STICHTING
MATHEMATISCH CENTRUM
2e BOERHAAVESTRAAT 49
AMSTERDAM

SP 81
(S 349)

Monte Carlo estimation of the powers of the distribution-free two sample tests of Wilcoxon, van der Waerden and Terry and comparison of these powers

P.v.d. Laan and J Oosterhoff

1965

Monte Carlo estimation of the powers of the distribution-free two-sample tests of Wilcoxon, van der Waerden and Terry and comparison of these powers *)

by P. van der Laan **) *and J. Oosterhoff* ***)

UDC 518.5 + 519.2

S u m m a r y

This paper describes a method for determining the power functions of the distribution-free two-sample tests of WILCOXON, VAN DER WAERDEN *and* TERRY *by a simulation technique, and for comparing these tests with each other and with parametric tests. For samples of size 6 and normal shift alternatives the numerical results are reproduced.*

Samenvatting

In dit artikel wordt een methode beschreven voor het bepalen van de onderscheidingsvermogens van de verdelingsvrije twee-steekproeven toetsen van WILCOXON, VAN DER WAERDEN en TERRY door middel van een Monte Carlo techniek en voor het vergelijken van deze toetsen onderling en met parametrische toetsen. Voor normale verschuivingsalternatieven en twee steekproeven ieder van 6 waarnemingen worden numerieke resultaten gegeven.

1. Introduction

Assume two independent random samples

$$\underline{x}_1{}^{1)}, \underline{x}_2, \cdots, \underline{x}_m \quad \text{and} \quad \underline{y}_1, \underline{y}_2, \cdots, \underline{y}_n \; (m \leqslant n) \tag{1}$$

are given from populations with continuous cumulative distribution functions $F(x)$ and $G(y)$ respectively. One wishes to test the null-hypothesis:

$$H_0 : F(x) \equiv G(x) \tag{2}$$

against the alternative hypothesis:

$$H_1 : F \neq G, \tag{3}$$

specifically "location" alternatives.

*) Presented at the "Statistische Dag 1965" of the V.V.S. (Mathematical section) at Amsterdam.
Report S 349 (SP 81) of the Stat. Dept., Mathematical Centre, Amsterdam.

**) N.V. Philips Research Laboratories, Eindhoven.
***) Mathematical Centre, Amsterdam.

[1]) Random variables will be distinguished from fixed numbers (e.g. from values they assume in an experiment) by underlining their symbols.

Three important distribution-free tests for this two-sample problem are:

a. The two-sample test of WILCOXON (MANN-WHITNEY)
b. The two-sample test of VAN DER WAERDEN (X-test)
c. The two-sample test of TERRY (FISHER-YATES, HOEFFDING).

Let the $N = m + n$ observations be ranked, $\underline{r}_1 < \underline{r}_2 < \ldots < \underline{r}_m$ being the set of ranks of the x-observations. The three test statistics are then defined as follows:

a. WILCOXON: $$\underline{W} = \sum_{i=1}^{m} \underline{r}_i \tag{4}$$

b. VAN DER WAERDEN: $$\underline{X} = \sum_{i=1}^{m} \Psi\left(\frac{\underline{r}_i}{N+1}\right) \tag{5}$$

c. TERRY: $$\underline{T} = \sum_{i=1}^{m} \mathscr{E}\underline{x}_{r_i, N} \tag{6}$$

where $\Psi(.)$ is the inverse of the standard normal distribution function and $\underline{x}_{r_i,N}$ is the $\underline{r}_i$ —th order statistic of a sample of size N from a population with a standard normal distribution.

Various investigations concerning efficiencies and power functions of these three tests have been made. In the case of an underlying normal distribution it is important to compare the power functions with the power of STUDENT's two-sample test, which is the uniformly most powerful unbiased test against normal shift alternatives (for short: under STUDENT's conditions). We shall give a brief, and far from complete, survey of the literature on this subject.

