Exact distributions of two-sample rank statistics and block rank statistics using computer algebra

Document Version:
Publisher's PDF, also known as Version of Record (includes final page, issue and volume numbers)

Please check the document version of this publication:

• A submitted manuscript is the version of the article upon submission and before peer-review. There can be important differences between the submitted version and the official published version of record. People interested in the research are advised to contact the author for the final version of the publication, or visit the DOI to the publisher's website.
• The final author version and the galley proof are versions of the publication after peer review.
• The final published version features the final layout of the paper including the volume, issue and page numbers.

[Link to publication](#)

Exact distributions of two-sample rank statistics and block rank statistics using computer algebra

M.A. van de Wiel

Department of Mathematics and Computing Science

Eindhoven University of Technology

P. O. Box 513

5600 MB Eindhoven, The Netherlands

`markvdw@win.tue.nl`

URL: http://www.win.tue.nl/math/bs/statistics/vdwiel

Abstract

We derive generating functions for various rank statistics and we use computer algebra to compute the exact null distribution of these statistics. We present various techniques for reducing time and memory space used by the computations. We use the results to write Mathematica notebooks for computing exact tail-probabilities and to extend tables of critical values for some well-known rank statistics.

Keywords   Two-sample linear rank tests, block rank tests, exact null distributions, generating functions, computer algebra.

1   Introduction

In this paper we derive generating functions of two-sample linear rank statistics and block rank statistics. We use these generating functions as a powerful tool for computing exact distributions. We implemented these generating functions directly in the computer algebra package Mathematica, which resulted in notebooks for computing exact tail-probabilities and critical values. We increased the efficiency of the notebooks by using certain techniques for cutting off parts of the expressions that arise during a computation. The notebooks enabled us to extend the existing tables with critical values for various test statistics. Among them are the Van der Waerden statistic, which is more powerful than the Wilcoxon statistic for distributions with light tails, and the well-known Friedman statistic. The majority of the existing tables date back to the fifties and the sixties. One often used recursions to compute the tables. Examples of such recursions can be found in Kendall and Stuart (1977) and Gibbons and Chakraborti (1992) for the Kendall rank correlation test and the Wilcoxon rank-sum test, respectively. The procedures for computing exact tail-probabilities are very useful when we deal with ties, because then the (conditional) distribution depends on the tie strucure, which makes it practically impossible to create tables with critical values for all possible tie structures.

A major advantage of using generating functions and computer algebra systems over other approaches is that one can work directly with mathematical objects like polynomials the way we are used to do as humans, as opposed to representations of these objects in arrays etc., which are suitable for computers only. Another advantage is that computer algebra systems use infinite precision, so that rounding errors during computations do not occur. One may find a discussion about the use of computer algebra within statistics and probability in Kendall (1993). Baglivo et al. (1993) and Van de Wiel et al. (1997) use computer algebra packages to implement generating functions.

This paper is organised as follows. In section 2 we deal with two-sample rank statistics. We present applications of a two-dimensional generating function derived in Streitberg and Röhmel (1986). Section 3 is devoted to one of the main results of this paper: a branch-and-bound algorithm for the computation of critical values. The algorithm is useful in those cases for which the direct application of the Streitberg-Röhmel formula is too time- and memory-consuming. In Section 4 we use the Streitberg-Röhmel formula to derive a closed form generating function for the Halperin statistic. We extend the Streitberg-Röhmel formula for two-sample rank statistics to a formula for block rank statistics in section 5. Section 4 contains new tables of critical values.

For further details about the presented tests we refer to Gibbons and Chakraborti (1992).

2 Applications of the Streitberg-Röhmel formula to the two-sample case

In this section we show how we use a two dimensional generating function to compute the null distribution or critical values of two-sample linear rank statistics.

2.1 The Streitberg-Röhmel generating function

Suppose two independent random samples $X_1, \ldots, X_m$ and $Y_1, \ldots, Y_n$ with distribution functions F and G, respectively, are given. We define a general two-sample rank statistic T and we present a two dimensional generating function of this rank statistic.

Definition 2.1 *A linear rank statistic T is a statistic of the form*

$$T_N = \sum_{\ell=1}^{N} a(\ell)\, Z_\ell, \tag{1}$$

where $a : \{1, \ldots, N\} \to \mathbf{R}$ is said to be a rank score function, $m + n = N$, $Z_\ell = 1$ if the ℓth order statistic of the combined sample $X_1, \ldots, X_m, Y_1, \ldots, Y_n$ is an X and $Z_\ell = 0$ otherwise.

Theorem 2.2 *Let T_N be a linear rank statistic and $H_{T_N}(x) = \sum_{t=0}^{\infty} \Pr(T_N = t)\, x^t$ and let $a(\ell)$ be as in definition 2.1, then under H_0*

$$H_{T_N}(x) = \frac{1}{\binom{N}{m}}(1 + x^{a(1)} y) \ldots (1 + x^{a(N)} y)[[y^m]], \tag{2}$$

where $P(x, y)[[y^m]]$ is the coefficient of y^m.

Proof: for a proof we refer to Streitberg and Röhmel (1986). □

We note that Streitberg and Röhmel derive (2) under the unnecessary assumption that the scores are nonnegative integers. In fact, their proof is also valid for arbitrary real scores. This formula was already known to Euler (1748) in the context of generating the frequencies of all outcomes of the sum of m integers that form a subset of N integers. Streitberg and Röhmel present several algorithms for computing the formula. These are quite complex and translate mathematical structures into computer structures. However, a computer algebra package deals directly with (2).

2.2 Ties

We emphasize that Theorem (2.2) also holds when ties are present. However, we have to apply a conditional argument to the probabilities. We are aware that testing is not distribution-free anymore, since the distribution functions determine the probability on a certain tie configuration and tail probabilities are generally different for different tie configurations. However, it is impossible

to compute exact unconditional tail probabilities, because we do not know the probability on a certain tie configuration, since F and G are unknown. We should note that the choice of the way one deals with the scores in the case of ties influences the power of the test. Arrenberg (1994) shows that natural ranks are better than midranks for the Wilcoxon rank-sum test. The principle of conditional testing may be found in Lehmann (1975) and (Hájek, 1969, Ch. VII).

2.3 Applications

We represent the rank statistics by their corresponding rank scores. We denote the ℓth rank score for test statistic T_N by $a_{T_N}(\ell)$. Substition of these scores into (1) yields the rank statistic, while substitution of these scores into (2) yields the generating function.

2.3.1 Two-sample location tests

We wish to test the null hypothesis

$$H_0 : \mu_X = \mu_Y \quad \text{against} \quad H_1 : \mu_X \neq \mu_Y,$$

where μ_X and μ_Y denote the expectations of X and Y, respectively. For all tests we consider the combined sample $(X_1, \ldots, X_m, Y_1, \ldots, Y_n)$ and $m + n = N$. Well-known scores are:

$$a_{W_N}(\ell) \quad = \quad \ell, \quad \text{the Wilcoxon (1945) rank sum scores} \tag{3}$$

$$a_{V_N}(\ell) \quad = \quad \Phi^{-1}\left(\frac{\ell}{N+1}\right), \quad \text{the Van der Waerden (1957) scores,} \tag{4}$$

where Φ^{-1} denotes the inverse of the standard normal distribution function.

