_1)(s_{2\beta} - \bar{s}_2)\right]}{\left\{\left[\sum_{\alpha=1}^{m} (r_{1\alpha} - \bar{r}_1)^2 + \sum_{\beta=1}^{n} (s_{1\beta} - \bar{s}_1)^2\right]\left[\sum_{\alpha=1}^{m} (r_{2\alpha} - \bar{r}_2)^2 + \sum_{\beta=1}^{n} (s_{2\beta} - \bar{s}_2)^2\right]\right\}^{1/2}}.$$

Wilcoxon computed Fisher's discriminant function from the ranks, the linear function that has the greatest variance between samples relative to the variance within samples. A quantity proportional to this discriminant function for an arbitrary point (t_1, t_2) is

(3.14)
$$z = t_1 + t_2 \tan\theta$$

where

(3.15)
$$\tan\theta = \frac{\Delta r_2 - r\,\Delta r_2}{\Delta r_1 - r\,\Delta r_1}$$

with

(3.16)
$$\Delta r_1 = \sum_{\alpha=1}^{m} r_{1\alpha} - \tfrac{1}{2}m(m+n+1), \qquad \Delta r_2 = \sum_{\alpha=1}^{m} r_{2\alpha} - \tfrac{1}{2}m(m+n+1).$$

Substitution of each rank vector $(r_{1\alpha}, r_{2\alpha})$ and $(s_{1\alpha}, s_{2\alpha})$ for (t_1, t_2) in (3.14) yields

$(m + n)$ values of z. The z's are then ranked and the ranks associated with the X- and Y-samples, the ranks being $R_1, \cdots, R_m, S_1, \cdots, S_n$. Wilcoxon again used the rank-sum statistic, but in the function

$$(3.17) \qquad \Delta^2 = 12\left[\sum_{\alpha=1}^{m} R_\alpha - \tfrac{1}{2}m(m + n + 1)\right]^2 \Big/ mn(m + n + 1).$$

The procedure is simple geometrically. The observation vectors are transformed to rank vectors as a scaling process, since first and second variates in an observation vector may not otherwise be commensurate. The rank vectors are plotted in the two-dimensional space, and a line with slope $\tan \theta$ of (3.15) is drawn, say, through the mean point $[\frac{1}{2}(m + n + 1), \frac{1}{2}(m + n + 1)]$. Each plotted rank-vector point is projected orthogonally onto the line, and the projection points are ranked along the line yielding the required ranks, $R_1, \cdots, R_m, S_1, \cdots, S_n$.

The distribution of Δ^2 is not known. It is formulated from the univariate problem in which Δ may be taken to be standard normal under H_0 for moderate sizes of m and n. In the bivariate problem, the line chosen for the final ranking gives a maximum or near maximum value for Δ^2, and univariate tables for the small-sample distribution of $\sum_{\alpha=1}^{m} R_\alpha$ are not satisfactory. Wilcoxon believed that Δ^2 had approximately a chi-square distribution with two, rather than one, degrees of freedom under H_0. This belief was based on empirical studies; his notebook contains many calculations of Δ^2 for various sets of data and some limited calculations of the randomization distribution of Δ^2 for special examples. Further study is needed to substantiate his belief.

3.4 *Wilcoxon's second bivariate method.* Again consider the bivariate rank vectors $(r_{1\alpha}, r_{2\alpha})$ and $(s_{1\beta}, s_{2\beta})$, $\alpha = 1, \cdots, m$; $\beta = 1, \cdots, n$, of subsection 3.3. Wilcoxon transformed the rank variates to yield new vectors, $(u_{1\alpha}, u_{2\alpha})$ and $(v_{1\beta}, v_{2\beta})$, wherein $u_{1\alpha} = r_{1\alpha} - r_{2\alpha}$, $u_{2\alpha} = r_{1\alpha} + r_{2\alpha}$; $v_{1\beta} = s_{1\beta} - s_{2\beta}$, $v_{2\beta} = s_{1\beta} + s_{2\beta}$. Note that a randomly selected vector (t_1, t_2) from the set of r- and s-vectors yields a correlation between t_1 and t_2, but that a randomly selected vector (w_1, w_2) from the set of u- and v-vectors has zero correlation between w_1 and w_2, a result that follows since t_1 and t_2 have equal variances.