PITMAN proved in [27] that the PITMAN asymptotic relative (local) efficiency of WILCOXON's test against STUDENT's test for the shifted normal distribution is equal to $3/\pi$. In [17], HODGES and LEHMANN showed that for all distributions this efficiency for shift alternatives is greater than or equal to .845. VAN DER VAART [33] found in 1950 that under STUDENT's conditions for one- and two-sided testing with $m + n \leqslant 5$ and $m + n \leqslant 6$, respectively, the difference in power of WILCOXON's and STUDENT's tests is small in the neighbourhood of the null-hypothesis, and he made it plausible that for large sample sizes the difference is not large either: for two-sided testing he has shown that the ratio of the second derivatives of the power functions in H_0 is about $3/\pi$. Using PITMAN's formula for the variance of WILCOXON's test statistic, SUNDRUM [29] determined the power of the test against normal and rectangular shift alternatives for large samples and compared this test with certain parametric tests. Numerical calculations of the power of WILCOXON's test for normal shift alternatives, and for $m = n = 3, 4, 5$ and various significance levels have been carried out by DIXON [7]. JEAN D. GIBBONS [12] also gives, besides other investigations, some numerical results for the powers of WILCOXON's and TERRY's tests for small sample sizes against normal alternatives differing only in location.

The asymptotic equivalence of VAN DER WAERDEN's and TERRY's test was established for a wide class of cases by CHERNOFF and SAVAGE [3]. They proved that if the underlying distribution has a density and a finite second moment the PITMAN asymptotic relative efficiency of the test for translation based on TERRY's statistic relative to the t-test is not smaller than 1, and PITMAN's efficiency is 1 only if the underlying distribution is normal. The test of VAN DER WAERDEN has under STUDENT's conditions for m fixed and $n \to \infty$ asymtotically the same power function as the test of STUDENT [39]. See for further information concerning VAN DER WAERDEN's test the papers [37, 38, 39].

For various underlying distributions HODGES and LEHMANN [18] determined the PITMAN asymptotic relative efficiencies against shifts of WILCOXON's test as compared to the normal scores test, of which VAN DER WAERDEN's and TERRY's tests are asymptotically equivalent versions. Their results are reproduced in Table 1.

TABLE 1

PITMAN-efficiencies of WILCOXON's test against the normal scores test.

F	$e_{W,N}(F)$
Rectangular	0
Exponential	0
Normal	$\frac{3}{\pi} \approx .955$
Logistic	$\frac{\pi}{3} \approx 1.05$
Double exponential	$\frac{3\pi}{8} \approx 1.18$
Cauchy	1.413

They stated that for bell-shaped densities $e_{W,N}(F)$ seems to increase as the tails of the underlying distributions increase in importance. This result was made precise by a theorem due to VAN ZWET [44]. Computations of exact power for the three tests for small sample sizes in the case of exponential and homogeneous shift alternatives (one- and two-sided) have been made by VAN DER LAAN [20].

2. Estimation of power functions with Monte Carlo methods

There are great mathematical difficulties determining the power functions of the three tests, mentioned in section 1, for small sample sizes. With Monte

Carlo methods one can try to find estimates of these powers. The simulation technique we used is as follows.

To get the power function of a test against shift alternatives the test statistic for pairs of samples of sizes m and n, respectively, from a certain distribution is computed after applying different shifts to the second sample of each pair. The numbers of rejections of H_0 for this test divided by M, the total number of pairs of samples, are unbiased estimates of the powers.

Several problems connected with such a simulation experiment are discussed in the following sections. But first we review some similar experiments made by other investigators.

DIXON and TEICHROEW [8] determined in 1953 for various sample sizes ($m = n = 5, 10, 20$; $m = 5, n = 10$; $m = 10, n = 20$ a.o.) the powers of some distribution-free two-sample tests (WILCOXON, KOLMOGOROV-SMIRNOV, etc.) against normal shift alternatives. The significance levels were .01, .05 and .10. They also made experiments with alternative hypotheses for which $\sigma \neq 1$. The estimates are based on 100 or 150 pairs of samples (one exception).