2.3.2 Two-sample scale tests

We wish to test the null hypothesis

$$H_0 : \sigma_X = \sigma_Y \quad \text{against} H_1 : \sigma_X \neq \sigma_Y,$$

where σ_X and σ_Y denote the standard deviations of the X- and Y-population, respectively. Again we consider the combined sample $(X_1, \ldots, X_m, Y_1, \ldots, Y_n)$, $m + n = N$. The following scores are often used:

$$a_{M_N}(\ell) \quad = \quad \left(\ell - \frac{N+1}{2}\right)^2, \quad \text{the Mood (1954) scores} \tag{5}$$

$$a_{A_N}(\ell) \quad = \quad \left|\left(\ell - \frac{N+1}{2}\right)\right|, \quad \text{the Freund-Ansari-Bradley (1957 and 1960) scores} \tag{6}$$

$$a_{K_N}(\ell) \quad = \quad \left[\Phi^{-1}\left(\frac{\ell}{N+1}\right)\right]^2, \quad \text{the Klotz (1962) scores,} \tag{7}$$

where Φ^{-1} denotes the inverse of the standard normal distribution function.

2.3.3 Percentile modified rank tests

The percentile modified rank statistics for location and scale were proposed by Gastwirth (1965). One has to select two numbers s and r, and to score only the data in the upper sth and lower rth percentiles of the combined sample. In other words, we assign the scores to the data and then we set the scores of the observations in the middle to zero. The freedom to choose s and r enables one to increase the power of the test statistic by incorperating knowledge of any characteristic of the underlying distribution functions (e.g. asymmetry) into the test statistic. Let $S = \lfloor N_s \rfloor$ and $R = \lfloor N_r \rfloor$, where $\lfloor N_q \rfloor$ denotes the largest integer not exceeding the number qN, $q = r, s$.

The percentile modified rank statistics for location and scale are denoted by $G_{N,r,s}$ and $J_{N,r,s}$, respectively. They are defined by the following scores, where $c_N = 1$ if N is odd and $c_N = 1/2$ if N is even.

$$a_{G_{N,r,s}}(\ell) = \begin{cases} -(R - \ell + c_N) & \text{if } 1 \le \ell \le R \\ \ell - (N - S) + c_N - 1 & \text{if } N - S + 1 \le \ell \le N \\ 0 & \text{otherwise} \end{cases} \tag{8}$$

and

$$a_{J_{N,r,s}}(\ell) = \begin{cases} R - \ell + c_N & \text{if } 1 \le \ell \le R \\ \ell - (N - S) + c_N - 1 & \text{if } N - S + 1 \le \ell \le N \\ 0 & \text{otherwise} \end{cases} \tag{9}$$

3 Branch-and-bound algorithm

For some test statistics the number of terms in expression (2) increases very fast when N increases. The rate of increase obviously depends on the scores. If sums of scores have relatively often the same value (e.g., for Wilcoxon scores), the number of terms increases relatively slowly. For rank tests that use inverse normal scores, i.e. the Van der Waerden test and the Klotz test, sums of scores have rarely the same value. With the aid of combinatorial arguments we proved (see Van de Wiel (1996)) that, if $m = N/2$, the number of different values of the Van der Waerden test statistic V_N equals $Q_m = \frac{1}{2}(3^m + 1)$. If for example $N = 30, m = 15$, then $Q_{15} = 7\,174\,454$. Computation of an expression with such a large number of terms gives memory and time problems.

We experienced that, using a Pentium 200 Mhz PC, it is possible to compute the critical values of a two-sample linear rank statistic for $N \le 20$ within reasonable time by simply expanding the product in (2) with a computer algebra system. This is generally not possible for larger N and therefore we present a branch-and-bound algorithm for computing critical values.

We note that Mehta et al. (1987) give a branch-and-bound algorithm for permutation test statistics that contains some of the principles of our algorithm. However, there are some important differences. First of all, instead of a generating function, which is a common mathematical object, they use a network representation. Therefore, they have to set up rules for multiplication, whereas this is trivial for the expressions that appear during the expansion of (2). Secondly, they compute tail probabilities which is a bit easier than the computation of critical values. Furthermore, they do not mention the order in which the scores are used in the algorithm, whereas this is important for the speed of the algorithm for some statistics (e.g. Van der Waerden and Klotz). Finally, they use dynamic programming for finding the largest and smallest sums of scores, whereas we use explicit formulas.

3.1 Principles of the algorithm

We first choose an interval of which we think that it contains the critical value. An approximation formula may help us to do so. If the interval does not contain the critical value, then this will be detected at the end of the algorithm, enabling us to restart the algorithm with a different interval. We rearrange the scores in decreasing order of their absolute values, so that the first scores have the largest contribution on the sum of scores. The expressions we deal with are sums of expressions of the form $cx^b y^k$, where c is a constant, b a sum of scores and k the number of scores that contribute to b. We denote this expression, which we consider a polynomial in y, by P. The term of degree k in P is denoted by $P[y^k]$. Note that $P[y^k]$ includes y^k.

Each loop we add the next score with the largest absolute value. Within each loop we use two types of bounding rules. The first bounding rule is to discard of all terms consisting of too many

or too few scores, for which we use that we only need the sums of exactly m scores. For the second bounding rule we first compute the maximum and the minimum value that can be added to b in further iteration loops. Then we discard of each term for which the sum of b and the maximum is smaller than the lower bound of the interval. After that, we add all coefficients of the terms for which the sum of b and the minimum is larger than the upper bound and discard of these terms too. We multiply this sum of coefficients with the number of ways to choose the next scores out of the unassigned scores. The sum of all these products times $1/\binom{N}{m}$ is the right-tail probability for the upper bound. With this probability and with the expression that is left after the last iteration loop we find the right critical value. Before giving a formal description of the algorithm, we present an example.

3.2   Example

This example is meant to illustrate the basic ideas of the algorithm. In this example we go through the algorithm for the Mood test (see subsection 2.3). For the sake of brevity we take small samples and we perform all steps only once.

Input: $N = 10, m = 5, \alpha = 0.05$, scores: $\{\frac{81}{4}, \frac{49}{4}, \frac{25}{4}, \frac{9}{4}, \frac{1}{4}, \frac{1}{4}, \frac{9}{4}, \frac{25}{4}, \frac{49}{4}, \frac{81}{4}\}$.
Output: one-sided right critical value.

We choose an interval $[l, u]$ of which we are sure that it contains the one-sided right critical value. One might do this with the aid of approximations. In this case we would need quite a large interval, because approximations are not good for such small samples. However, as we discussed in the beginning of this section, the algorithm is designed to compute critical values for the cases $N \geq 20$. For those cases, approximations are better, so that one can use a small interval. For this example, let $[l, u] = [56, 62]$. To find the critical value it is sufficient to know $\Pr(T_N \geq u)$ and $\Pr(T_N = t), l \leq t \leq u$.

We rearrange the scores in descending order of their absolute values:

$$\{\frac{81}{4}, \frac{81}{4}, \frac{49}{4}, \frac{49}{4}, \frac{25}{4}, \frac{25}{4}, \frac{9}{4}, \frac{9}{4}, \frac{1}{4}, \frac{1}{4}\}$$

Observe that the differences between the large unequal values in this list are larger than for the small unequal values (compare $\frac{81}{4} - \frac{49}{4}$ with $\frac{9}{4} - \frac{1}{4}$). Therefore, the larger scores have a larger relative contribution on the value of the test statistic than the smaller ones. This would not be the case for Wilcoxon scores, because these scores are equidistant.