Wilcoxon supposed that lack of correlation might justify an approximate procedure, properly valid when w_1 and w_2 are independent. Thus he suggested reranking of the variates in the u- and v-vectors leading to new rank vectors, say, $(R_{1\alpha}, R_{2\alpha})$ and $(S_{1\alpha}, S_{2\alpha})$. Then he computed $\sum_{\alpha=1}^{m} R_{1\alpha}$ and $\sum_{\alpha=1}^{m} R_{2\alpha}$, Δ_1^2 and Δ_2^2 from (3.17) by substituting the two new rank sums for $\sum_{\alpha=1}^{m} R_\alpha$ in that formula, and W^2, his proposed statistic where

$$(3.18) \qquad W^2 = \Delta_1^2 + \Delta_2^2.$$

On the basis of the assumed independence, he took W^2 to have the chi-square distribution with two degrees of freedom under H_0.

This method should be correct asymptotically with large m and n. It does not provide easy generalization beyond the bivariate case.

4. Other research

Wilcoxon had a third major interest in research in rank-order statistics. This was in regard to the distributions of ranges of rank totals in a two-way classification. His notes contain much preliminary study of the problem, and he worked with Dunn-Rankin [9] in the development of a dissertation on the topic in the School of Education. In addition, in research in progress, McDonald and Thompson have developed new results in this area at the Florida State University.

Daniel and Wilcoxon [8] have a paper on the design of factorial experiments scheduled for publication in *Technometrics* in the near future.

REFERENCES

[1] R. H. BERK, "Asymptotic properties of sequential probability ratio tests," Ph.D. thesis Harvard University, 1964.

[2] R. A. BRADLEY, "Some statistical methods in taste testing and quality evaluation," *Biometrics*, Vol. 9 (1953), pp. 22–38.

[3] R. A. BRADLEY, D. C. MARTIN, and F. WILCOXON, "Sequential rank tests, I. Monte Carlo studies of the two-sample procedure," *Technometrics*, Vol. 7 (1965), pp. 463–483.

[4] R. A. BRADLEY, S. D. MERCHANT, and F. WILCOXON, "Sequential rank tests, II. A modified two-sample procedure," *Technometrics*, Vol. 8 (1966).

[5] R. A. BRADLEY and K. M. PATEL, "A randomization test for equality of means of two multivariate populations with common covariance matrix," *Ann. Math. Statist.*, Vol. 36 (1965), pp. 730–731. (Abstract.)

[6] S. K. CHATTERJEE and P. K. SEN, "Non-parametric tests for the bivariate two-sample location problem," *Calcutta Statist. Assoc. Bull.*, Vol. 13 (1964), pp. 18–58.

[7] ———, "Nonparametric tests for the multisample multivariate location problem," NSF Technical Report, University of California, Berkeley, 1965.

[8] C. DANIEL and F. WILCOXON, "2^{p-q} plans robust against linear and quadratic trends," *Technometrics*, Vol. 8 (1966), pp. 259–278.

[9] P. DUNN-RANKIN, "The true probability distribution of the range of rank totals and its application to psychological scaling," Ed.D. thesis, Florida State University, Tallahassee, 1965.

[10] Z. GOVINDARAJULU, "Asymptotic normality of a class of nonparametric test statistics," AFOSR Technical Report, Case Institute of Technology, Cleveland and University of California, Berkeley, 1965.

[11] W. J. HALL, R. A. WIJSMAN, and J. K. GHOSH, "The relationship between sufficiency and invariance with applications in sequential analysis," *Ann. Math. Statist.*, Vol. 32 (1965), pp. 575–614.