In 1961 HEMELRIJK [16] made experiments in order to estimate and compare the power functions of WILCOXON's and STUDENT's tests. He used 50 pairs of samples with $m = n = 10$ and with normal shift alternatives and significance level $\alpha = .025$. He also considered outlyers and the exponential distribution, in which case the alternative hypotheses were not chosen in the form of shifts but of multiplication factors ($G(x) \equiv F(x/k)$).

Extensive experiments were made by DOKSUM [9], THOMPSON [32] in 1964. Some of THOMPSON's results for exponential and uniform shifts can be compared with the exact results obtained by VAN DER LAAN [20].

3. Generation of pseudo-random numbers

We used pseudo-random numbers, generated with the mixed congruential method. A sequence of integers $t_1, t_2, \ldots$ is generated, defined by

$$t_{k+1} = \lambda t_k + a \pmod{2^P}. \tag{7}$$

After dividing by 2^P one can consider this sequence of numbers as a sequence of pseudo-random numbers between 0 and 1. If $\lambda \equiv 1 \pmod 4$ the period of this sequence is equal to 2^P; for a suitable choice of λ the serial correlation is particularly small. This method is a modification of the multiplicative congruential method ($a = 0$) of LEHMER [21]. COVEYOU [4], GREENBERGER [13] and ROTENBERG [28] have proposed and investigated this method.

This generator with parameters $a = 1$, $\lambda = 26\,353\,589$ and $P = 26$ is a function procedure of the ELECTROLOGICA X — 1 computer at the Mathematical Centre, so that one can get very rapidly successive pseudo-random deviates.

A sequence of 1 000 000 successive numbers was tested at length [1]. A perfunctory inspection of the test results gave the impression that this sequence was acceptable.

To obtain pseudo-random normal deviates these numbers are transformed by an approximation formula for the inverse of the standard normal distribution function, for example the formula (HASTINGS [15]):

$$u = z - \frac{a_0 + a_1 z + a_2 z^2}{1 + b_1 z + b_2 z^2 + b_3 z^3}$$

where a_i $(i = 0, 1, 2)$ and b_i $(i = 1, 2, 3)$ are given constants (see [15]) and

$$z = \sqrt{-2 \ln h} \qquad 0 < h \leqslant \cdot 5 \tag{8}$$

h being an observation from a homogeneous distribution on (0,1). If $.5 < h < 1$, then one must take $1 - h$ and $-u$ instead of h and u, respectively.

The difference between this approximation and the true value of u is in absolute value smaller than about .000 4. For practical applications this is sufficiently accurate, while this is a relatively fast procedure.

We use for an underlying normal distribution a χ^2-test for goodness of fit to compare the empirical power function of STUDENT's test with the theoretical one. This is another check whether the pseudo-random numbers used are more or less random.

4. Determination of the number of pairs of samples

To choose M, the number of pairs of samples for certain m and n and a fixed alternative we use two criteria.

a. The expected length of a confidence interval for the probability of rejecting H_0 if H_1 is true for a fixed test must be sufficiently small.
b. We want to test the equality of powers of a pair of tests at a certain confidence level. The power of such a test must be sufficiently large (see section 6).

Ad a. For a fixed test one can construct a central two-sided confidence interval with confidence level 2α for p, the probability of rejecting H_0, on the basis of the normal approximation with continuity correction (p not too close to 0 or 1). The left and right limit are

$$p_{l,r} = \frac{1}{M + u_\alpha^2}\left\{(x \mp \tfrac{1}{2}) + \tfrac{1}{2}u_\alpha^2 \mp u_\alpha \sqrt{\frac{(x \mp \frac{1}{2})(M - x \pm \frac{1}{2})}{M} + \tfrac{1}{4}u_\alpha^2}\right\} \tag{9}$$

where u_α is defined by

$$\frac{1}{\sqrt{2\pi}} \int_{u_\alpha}^{\infty} e^{-\frac{t^2}{2}} \mathrm{d}t = \alpha \qquad (0 < \alpha < \tfrac{1}{2}) \tag{10}$$

and x denotes the number of rejections of H_0.