We expand the product in (2) for $\ell \leq 6$. Then we obtain:

$$
\begin{aligned}
P &= 1 + \left(2\,x^{\frac{25}{4}} + 2\,x^{\frac{49}{4}} + 2\,x^{\frac{81}{4}}\right) y + \left(x^{\frac{25}{2}} + 4\,x^{\frac{37}{2}} + x^{\frac{49}{2}} + 4\,x^{\frac{53}{2}} + 4\,x^{\frac{65}{2}} + x^{\frac{81}{2}}\right) y^2 \\
&+ \left(2\,x^{\frac{99}{4}} + 2\,x^{\frac{123}{4}} + 2\,x^{\frac{131}{4}} + 8\,x^{\frac{155}{4}} + 2\,x^{\frac{179}{4}} + 2\,x^{\frac{187}{4}} + 2\,x^{\frac{211}{4}}\right) y^3 \\
&+ \left(x^{37} + 4\,x^{45} + 4\,x^{51} + x^{53} + 4\,x^{59} + x^{65}\right) y^4 + \left(2\,x^{\frac{229}{4}} + 2\,x^{\frac{261}{4}} + 2\,x^{\frac{285}{4}}\right) y^5 + x^{\frac{155}{2}} y^6
\end{aligned}
$$

To compute the critical value we only need the coefficient of y^5. Since there are 4 scores left to add, we know that the terms that arise from multiplication with the term '1' in P will not contribute to this coefficient. Also, the terms that arise from multiplication with the term '$x^{\frac{155}{2}} y^6$' in P will not not contribute to this coefficient. Therefore, we may remove '1' and '$x^{\frac{155}{2}} y^6$' from P.

$$
\begin{aligned}
P &= \left(2\,x^{\frac{25}{4}} + 2\,x^{\frac{49}{4}} + 2\,x^{\frac{81}{4}}\right) y + \left(x^{\frac{25}{2}} + 4\,x^{\frac{37}{2}} + x^{\frac{49}{2}} + 4\,x^{\frac{53}{2}} + 4\,x^{\frac{65}{2}} + x^{\frac{81}{2}}\right) y^2 \\
&+ \left(2\,x^{\frac{99}{4}} + 2\,x^{\frac{123}{4}} + 2\,x^{\frac{131}{4}} + 8\,x^{\frac{155}{4}} + 2\,x^{\frac{179}{4}} + 2\,x^{\frac{187}{4}} + 2\,x^{\frac{211}{4}}\right) y^3 \\
&+ \left(x^{37} + 4\,x^{45} + 4\,x^{51} + x^{53} + 4\,x^{59} + x^{65}\right) y^4 + \left(2\,x^{\frac{229}{4}} + 2\,x^{\frac{261}{4}} + 2\,x^{\frac{285}{4}}\right) y^5.
\end{aligned}
$$

5

Let us now consider the term of degree 4 in P, denoted by $P[y^4]$. Thus,

$$P[y^4] = \left(x^{37} + 4\,x^{45} + 4\,x^{51} + x^{53} + 4\,x^{59} + x^{65}\right)\,y^4.$$

Let cx^b be a term of the coefficient of y^4. Because $m = 5$, precisely one more score has to be added to b. The maximum M^+ of the scores that are left is $\frac{9}{4}$, the minimum M^- is $\frac{1}{4}$. We know that for computing the critical value it is sufficient to know the probabilities $\Pr(T_N \geq u)$ and $\Pr(T_N = t), l \leq t \leq u$. The terms cx^b for which $b + M^+ \leq l$ will not contribute to any of these probabilities and they may therefore be deleted. In our case we have $b \leq 56 - \frac{9}{4} = 53\frac{3}{4}$. So we delete $x^{37} + 4\,x^{45} + 4\,x^{51} + x^{53}$, which results in:

$$P[y^4] = \left(4\,x^{59} + x^{65}\right)\,y^4.$$

We also know that only those terms for which $b + M^- \geq u$ will contribute to $\Pr(T_N \geq u)$. In this case $b \geq 62 - \frac{1}{4} = 61\frac{3}{4}$. This is true for the term x^{65} and its contribution to $\Pr(T_N \geq u)$ equals $c\binom{4}{1}/\binom{N}{m} = 4/252 = 0.016$, where $\binom{4}{1}$ is the number of ways to choose 1 score out of the 4 scores that are not used yet. We store this contribution (or add it to the total) and we may then delete x^{65}. The result of the removal is:

$$P[y^4] = 4\,x^{59}\,y^4.$$

It is clear that we may apply the principle of deleting parts to any term $P[y^d]$. The result is a new polynomial P. We use the next score, $\frac{9}{4}$, and multiply P by $1 + x^{\frac{9}{4}}y$. For the resulting polynomial we repeat the previous steps until all scores are used. When all scores are used, we know $\Pr(T_N \geq u)$ and $\Pr(T_N = t), l \leq t \leq u$ which suffices to find the one-sided right critical value, i.e. the smallest value v for which $\Pr(T_N \geq v) \leq \alpha = 0.05$.

3.3 Formal description of the algorithm

Input: N integer, m integer, α: $\alpha \in [0,1]$, v: integer, $[l,u]$: interval that contains the right critical value, rank scores: reals.
Output: one-sided right α critical value of the statistic T_N with arbitrary scores.

1. Rearrange the scores such that $|a(1)| \geq \ldots \geq |a(N)|$.

2. Expand $P = \prod_{i=1}^{v} \left(1 + x^{a(i)}y\right)$.

3. $j \to v$, $q \to 0$.

4. Let k be the exponent of y in an arbitrary term of P, then $P \to P - I$, where I consists of all terms $P[y^k]$ of P for which $k > m$ or $k < m - (N - j)$.

5. Let r be the smallest exponent of y in $P, k \to r$.

6. Compute M^+, i.e. the sum of the $m - k$ largest scores and compute M^- i.e. the sum of the $m - k$ smallest scores.

7. Let b be the exponent of x in an arbitrary term of $P[y^k]$ then $P \to P - J_1 y^k$, where J_1 consists of all terms of $P[y^k]$ for which $b + M^+ \leq l$.

8. Let J_2 consist of all terms of $P[y^k]$ for which $b + M^- \geq u$ and add the coefficients of these terms.

9. Multiply this sum with $\binom{N-j}{m-k}$ and denote this product by t.

10. $q \to q + t$.

11. $P \to P - J_2 y^k$.

12. Let s be the degree of P. If $k < s$, $k \to k + 1$ and go to step 5.

13. $j \to j + 1$.

14. If $j \leq N$, $P \to P\left(1 + x^{a(j)}y\right)$ and go to step 4.

15. If $j > N$ compute $\Pr(T_N \geq u) = \frac{1}{\binom{N}{m}}q$, where T_N is the two-sample rank statistic.

16. Let $b(i)$ be the exponent of x in the ith term of $\frac{P}{y^m}$. Furthermore, let $c(i)$ be the coefficient in the ith term of $\frac{P}{y^m}$, where P is given in increasing order of $b(i)$ and let $d(i) = \sum_{j \geq i} c(j)$. Now determine the first term i^* for which $d(i^*) \leq \binom{N}{m}(\alpha - \Pr(T_N \geq u))$. Then the one-sided right α critical value is equal to $b(i^*)$.