[12] J. E. JACKSON and R. A. BRADLEY, "Sequential χ^2- and T^2-tests," *Ann. Math. Statist.*, Vol. 32 (1961), pp. 1063–1077.

[13] C. Y. KRAMER, "A method of choosing judges for a sensory experiment," *Food Research*, Vol. 20 (1955), pp. 492–496.

[14] ———, "Additional tables for a method of choosing judges for a sensory experiment," *Food Research*, Vol. 21 (1956), pp. 598–600.

[15] E. L. LEHMANN, "The power of rank tests," *Ann. Math. Statist.*, Vol. 24 (1953), pp. 23–43.

[16] L. LYNCH and J. E. FREUND, "On the analysis of paired ranked observations," ONR Technical Report No. 32, Virginia Polytechnic Institute, Blacksburg, 1957.

[17] R. C. MILTON, "Rank order probabilities; two-sample normal shift alternatives," *Ann. Math. Statist.*, Vol. 36 (1965), pp. 1613–1614. (Abstract.)

[18] ———, "Power of two-sample nonparametric tests against the normal shift alternative," *Ann. Math. Statist.*, Vol. 36 (1965), p. 1614. (Abstract.)

[19] ———, "Sequential two-sample rank tests of the normal shift hypothesis," *Ann. Math. Statist.*, Vol. 36 (1965), p. 1614. (Abstract.)

[20] E. A. PARENT, JR., "Sequential ranking procedures," ONR Technical Report, Stanford University, Stanford, 1965.

[21] D. S. ROBSON, "A rank-sum test of whether two multivariate samples were drawn from the same population," ONR Technical Report No. 14, Cornell University, Ithaca, 1964.

[22] I. R. SAVAGE and L. J. Savage, "Finite stopping time and finite expected stopping time," *J. Roy. Statist. Soc. Ser. B*, Vol. 27 (1965), pp. 284–289.

[23] I. R. SAVAGE and J. SETHURAMAN, "Properties of the generalized SPRT based on ranks for Lehmann alternatives," ONR Technical Report No. M-84, Florida State University, Tallahassee and NSF Technical Report No. 7, Stanford University, Stanford, 1965.

[24] P. K. SEN, "On a class of multisample multivariate non-parametric tests," USPH Technical Report, University of North Carolina, Chapel Hill, 1965.

[25] ———, "On a class of two sample bivariate non-parametric tests," NSF Technical Report, University of California, Berkeley, 1965.

[26] P. K. SEN and Z. GOVINDARAJULU, "On a class of C-sample weighted rank-sum tests for location and scale," NSF and AFOSR Technical Report, University of California, Berkeley, 1965.

[27] A. WALD, *Sequential Analysis*, New York, Wiley, 1947.

[28] A. WALD and J. WOLFOWITZ, "Statistical tests based on permutations of the observations," *Ann. Math. Statist.*, Vol. 15 (1944), pp. 358–372.

[29] F. WILCOXON, "Individual comparisons by ranking methods," *Biometrics*, Vol. 1 (1945), pp. 80–83.

[30] ———, "Probability tables for individual comparisons by ranking methods," *Biometrics*, Vol. 3 (1947), pp. 119–122.

[31] F. WILCOXON, S. K. KATTI, and R. A. WILCOX, *Critical Values and Probability Levels for the Wilcoxon Rank Sum Test and the Wilcoxon Signed Rank Test*, Pearl River and Tallahassee, American Cyanamid Company and Florida State University, 1963.

[32] F. WILCOXON, L. J. RHODES, and R. A. BRADLEY, "Two sequential two-sample grouped rank tests with applications to screening experiments," *Biometrics*, Vol. 19 (1963), pp. 58–84.

[33] F. WILCOXON and R. A. BRADLEY, "A note on the paper 'Two sequential two-sample grouped rank tests with application to screening experiments'," *Biometrics*, Vol. 20 (1964) pp. 892–895.