It is easy to show that an upper bound for the expected length of this confidence interval is given by

$$\frac{1}{M + u_\alpha^2}\left(1 + u_\alpha \sqrt{M + 2 + u_\alpha^2}\right) \tag{11}$$

If one chooses $\alpha = .01$, as we do, we obtain for $M = 2000$:

$$\mathscr{E}\,\{p_r(\underline{x}) - p_b(\underline{x})\} \leqslant \cdot 0525$$

which we considered an acceptable value.

5. Determination of the distributions of the test statistics

In order to determine power functions at various significance levels, we have computed the distributions of the three test statistics under H_0 for $m = n = 6$, 8, 10, $m = 8$, $n = 12$ and $m = 5$, $n = 15$.

For the test of WILCOXON we used the generating function of the test statistic [22, vol II; p. 10] to compute the distribution of $\underline{W}$. For m, $n \leqslant 10$ one can use the tables in [36]. It is also possible (for m, $n \leqslant 20$) to use the tables determined by AUBLE [2]. The distributions of $\underline{X}$ and $\underline{T}$ were determined by computing all possible positive values of $\underline{X}$ and $\underline{T}$, respectively, using properties of symmetry of the problem. For this method we are indebted to Dr. VAN ZWET of the Mathematical Centre, Amsterdam, who proposed it to us. Later we came across a recent paper of KLOTZ [19] in which he also used properties of symmetry to compute critical values of T.

We determined critical values of the non-central $\underline{t}$ for certain significance levels using an approximation formula ([10, p. 13, formula IIa] with one extra term for $t\,(f, o, \alpha)$, which can be found in [35]). With this approximation formula we determined tail probabilities of the non-central $\underline{t}$ applying an iterative method. Using these values we determined shift alternatives in such a way that the power values covered more or less the whole range.

6. Testing the equality of power functions

To test the equality of power functions of a pair of tests A and B, one cannot use the well-known hypergeometrical test in a 2×2-table, because the rejection of H_0 for both tests depends on the same observations, so the two sequences of observations are strongly dependent.

In fact one has observations from a multinomial distribution with four classes.

Class 1: Both tests reject. Probability p_1 and number of occurences $\underline{n}_1$

Class 2: Both tests don't reject. Probability p_2 and number of occurences $\underline{n}_2$

Class 3: A rejects, B doesn't reject. Probability p_3 and number of occurences $\underline{n}_3$
Class 4: A doesn't reject, B rejects. Probability p_4 and number of occurences $\underline{n}_4$
with $\underline{n}_1 + \underline{n}_2 + \underline{n}_3 + \underline{n}_4 = M$.

One wants to test the hypothesis: $p_1 + p_3 = p_1 + p_4$. So it is sufficient to test the null-hypothesis:

$$G_0 : p_3 = p_4 \tag{13}$$

It is possible to test G_0 a.o. in one of the following two ways:

a. With a binomial test applied to $\underline{n}_3$ and $\underline{n}_4$, given that $\underline{n}_3 + \underline{n}_4 = n (= n_3 + n_4)$
b. With a likelihood ratio test using the fact that $-2 \ln \lambda$ is asymptotically distributed as $\underline{\chi}^2$.

Both tests are asymptotically ($M \to \infty$) equivalent. The method given under a. is uniformly most powerful unbiased and is more simple, so we shall apply this method.

To get an impression of the power, we considered the following two alternatives:

$$| p_3 - p_4 | = .02 \quad \text{and} \quad | p_3 - p_4 | = .04 \tag{14}$$

For a number of values of $p\,(= p_3 + p_4)$ between .04 and .4 we determined values of n such that $P[\underline{n} > n \mid p, M] \approx .9$. For every p considered one can compute the power of the test with this value of n. This value will be a lower bound for the actual power, which may be assumed, because $\underline{n}$ will be usually larger than the value of n, determined in the way mentioned above. With $M = 2000$ and a one-sided significance level of .01 the results run something like this:

$\lvert p_3 - p_4 \rvert$ \ p	.04	.05	.066...	.10	.125	.166...	.2	.25	.4
.02	.97	.92	.81	.64	.53	.42	.33	.29	.17
.04		≈ 1	≈ 1	.995	.99	.98	.94	.89	.67

The test statistics are positively correlated, so p will probably be small. For small values of $p\,(p < .25)$ the power for $| p_3 - p_4 | = .04$ is very large and for $p < .10$ the power for $| p_3 - p_4 | = .02$ is reasonably good. This was also an argument to take $M = 2000$. One must take M very much larger to get an appreciably larger power.