3.3.1 Notes to formal description

Input: $N = m + n$, the total sample size, m is the sample size of the smallest sample, α is the one-sided confidence level. One should not choose v too large, because then the expression given in step 2 becomes too large. We experienced that $v = 6$ works quite well.
One may choose the interval with the aid of an approximation formula. The smaller the interval, the faster the algorithm. If one is not sure whether the interval contains the critical value, one can still use the algorithm. If the algorithm fails to give an answer then the real value will be larger than the upper bound and one should choose a new interval for which the lower bound is equal to the upper bound of the previous interval. If the algorithm gives an answer that is very close to the lower bound, one should check whether the answer is equal to the exponent of the first term of P. If this is the case then one should re-run with an new interval for which the upper bound is equal to the lower bound of the previous interval. In fact, one can always find the exact critical value with this method.

ad 1. The order in which the scores are used may influence the speed of the algorithm. As we argued in the example, the larger scores may have relatively more influence on the value of the test statistic than the smaller ones. In fact, we minimize the difference between M^+ and M^- (step 6.), so that the bounding procedure works optimally.

ad 3. We consider j as the number of scores that is used in the iteration so far. Because v scores are used in step 2 of the algorithm the initilization value of j is v. We need the variable q to compute the right-tail probability of the upper bound. This step is the start of the outer loop that ends at step 15.

ad 4. To avoid unnecessary computations, we discard of all terms for which $k > m$ or $k < m - (N - j)$, because we only need $P[y^m]$. Now, $N - j$ is the maximum that could be added to k, because this is the number of scores that are not used yet. Therefore, we only have to consider those terms for which $m - (N - j) \leq k \leq m$.

ad 6. Note that M^- may be negative as well. The $m - k$ scores have to be chosen from the $N - j$ scores that are not yet used.

ad 7. We discard of all those terms for which the sum of scores is too small.

ad 8. We discard of all those terms for which the sum of scores is too large.

ad 9,10. The number $\binom{N-j}{m-k}$ is the number of ways to choose $m-k$ scores out of the $N-j$ unassigned scores. We consider q as the number of sums of scores that exceed u and t as the contribution to q of those terms of $P[y^k]$ that satisfy the inequality in 8.

ad 14. Here we introduce the next score according to the sequence after step 1.

ad 16. Note that P now only consists of the term y^m and its coefficient. Note that

$$d(i) \leq \binom{N}{m}(\alpha - \Pr(T_N \geq u)) \Longleftrightarrow \sum_{j \geq i} c(j)/\binom{N}{m} + \Pr(T_N \geq u) \leq \alpha$$

$$\Longleftrightarrow \Pr(u > T_N \geq b(i)) + \Pr(T_N \geq u) \leq \alpha \Longleftrightarrow \Pr(T_N \geq b(i)) \leq \alpha.$$

The efficiency of this algorithm depends on the differences between the absolute values of the successive scores. If these differences are relatively large then the algorithm is very efficient, because then many terms will be removed in steps 7 and 11 in an early stage. This is the case when the scores are quadratic or when they are quantiles of the normal distribution. Hence, the algorithm is very efficient for the Mood scale test and the Van der Waerden test and especially for the Klotz test, because the Klotz scores have both characteristics.

For some cases it may be more efficient to perform steps 4 to 12 (the bounding) only for values of j which are a multiple of a certain integer d. This depends on the scores and the sample sizes. When one deals with large sample sizes, we know from practical experience that one should choose $d = 1$ or $d = 2$, because otherwise the expressions will be too large.

4 A generating function for the Halperin statistic

We show how we use (2) to derive a generating function with a closed form for the Halperin statistic. This statistic is used for data that are censored at a fixed point, also known as type I censoring.

4.1 The Halperin statistic

Let T be a fixed point and suppose we deal with mutually independent data $X_1, \ldots, X_m$ and $Y_1, \ldots, Y_n$ from which we observe

$$X_{(1)}, \ldots, X_{(m-r_m)} \quad \text{and } r_m \text{ X-observations after point } T,$$
$$Y_{(1)}, \ldots, Y_{(n-r_n)} \quad \text{and } r_n \text{ Y-observations after point } T,$$

where $X_{(i)}$ and $Y_{(j)}$ are order statistics. So we deal with two samples which are censored at the same fixed point T. We denote the distribution functions of the X-observations and the Y-observations by F and G, respectively. We wish to test the null hypothesis

$$H_0 : F(x) \equiv G(x) \quad \text{for all } x \leq T.$$

Halperin (1960) proposed a statistic for which he proves that it is consistent for alternatives of the form

$$H_1 : \frac{F(x)}{F(T)} > \frac{G(x)}{G(T)}, F(T) > G(T) \quad \text{for all } x \leq T.$$

The test statistic U_c is given by

$$U_c = M(m - r_m, n - r_n) + r_m(n - r_n), \tag{10}$$

where $M(m - r_m, n - r_n)$ is the Mann-Whitney statistic of the uncensored observations of the two samples.

4.2   Generating function

For U_c we define the following conditional probability generating function.

Definition 4.1 *Let U_c be the test statistic given in (10), then*

$$H_{U_c}(x|r) = \sum_{u=0}^{\infty} \Pr(U_c = u | r_m + r_n = r)\, x^u. \tag{11}$$

Lemma 4.2 *Let A and B be events and assume that $Pr(B) > 0$. Furthermore, let $C_1, \ldots, C_N$, be events that are mutually disjoint, whereas $\Pr(C_i) > 0, \Pr(B \cap C_i) > 0$, $k = 1, \ldots, N$, and $\sum_{k=1}^{N} \Pr(C_i) = 1$. Then,*

$$\Pr(A|B) = \sum_{i=1}^{N} \Pr(A|B \cap C_i) \Pr(C_i|B).$$

Proof: by straightforward verification.   □

The following theorem, which is not trivial because of a tricky conditioning argument, helps us to obtain the conditional distribution of the Halperin statistic from the distribution of the Mann-Whitney statistic.

Theorem 4.3 *Let U_c denote the Halperin statistic for sample sizes m and n and let $M(a,b)$ denote the Mann-Whiney statistic for sample sizes a and b. Then, under $H_0 : F = G$, with $\bar{r}_m = m - r_m$ and $\bar{r}_n = n - r_n$,*

$$\Pr(U_c = u | \bar{r}_m = a, \bar{r}_n = b) = \Pr(M(a,b) = u + (m-a)b). \tag{12}$$

Proof: see Appendix A.

Definition 4.4 *Let $P = \sum_{k=0}^{N} \sum_{\ell} c(\ell) x^{b(\ell)} y^k$, where $c(\ell)$ and $b(\ell)$ are arbitrary real numbers. Then,*

$$\Psi(P, k) = P[[y^k]] = \sum_{\ell} c(\ell) x^{b(\ell)}. \tag{13}$$

Theorem 4.5 *Let $H_{U_c}(x|r)$ be the probability generating function (11) of the Halperin statistic as defined in (10) and let Ψ be the operator as defined in Definition 4.4. Then, under $H_0 : F = G$,*

$$H_{U_c}(x|r) = \sum_{i=0}^{m} \frac{1}{\binom{r}{i}\binom{m+n}{m}} x^{g(i)} \Psi\left(\prod_{\ell=1}^{M=m+n-r} (1 + x^\ell y), m - i \right), \tag{14}$$

where $g(i) = i(n - r + i) - \frac{1}{2}(m-i)(m-i+1), i = 0, \ldots, m$.