7. Results for m = n = 6

In this paper we shall only give a part of our results for $m = n = 6$. Further research is still in progress. In the tables below one can find estimated power

TABLE 2

Monte Carlo estimates. $m = n = 6$, one-sided, sign. level .01. Randomisation was used. Normal shift alternatives.

Shift	STUDENT exact	STUDENT	WILCOXON	VAN DER WAERDEN	TERRY
.2	.0215	.0205	.0210	.0220	.0210
.4	.0425	.0380	.0390	.0410	.0400
.6	.0777	.0740	.0740	.0775	.0765
.8	.1312	.1420	.1330	.1315	.1295
1.0	.2057	.2040	.1875	.1920	.1895
1.2	.3004	.2990	.2740	.2800	.2780
1.4	.4107	.4040	.3765	.3785	.3755
1.6	.5286	.5160	.4770	.4830	.4800
1.8	.6437	.6375	.5965	.5960	.5985
2.0	.7469	.7415	.7085	.7020	.7030
2.2	.8318	.8360	.7970	.7985	.7930
2.5	.9204	.9215	.8885	.8850	.8825

TABLE 3

Monte Carlo estimates. $m = n = 6$, one-sided, sign. level .05. Radomisation was used. Normal shift alternatives.

Shift	STUDENT exact	STUDENT	WILCOXON	VAN DER WAERDEN	TERRY
.2	.0931	.0890	.0880	.0845	.0845
.4	.1590	.1590	.1490	.1535	.1545
.6	.2495	.2535	.2425	.2460	.2450
.8	.3617	.3670	.3415	.3555	.3540
1.0	.4874	.4890	.4660	.4715	.4715
1.2	.6143	.6165	.5860	.5890	.5855
1.4	.7297	.7125	.6820	.6830	.6825
1.6	.8245	.8130	.7815	.7870	.7815
1.8	.8949	.8930	.8705	.8745	.8745
2.0	.9422	.9380	.9255	.9275	.9265
2.2		.9685	.9605	.9620	.9620
2.5		.9880	.9835	.9830	.9830

values for one-sided normal shift alternatives with significance levels .01, .05 and .046 537 respectively. Randomisation was used. However in the last case .046 537 is an exact significance level for WILCOXON's test, so in this case no randomisation for WILCOXON's test was necessary. For $\alpha = .01$, $\alpha = .05$ and .046 537 the rejection regions of the tests of VAN DER WAERDEN and TERRY

TABLE 4.

Monte Carlo estimates. $m = n = 6$, one-sided, sign. level .046 537 was chosen such that no randomisation for the test of WILCOXON was necessary.

Shift	STUDENT exact	STUDENT	WILCOXON	VAN DER WAERDEN	TERRY
.2	.0874	.0840	.0805	.0785	.0780
.4	.1503	.1480	.1415	.1450	.1440
.6	.2377	.2405	.2335	.2350	.2350
.8	.3474	.3500	.3315	.3370	.3385
1.0	.4717	.4675	.4505	.4555	.4545
1.2	.5988	.6040	.5720	.5735	.5695
1.4	.7158	.6915	.6700	.6735	.6725
1.6	.8133	.7995	.7715	.7745	.7740
1.8	.8868	.8850	.8640	.8680	.8680
2.0	.9369	.9350	.9215	.9205	.9220
2.2	.9677	.9665	.9585	.9590	.9595
2.5	.9899	.9870	.9830	.9820	.9825

are identical. Possible differences for the pairs in these cases are due to randomisation.

Comparing the power of STUDENT's test with the distribution-free tests for not too small shifts we found strongly significant results (significance level .01). Differences between the powers of WILCOXON's test and the test of VAN DER WAERDEN and TERRY, respectively, were rarely significant (significance level .01).