Proof: Using Lemma 4.2 we have

$$\Pr(U_c = u | r_m + r_n = r) = \sum_{i=0}^{m} \Pr(U_c = u | r_m + r_n = r, r_m = i) \Pr(r_m = i | r_m + r_n = r). \tag{15}$$

Under H_0, the X_i's and Y_j's are equally distributed and hence we have $\Pr(X_i > T) = \Pr(Y_j > T) = 1 - F(T)$. Furthermore, r_m and r_n are independent. This implies that

$$\Pr(r_m = i | r_m + r_n = r) = \frac{\Pr(r_m = i, r_n = r - i)}{\Pr(r_m + r_n = r)} = \frac{\Pr(r_m = i)\Pr(r_n = r - i)}{\Pr(r_m + r_n = r)}$$

$$= \frac{\binom{m}{i}(F(T))^{m-i}(1 - F(T))^i \binom{n}{r-i}(F(T))^{n-r+i}(1 - F(T))^{r-i}}{\binom{m+n}{r}(F(T))^{m+n-r}(1 - F(T))^r} = \frac{\binom{m}{i}\binom{n}{r-i}}{\binom{m+n}{r}}. \tag{16}$$

So, r_m given $r_m + r_n$ is hypergeometrically distributed. We now set out to rewrite (12) in a form that suits for computing (11). We use Theorem 4.3 and the well-known relation between the Mann-Whitney statistic and the Wilcoxon statistic for uncensored samples: $M(m,n) = W(m,n) - \frac{1}{2}m(m+1)$ to write

$$\Pr(U_c = u | r_m + r_n = r, r_m = i) = \Pr(U_c = u | \bar{r}_m = m - i, \bar{r}_n = n - r + i)$$

$$= \Pr(M(m-i, n-r+i) = u + i(n-r+i) = \Pr(W(m-i, n-r+i) = u - g(i))$$

$$= \frac{1}{\binom{m+n-r}{m-i}} h(u - g(i), m-i, n-r+i), \tag{17}$$

where $h(u - g(i), m-i, n-r+i), i = 0, \ldots, m$ is the number of rank configurations of $m - i$ uncensored X-observations and $n - r + i$ uncensored Y-observations for which $W = u - g(i)$. Substituting (16) and (17) into (15) and (15) into (11) gives

$$H_{U_c}(x|r) = \sum_{u=0}^{\infty} \sum_{i=0}^{m} \frac{\binom{m}{i}\binom{n}{r-i}}{\binom{m+n-r}{m-i}\binom{m+n}{r}} h(u - g(i), m-i, n-r+i)x^u$$

$$= \sum_{i=0}^{m} \frac{1}{\binom{r}{i}\binom{m+n}{m}} x^{g(i)} \sum_{w=0}^{\infty} h(w, m-i, n-r+i)x^w, \tag{18}$$

where we set $w := u - g(i)$.

We use (2) with the Wilcoxon scores to observe that

$$\sum_{w=0}^{\infty} h(w, m-i, n-r+i)x^w = \Psi\left(\prod_{\ell=1}^{m+n-r}(1 + x^\ell y), m-i\right). \tag{19}$$

After substituting (19) into (18) we obtain the desired result. $\qquad\square$

Note on efficiency

From (17) we observe that, in order to compute $H_{U_c}(x|r)$, we might also use the well-known probability generating function of the Mann-Whitney statistic, i.e.

$$\sum_{w=0}^{\infty} \Pr(M(a,b) = k)x^k = \frac{1}{\binom{a+b}{a}} \frac{\prod_{i=a+1}^{a+b}(1 - x^i)}{\prod_{i=1}^{b}(1 - x^i)}. \tag{20}$$

In Halperin (1960) Halperin gives a recursive formula for computing $\Pr(U_c = u | r_m + r_n = r)$ that is also based on the null distribution of the Mann-Whitney statistic $M(m-i, n-r+i)$ for all $i, i = 1, \ldots, m$. From (15) we conclude that the computation of $H_{U_c}(x|r)$ with the aid of (20) requires the expansion of (20) for $a = m - i, b = n + r - i, i = 0, \ldots, m$. We experienced that this is much more time-consuming than the computation of $H_{U_c}(x|r)$ with the aid of (19). The first reason for this is that (20) is a quotient, which is harder to expand than (19). The second reason is that (20) has to be expanded for various values of a and b, whereas the product in (19) has to be expanded only once. For this last reason Halperin's formula is also less efficient than (19).

5 Generating function for block rank statistics

In this section we show how to compute the null distribution of block rank test statistics. We derive a generating function for these rank test statistics which may be seen as a generalization of (2). The main idea is to derive a generating function for the distribution of the ranks within the blocks and multiply these to obtain the generating function of the sum of the ranks. We note that Streitberg and Röhmel (1987) give recursive formulas for computing the null distribution of block rank statistics. They do not deal with the Friedman statistic.

5.1 Block rank tests

Suppose we deal with data that are presented in the form of a two-way layout of b rows and t columns. The rows indicate blocks and the columns indicate treatments. Let Y_{ij} be the observation in row i and column j. We consider the following model:

$$Y_{ij} = \mu + \beta_i + \gamma_j + \varepsilon_{ij},$$

$i = 1, \ldots, b, j = 1, \ldots t$, where μ is a general mean, β_i is a block effect, γ_j a treatment effect and ϵ_{ij} is a random error component, mutually independent and identically distributed for all i and j. We want to test the null hypothesis

$$H_0 : \gamma_1 = \ldots = \gamma_t = 0 \tag{21}$$

against the alternative hypothesis

$$H_1 : \gamma_j \neq 0 \text{ for at least one } j.$$

Within each block rank scores are assigned to the observations. The blocks are independent and may have different rank scores. We denote the rank score for the observation in the ith block belonging to the jth treatment by R_{ij}. Furthermore, let T_j denote the treatment totals $\sum_{i=1}^{b} R_{ij}$, B_i the block totals $\sum_{j=1}^{t} R_{ij}$ and E the total of all rank scores $\sum_{i=1}^{b} \sum_{j=1}^{t} R_{ij}$. We define $S = \sum_{j=1}^{t} T_j^2$. To test H_0 we consider test statistics of the form $c_1 S + c_2$, where c_1 and c_2 are constants and $c_1 > 0$. We can compute the distribution of S under H_0 if we can compute the joint distribution of the T_j's under H_0. Generating functions turn out to be a convenient tool for computing the joint distribution of the T_j's under H_0.

5.2 A generating function for block rank statistics

As in the two-sample case we need rank score functions. The rank score function for block i is defined as

$$a_i : \{1, \ldots, t\} \to \mathbf{R}. \tag{22}$$

If ties occur, then again the test becomes a conditional one (cf. our remarks in Section 2.2). We now introduce a probability generating function for observations in one block.

Definition 5.1 *The probability generating function of the row vector of the ranks within block i, $R_i = (R_{i1}, \ldots, R_{it})$, is defined by*

$$H_{R_i}(\vec{x}) = \sum_{\vec{v}=\vec{0}}^{\vec{\infty}} \Pr(R_{i1} = v_1, \ldots, R_{it} = v_t)\, x_1^{v_1} \ldots x_t^{v_t}, \tag{23}$$

where $\vec{x} = (x_1, \ldots, x_t), \vec{v} = (v_1, \ldots, v_t), \vec{0} = (0, \ldots, 0)$ and $\vec{\infty} = (\infty, \ldots, \infty)$, which are vectors of length t.

Theorem 5.2 *Let $\mathcal{S}_t$ be the symmetric group of all these permutations of $\{1, \ldots, t\}$. Under $H_0 : \gamma_1 = \ldots = \gamma_t = 0$ we have*

$$H_{R_i}(\vec{x}) = \frac{1}{t!} \sum_{\sigma \in \mathcal{S}_t} x_1^{a_i(\sigma(1))} \ldots x_t^{a_i(\sigma(t))}. \tag{24}$$

Proof: We know that $\Pr(R_{i1} = v_1, \ldots, R_{it} = v_t) = 0$ if $\{v_1, \ldots, v_t\} \neq \{a_i(1), \ldots, a_i(t)\}$. Under H_0, we have $\Pr(R_{ij} = a_i(\ell)) = \Pr(R_{ik} = a_i(\ell))$ for all $j \neq k$, for all ℓ and for all i. Therefore, we have $\Pr(R_{i1} = a_i(\sigma(1)), \ldots, R_{it} = a_i(\sigma(t))) = \frac{1}{t!}$ for all $\sigma \in \mathcal{S}_t$ and the result follows. $\square$

We use generating function (24) to derive a formula for the probability generating function of the vector $(T_1, \ldots, T_t)$.

Definition 5.3 *The probability generating function of the row vector of treatment totals $T = (T_1, \ldots, T_t)$ is defined by*

$$H_T(\vec{x}) = \sum_{\vec{u}=\vec{0}}^{\vec{\infty}} \Pr(T_1 = u_1, \ldots, T_t = u_t) x_1^{u_1} \ldots x_t^{u_t}, \tag{25}$$

where $\vec{u} = (u_1, \ldots, u_t)$ and $\vec{x}, \vec{0}$ and $\vec{\infty}$ are as in Definition 5.1.

The following theorem shows that computing the null distribution of block rank statistics essentially reduces to computing the null distribution within blocks and multiplication.

Theorem 5.4 *The generating function of the vector $T = (T_1, \ldots, T_t)$ factors under $H_0 : \gamma_1 = \ldots = \gamma_t = 0$ as*

$$H_T(\vec{x}) = \prod_{i=1}^{b} H_{R_i}(\vec{x}), \tag{26}$$

where $\vec{x}$ as in Definition 5.1.

Proof: Let $\vec{u} = (u_1, \ldots, u_t), \vec{v} = (v_1, \ldots, v_t), \vec{0} = (0, \ldots, 0)$ and $\vec{\infty} = (\infty, \ldots, \infty)$, which are vectors of length t and let $T_j^s = \sum_{i=1}^{s} R_{ij}, 1 \leq s \leq b$. Since we assume that there is no block effect, R_{ij} and R_{kj} are mutually independent for all $i \neq k, i = 1, \ldots, b, k = 1, \ldots, b$ and therefore T_j^{s-1} and R_{sj} are mutually independent too.

$$
\begin{aligned}
H_T(\vec{x}) &= \sum_{\vec{u}=\vec{0}}^{\vec{\infty}} \Pr(T_1 = u_1, \ldots, T_t = u_t) \, x_1^{u_1} \ldots x_t^{u_t} \\
&= \sum_{\vec{u}=\vec{v}}^{\vec{\infty}} \sum_{\vec{v}=\vec{0}}^{\vec{\infty}} \Pr(T_1^{b-1} = u_1 - v_1, \ldots, T_t^{b-1} = u_t - v_t) \\
&\quad \Pr(R_{b1} = v_1, \ldots, R_{bt} = v_t) \, x_1^{u_1} \ldots x_t^{u_t} \\
&= \sum_{\vec{v}=\vec{0}}^{\vec{\infty}} \sum_{\vec{u}=\vec{v}}^{\vec{\infty}} \Pr(T_1^{b-1} = u_1 - v_1, \ldots, T_t^{b-1} = u_t - v_t) \, x_1^{u_1-v_1} \ldots x_t^{u_t-v_t} \\
&\quad \Pr(R_{b1} = v_1, \ldots, R_{bt} = v_t) \, x_1^{v_1} \ldots x_t^{v_t} \\
&= \sum_{\vec{v}=\vec{0}}^{\vec{\infty}} \Pr(R_{b1} = v_1, \ldots, R_{bt} = v_t) \, x_1^{v_1} \ldots x_t^{v_t} \\
&\quad \sum_{\vec{u}=\vec{0}}^{\vec{\infty}} \Pr(T_1^{b-1} = u_1, \ldots, T_t^{b-1} = u_t) \, x_1^{u_1} \ldots x_t^{u_t} \\
&= H_{R_b}(\vec{x}) H_{T^{b-1}}(\vec{x}).
\end{aligned}
$$

Repeating this argument, we obtain (26). $\qquad\square$

Expansion of (26) yields the simultaneous distribution of $T_1, \ldots, T_t$. Since we are interested in the distribution of $S = \sum_{j=1}^{t} T_j^2$, it suffices to replace all terms of the form $k x_1^{u_1} \ldots x_t^{u_t}$ by $k x^U$ with $U = \sum_{j=1}^{t} u_j^2$. Let $k^* x^{U^*}$ be a term of the new expression, then $\Pr(S = U^*) = k^*$. So we are able to determine the distribution of S and therefore also of $c_1 S + c_2$.

The problem is that the expansion of (26) might cost a lot of memory space, due to the length of the expressions that arise. As a solution to this problem we introduce a recursion that is based on the equiprobability of the values of $T_1, \ldots, T_t$ and $R_{i1}, \ldots, R_{it}$ under H_0. We define the function $H_{T^s}^*(\vec{x})$:

Definition 5.5 *Let* $T_j^s = \sum_{i=1}^{s} R_{ij}$, $1 \le s \le b$, *let* $T^s = (T_1^s, \ldots, T_t^s)$, *let* $\vec{u} = (u_1, \ldots, u_t)$ *and let* $d(\vec{u})$ *be the number of distinct permutations of* $\vec{u}$, *then*

$$H_{T^s}^*(\vec{x}) = \sum_{u_1 \ge \ldots \ge u_t} d(\vec{u}) \Pr(T_1^s = u_1, \ldots, T_t^s = u_t)\, x_1^{u_1} \ldots x_t^{u_t}, \qquad (27)$$

where $\vec{x}$ *is as in Definition 5.1.*

Since all treatment effects are assumed to be equal under H_0, the coefficients in $H_{T^s}^*(\vec{x})$ equal the probability that T equals one of the permutations of $\vec{u}$. We present a recursive formula for computing $H_{T^s}^*(\vec{x})$:

Theorem 5.6 *Let* T_j^s *and* T^s *be as in Definition 5.5, then under* $H_0 : \gamma_1 = \ldots = \gamma_t = 0$

$$H_{T^s}^*(\vec{x}) = I\left(H_{T^{s-1}}^*(\vec{x})\, H_{R_s}(\vec{x}) \right), \qquad (28)$$

where I *is the linear operator defined by* $I(c x_1^{u_1} \ldots x_t^{u_t}) = c x_1^{\max(u_1,\ldots,u_t)} \ldots x_t^{\min(u_1,\ldots,u_t)}$.

Proof: As noted before, T_j^{s-1} and R_{sj} are mutually independent for all $i \ne k, i = 1, \ldots, b, k = 1, \ldots, b$. Let $\vec{x}, \vec{v}, \vec{0}$ and $\overline{\infty}$ be as in Definition 5.1 and let $d(\vec{u})$ be as in Definition 5.5.

$$
\begin{aligned}
H_{T^{s-1}}^*(\vec{x})\, H_{R_s}(\vec{x}) &= \sum_{u_1 \ge \ldots \ge u_t} d(\vec{u}) \Pr(T_1^{s-1} = u_1, \ldots, T_t^{s-1} = u_t)\, x_1^{u_1} \ldots x_t^{u_t} \\
&\qquad \sum_{\vec{v}=\vec{0}}^{\overline{\infty}} \Pr(R_{s1} = v_1, \ldots, R_{st} = v_t)\, x_1^{v_1} \ldots x_t^{v_t} \\
&= \sum_{\substack{\vec{v}=\vec{0} \\ u_1 \ge \ldots \ge u_t}}^{\overline{\infty}} d(\vec{u}) \Pr(T_1^s = u_1 + v_1, \ldots, T_t^s = u_t + v_t)\, x_1^{u_1+v_1} \ldots x_t^{u_t+v_t}.
\end{aligned}
$$