Acknowledgements

We wish to thank Prof. J. HEMELRIJK of the Mathematical Centre, Amsterdam for many helpful and challenging discussions. Thanks are due to Miss S. J. CHRISTEN for writing the simulation program, to Mr. J. M. ANTHONISSE (both of the Math. Centre, Amsterdam) for writing the program for the determination of the distribution of WILCOXON's test statistic and to Mr. H. C. E. KOOYMAN v. GULDENER and Mr. J. v. D. DRIESSCHE (both of N.V. Philips, Computing Centre, Eindhoven) for writing programs concerning the distributions of VAN DER WAERDEN and TERRY's test statistics.

Bibliography

[1] ANTHONISSE, J. M. and J. OOSTERHOFF (1964), Testing a sequence of one million pseudo-random numbers produced by a generator of the mixed congruential type. Report S 329 of the Stat. Department, Mathematical Centre, Amsterdam.

[2] AUBLE, D. (1953), Extended tables for the Mann-Whitney statistic. Bull. of the Inst. of Educational Res. Indiana Univ. **1** no. 2.

[3] CHERNOFF, H. and I. R. SAVAGE (1958), Asymptotic normality and efficiency of certain nonparametric test statistics. A.M.S. **29**, 972–994.

[4] COVEYOU, R. R. (1959), Serial correlation in the generation of pseudo-random numbers. J. Assoc. Comp. Mach. **6**, 72–74.

[5] DANTZIG, D. VAN (1951), On the consistency and the power of Wilcoxon's two-sample test. Proc. K.N.A. van W. A **54**, 1–8, Ind. Math **13**, 1–8.

[6] DANTZIG, D. VAN, and J. HEMELRIJK (1954), Statistical methods based on few assumptions. Bull. of the Intern. Stat. Inst. **34**, 239–267.

[7] DIXON, W. J. (1954), Power under normality of several nonparametric tests. A.M.S. **25**, 610–614.

[8] DIXON, W. J. and D. TEICHROEW (1953), Some sampling results on the power of nonparametric tests against normal alternatives. Report U.S. Dept. of Commerce. N.B.S., Abstract A.M.S. **25** (1954) 175.

[9] DOKSUM, K. A. (1964), Unpublished.

[10] EEDEN, CONSTANCE VAN (1961), Some approximations to the percentage points of the non-central t-distribution. Revue Inst. Int. de Stat. **29**, 4–31.

[11] FIX, E. and J. L. HODGES, Jr. (1955), Significance probabilities of the Wilcoxon test. A.M.S. **26**, 301–312.

[12] GIBBONS, JEAN D. (1963), On the power of rank tests on the equality of two distribution functions. Abstract A.M.S. **34**, 355.

[13] GREENBERGER, M. (1961), Notes on a new pseudo-random number generator. J. Assoc. Comp. Mach. **8**, 163–167.

[14] HAMMERSLEY, J. M. and D. C. HANDSCOMB (1964), Monte Carlo methods. Methuen & Co, Ltd., London.

[15] HASTINGS, Jr., C. (1955), Approximations for digital computers. Princeton Univ. Press, Princeton, New Jersey.

[16] HEMELRIJK, J. (1961), Experimental comparison of Student's and Wilcoxon's two-sample tests. Quant. Methods in Pharmacology, North Holland Publ. Co, 118–134. Rep. SP 74 Stat. Dept., Mathematical Centre, Amsterdam.

[17] HODGES, Jr., J. L. and E. L. LEHMANN (1956), The efficiency of some nonparametric competitors of the t-test. A.M.S. **27**, 324–335.

[18] HODGES, Jr., J. L. and E. L. LEHMANN (1961), Comparison of the Normal Scores and Wilcoxon tests. Proc. of the 4th Berkeley Symp. Vol 1, 307–317.

[19] KLOTZ, J. H. (1964), On the Normal Scores two-sample test. J.A.S.A. **59**, 652–664.

[20] LAAN, P. VAN DER (1964), Exact power of some rank tests. Publ. de l'Inst. Stat. de l'Univer. de Paris **13**, 211–234. Rep. S 331 Stat. Dept., Math. Centre, Amsterdam.

[21] LEHMER, D. H. (1951), Mathematical methods in large-scale computing units. Ann. Comp. Lab. Harvard Univ. **26**, 141–146.

[22] MACMAHON, P. A. (1960), Combinatory Analysis I, II. Chelsea Publ. Comp., N.Y.

[23] MANN, H. B. and D. R. WHITNEY (1947), On a test of whether one of two random variables is stochastically larger than the other. A.M.S. **18**, 50–60.

[24] MARSAGLIA, G., M. D. MACLAREN and T. A. BRAY (1964), A fast procedure for generating normal random variables. Techniques **7**, 4–10.

[25] MEYER, H. A. (editor) (1956), Symposium on Monte Carlo methods. Methuen & Co, Ltd., London.

[26] Nat. Bur. of Stand., Applied Mathematics Series 23 (1953), Tables of normal probability functions. U.S. Government Printing Office, Washington, D.C.

[27] PITMAN, E. J. G. (1948), Lecture notes on nonparametric statistical inference. Lectures given for the Univ. of North Carolina, Inst. of Statist.
[28] ROTENBERG, A. (1960), A new pseudo-random number generator. J. Assoc. Comp. Mach. 7, 75–77.
[29] SUNDRUM, R. M. (1953), The power of Wilcoxon's 2-sample test. J.R.S.S. Series B, **15**, 246–252.
[30] TEICHROEW, D. (1956), Tables of expected values of order statistics and products of order statistics. A.M.S. **27**, 410–426.
[31] TERRY, M. E. (1952), Some rank order tests which are most powerful against specific parametric alternatives. A.M.S. **23**, 346–366.
[32] THOMPSON, R. (1964), Unpublished.
[33] VAART, H. R. VAN DER (1950), Some remarks on the power of Wilcoxon's test for two samples I, II. Proc. K.N.A. van W. A **53**, 494–520, Ind. Math. **12**, 146–172.
[34] VAART, H. R. VAN DER (1953), An investigation on the power of Wilcoxon's two-sample test if the underlying distributions are not normal. Proc. K.N.A. van W. A **56**, 438–448, Ind. Math. **15**, 438–448.
[35] VESELÁ, A. (1964), Kritische Werte der Studentschen t-Verteilung für die Freiheitsgrade zwischen 30 und 120. Biometrische Zeitschrift **6**, 123–127.
[36] WABEKE, Ir. DORALINE en CONSTANCE VAN EEDEN (1955), Handleiding voor de toets van Wilcoxon. Report S 176 (M 65) Statist. Dept., Mathematical Centre, Amsterdam.
[37] WAERDEN, B. L. VAN DER (1952), Order tests for the two-sample problem and their power. Proc. K.N.A. van W A **55**, 453–458, Ind. Math. **14**, 453–458.
[38] WAERDEN, B. L. VAN DER (1953), Order tests for the two-sample problem. Proc. K.N.A. van W. A **56**, 303–310 and 311–316, Ind. Math. **15**, 303–310 and 311–316.
[39] WAERDEN, B. L. VAN DER (1953), Ein neuer Test für das Problem der zwei Stichproben. Math. Annalen **126**, 93–107.
[40] WAERDEN, B. L. VAN DER (1956), The computation of the X-distribution. Proc. of the third Berkeley Symp. on Math. Stat. and Prob. **1**, 207–208.
[41] WAERDEN, B. L. VAN DER (1957), Mathematische Statistik, Springer-Verlag, Berlin.
[42] WAERDEN, B. L. VAN DER, und E. NIEVERGELT (1956), Tafeln zum Vergleich zweier Stichproben mittels X-test und Zeichentest. Springer-Verlag, Berlin.
[43] WILCOXON, F. (1945), Individual Comparisons by ranking methods. Bio^{s}. **1**, 80–83.
[44] ZWET, W. R. VAN (1964), Convex transformations of random variables. Mathematical Centre Tracts 7, Math. Centre, Amsterdam.