We let $(\tilde{u}_1, \ldots, \tilde{u}_t)$ be a tuple for which $\tilde{u}_1 \ge \ldots \ge \tilde{u}_t$. Applying I to the entire polynomial has the effect that those terms $c x_1^{u_1+v_1} \ldots x_t^{u_t+v_t}$ for which $(u_1 + v_1, \ldots, u_t + v_t)$ is a permutation of $(\tilde{u}_1, \ldots, \tilde{u}_t)$ sum up to one term. The sum of coefficients is the probability that T^s equals one of the permutations of $\tilde{u}_1 \ge \ldots \ge \tilde{u}_t$. Hence,

$$
\begin{aligned}
I\left(H_{T^{s-1}}^*(\vec{x})\, H_{R_s}(\vec{x}) \right) &= \sum_{\tilde{u}_1 \ge \ldots \ge \tilde{u}_t} d(\vec{\tilde{u}}) \Pr(T_1^s = \tilde{u}_1, \ldots, T_t^s = \tilde{u}_t)\, x_1^{\tilde{u}_1} \ldots x_t^{\tilde{u}_t} \\
&= H_{T^s}^*(\vec{x}). \qquad\qquad \square
\end{aligned}
$$

Example

Let $H_{T^1}^*(\vec{x}) = \frac{1}{2}(x_1^3 x_2^2 x_3 + x_1^2 x_2^2 x_3^2)$ and $H_{R_2}(\vec{x}) = \frac{1}{3}(x_1^3 x_2 x_3 + x_1 x_2^3 x_3 + x_1 x_2 x_3^3)$, then

$$
\begin{aligned}
H_{T^2}^*(\vec{x}) &= I\left(\frac{1}{6}(x_1^6 x_2^3 x_3^2 + x_1^4 x_2^5 x_3^2 + x_1^4 x_2^3 x_3^4 + x_1^5 x_2^3 x_3^3 + x_1^3 x_2^5 x_3^3 + x_1^3 x_2^3 x_3^5)\right) \\
&= \frac{1}{6}(x_1^6 x_2^3 x_3^2 + x_1^5 x_2^4 x_3^2 + 3\, x_1^5 x_2^3 x_3^3 + x_1^4 x_2^4 x_3^3).
\end{aligned}
$$

Using $H_{T^s}^*(\vec{x})$ and I instead of $H_T(\vec{x})$ we gain time and memory, because $H_{T^s}^*(\vec{x})$ contains much less terms than $H_{T^s}(\vec{x})$. Since the value of the test statistic is the same for all permutations of $(u_1, \ldots, u_t)$, $H_{T^b}^*(\vec{x})$ suffices for computing the null distribution of the test statistic.

b	$t = 3$	$t = 4$	$t = 5$
2	0.12	0.68	11.00
4	0.47	6.53	168.22
7	1.53	33.22	
10	3.43	99.30	
20	19.17		
30	56.57		
40	124.60		

Table 1: Computing times in seconds on a SUNSparc 5 for the Friedman test

5.3 Applications

5.3.1 The Friedman test

A well-known block rank test is the Friedman (1937) test. For the Friedman test we have $\{a_i(1), \ldots, a_i(t)\} = \{1, \ldots, t\}$ for all i. The Friedman test statistic is:

$$
\begin{aligned}
Q &= \frac{12}{bt(t+1)} \sum_{j=1}^{t} \left(T_j - \frac{b(t+1)}{2} \right)^2 \\
&= \frac{12}{bt(t+1)} \sum_{j=1}^{t} \left(T_j^2 + \frac{b^2(t+1)^2}{4} - b(t+1)T_j \right) \\
&= c_1 S + 3b(t+1) - \frac{12}{t} \sum_{j=1}^{t} T_j \\
&= c_1 S + 3b(t+1) - \frac{12}{t} \frac{bt(t+1)}{2} \\
&= c_1 S + c_2,
\end{aligned}
\tag{29}
$$

with $c_1 = \frac{12}{bt(t+1)}$ and $c_2 = -3b(t+1)$. So Q is a statistic of the form we dealt with in the previous section. Let T_F be the vector of treatment totals for this test. By using (26) and (24) we obtain:

$$
H_{T_F}(\vec{x}) = \prod_{i=1}^{b} H_{R_i}(\vec{x}) = (H_{R_i}(\vec{x}))^b = \left(\frac{1}{t!} \sum_{p=1}^{t!} x_1^{\sigma_p(1)} \ldots x_t^{\sigma_{ip}(t)} \right)^b .
\tag{30}
$$

5.3.2 Computing times

In Table 1 we give computing times for the Friedman test. One may observe that the computing times increase very quickly as the number of treatments increases. This is because the number of configurations within a block increases by a factor $t + 1$, when we add one treatment to t treatments.

5.4 The Cochran test

The Cochran and Cox (1957) test is an example of a test where ties occur in a natural way. Consider an experiment for which the observations can have only two values: 0 and 1. Here, 0 stands for "no success" and 1 for "success". We deal with blocks and for each observation within a block the rank score does not depend on the scores of the other observations within that block. Let $T_j = \sum_{i=1}^{b} R_{ij}$, $B_i = \sum_{j=1}^{t} R_{ij}$ and $E = \sum_{i=1}^{b} B_i$. If we wish to compare the treatments it suffices to condition on the number of blocks b_j with block total $j, j = 0, \ldots, t$, since a score

b^*	$t = 3$	$t = 4$	$t = 5$	$t = 6$
1	0.10	0.65	3.62	28.00
2	0.25	1.85	19.87	246.37
3	0.38	3.90	55.75	
4	0.62	8.08	133.23	
10	4.48	130.88		
20	27.20			
30	84.52			

Table 2: Computing times in seconds on a SUNSparc 5 for the Cochran test, $b_j = b^*, j = 0, \ldots t$.

equals either 0 or 1. Let us first define Cochran's test statistic C.

$$
\begin{aligned}
C &= t(t-1)\frac{\sum_{j=1}^{t}\left(T_j - \frac{E}{t}\right)^2}{\sum_{i=1}^{b} B_i(t - B_i)} \\
&= t(t-1)\frac{\sum_{j=1}^{t}\left(T_j - \frac{E}{t}\right)^2}{\sum_{j=0}^{t} j(t-j)b_j} \\
&= t(t-1)\frac{\sum_{j=1}^{t}\left(T_j - \frac{E}{t}\right)^2}{t\sum_{j=1}^{t} jb_j - \sum_{j=1}^{t} j^2 b_j} \\
&= (t-1)\frac{t\sum_{j=1}^{t} T_j^2 - E^2}{tE - \sum_{j=1}^{t} j^2 b_j} \\
&= c_1 S + c_2,
\end{aligned}
\tag{31}
$$

with $S = \sum_{j=1}^{t} T_j^2, c_1 = \frac{t(t-1)}{tE - \sum_{j=1}^{t} j^2 b_j}$ and $c_2 = \frac{-E^2(t-1)}{tE - \sum_{j=1}^{t} j^2 b_j}$.

So C is linear in S and we therefore use Theorem 5.4 with scores 0 and 1. We observe that $H_{R_i}(\vec{x})$ has to be computed for all values of j for which $b_j > 0, j = 0 \ldots, t$. We apply Theorem 5.6 and (28) to speed up the computations.

5.5 Computing times for the Cochran test

For the case $t = 2$ we used the expansion of formula (26). In this case the computing times are small: less than 1.5 sec. for $b_1 \leq 1000$ and less than 20 sec. if $b_1 \leq 8000$. If b_1 is larger than 8000, memory problems occur. In Table 2 we give computing times for the Cochran test for the cases $t > 2$. We used (28) for the cases $t > 2$. We did not apply I after every step in this recursion, but after each third step. This is because the operator I is not so fast and it has not that much effect on the length of the expression when it was used just a few steps before in the recursion. We present tables for the cases $b_j = b_k = b^*, 0 \leq j, k \leq t$, so the number of blocks $b = (t+1)b^*$. We should note that this is generally not the worst case. Bad cases will usually occur as b_j is relatively large for $j \approx \frac{1}{2}t$.

A   Proof of Theorem 4.3

In order to prove the theorem we have to define the test statistics more formally. We first define

$$
\xi = (\xi_{Z_1} \ldots \xi_{Z_i}) \equiv \{Z_1 < \ldots < Z_i \leq T\} \cap \{Z_{i+1} > T\}, \ldots, \{Z_{n+m} > T\},
$$

where $\xi_{Z_j} = 0$ if Z_j is an X-observation and $\xi_{Z_j} = 1$ if Z_j is an Y-observation. So when we know ξ, we know how many X's and Y's are smaller or equal to T and we know how they are configurated.

Let $\bar{r}_m(\xi)$ and $\bar{r}_n(\xi)$ be the number of 0's and 1's in ξ, respectively, then we introduce two classes

$$\mathcal{A}_{a,b} = \{\xi : \bar{r}_m(\xi) = a, \bar{r}_n(\xi) = b\} \quad \text{and}$$

$$\mathcal{A} = \bigcup_{a=1}^{m} \bigcup_{b=1}^{n} \mathcal{A}_{a,b}$$

Now we formally define the statistic U_c and the Mann-Whitney statistic $M_{a,b}$:

$$U_c : \mathcal{A} \to \mathbf{N}, \qquad U_c(\xi) = \sum_{j<k} I_{\{\xi_{Z_j}=1, \xi_{Z_k}=0\}} + (m - \bar{r}_m(\xi))\bar{r}_n(\xi)$$

$$M_{a,b} : \mathcal{A}_{a,b} \to \mathbf{N}, \quad M_{a,b}(\xi) = \sum_{j<k} I_{\{\xi_{Z_j}=1, \xi_{Z_k}=0\}},$$

where I_E is the indicator function of the event E. We define $\mu : \mathcal{A} \to [0, 1]$ as the normalized counting measure. Then,

$$\Pr(U_c = u | \bar{r}_m = a, \bar{r}_n = b) = \frac{\Pr(U_c = u, \bar{r}_m = a, \bar{r}_n = b)}{\Pr(\bar{r}_m = a, \bar{r}_n = b)} \tag{32}$$

and

$$\Pr(U_c = u, \bar{r}_m = a, \bar{r}_n = b) = \sum_{\xi \in \mathcal{A} : U_c(\xi) = u} \mu(\xi, \bar{r}_m = a, \bar{r}_n = b)$$

$$= \sum_{\substack{\xi \in \mathcal{A}_{a,b} : \\ M_{a,b}(\xi) = u - (m-a)b}} \mu(\xi, \bar{r}_m = a, \bar{r}_n = b), \tag{33}$$

because $\mu(\xi, \bar{r}_m = a, \bar{r}_n = b) = 0$ if $\xi \notin \mathcal{A}_{a,b}$. Since we have m X's and n Y's, there exist $\binom{m}{a}\binom{n}{b}$ equiprobable realizations of $X_1, \ldots, X_m$ and $Y_1, \ldots, Y_n$ that lead to the same value of $\xi \in \mathcal{A}_{a,b}$. With $N = m + n$ and $B = a + b$ we denote such a realization by $(w_1, \ldots, w_B, w_{B+1}, \ldots, w_N)$. Then for $\xi \in \mathcal{A}_{a,b}$,

$$\mu(\xi, \bar{r}_m = a, \bar{r}_n = b)$$

$$= \binom{m}{a}\binom{n}{b} \Pr(w_1 < \ldots < w_B \leq T, w_{B+1} > T, \ldots, w_N > T)$$

$$= \binom{m}{a}\binom{n}{b} \left(\int_{-\infty}^{T} \ldots \int_{-\infty}^{w_2} dF(w_1) \ldots dF(w_B) \right) \left(1 - F(T) \right)^{N-B} \tag{34}$$

$$= \frac{\binom{m}{a}\binom{n}{b}}{B!} \left(F(T) \right)^{B} \left(1 - F(T) \right)^{N-B}.$$

Substitution of (34) into (33) results in:

$$\Pr(U_c = u, \bar{r}_m = a, \bar{r}_n = b) = \sum_{\substack{\xi \in \mathcal{A}_{a,b} : \\ M_{a,b}(\xi) = u - (m-a)b}} \frac{\binom{m}{a}\binom{n}{b}}{B!} \left(F(T) \right)^{B} \left(1 - F(T) \right)^{N-B}. \tag{35}$$

We return to equality (32) and note that

$$\Pr(\bar{r}_m = a, \bar{r}_n = b) = \Pr(\bar{r}_m = a) \Pr(\bar{r}_n = b) = \binom{m}{a}\binom{n}{b} \left(F(T) \right)^{B} \left(1 - F(T) \right)^{N-B}. \tag{36}$$

Substitution of (35) and (36) into (32) completes the proof:

$$\Pr(U_c = u | \bar{r}_m = a, \bar{r}_n = b) = \sum_{\substack{\xi \in \mathcal{A}_{a,b} : \\ M_{a,b}(\xi) = u - (m-a)b}} \frac{1}{B!} = \Pr(M_{a,b} = u - (m-a)b). \tag{37}$$

$\square$

B Tables

As a result of the theorems in the previous sections, we present extended tables of critical values. For the two-sample rank statistics we only give tables for the balanced cases $n = m$ because of practical reasons. The confidence levels α are *two-sided*, except for the Friedman statistic for which only one-sided confidence levels exist. We give the dimensions of the new and existing tables and the references:

- **Freund-Ansari-Bradley**: $N \leq 80$, $N \leq 20$ in Ansari and Bradley (1960).

- **Friedman**: $t = 3, b \leq 25; t = 4, b \leq 20; t = 5, b \leq 12$.
  $t = 3, b \leq 15; t = 4, b \leq 8$ in Hollander and Wolfe (1973), $t = 5, b \leq 8; t = 6, b \leq 6$ in Odeh (1977).

- **Mood**: $N \leq 62$, $N \leq 20$ in Laubscher et al. (1968).

- **Klotz**: $N \leq 40$, $N \leq 20$ in Klotz (1962)

- **Van der Waerden**: $N \leq 32$, $N \leq 20$ in Van der Waerden (1957).

Since the exact distributions of the Halperin test and the Cochran test do not only depend on the sample sizes, but also on conditional arguments, it would be very paper-consuming to present tables for these tests. For the Halperin statistic, tables are available in Van der Laan and Van Putten (1987) for $N \leq 24$. We note that, with our method, we are able to compute p-values or critical values of the Halperin statistic within a few seconds for $N \leq 50$.

Packages for exact values
For all test statistics presented in this paper, we wrote Mathematica packages for computing exact p-values and critical values. These packages are available from the author.

Acknowledgement
I would like to thank Alessandro Di Bucchianico for correcting preliminary versions of this paper and for helpful discussions.

References

Ansari, A.R., and R.A. Bradley (1960). Rank sum tests for dispersion. *Annals of Mathematical Statistics*, **31**, 1174–1189.

Arrenberg, J. (1994). Natural ranks in the conditional Wilcoxon rank-sum test. *Computational Statistics Data Analysis*, **17**, 141–152.

Baglivo, J., M. Pagano, and C. Spino (1993). Symbolic computation of permutation distributions. In *Proceedings of the Statistical Computing Section*, pages 218–223.

Cochran, W.J., and G.M. Cox (1957). *Experimental Designs*. Wiley & Sons.

Euler, L. (1748). *Introductio in analysin infinitorum*, volume 1. Bousquet. English edition: *Introduction to analysis of the infinite*, Springer, 1988.

Freund, J.E., and A.R. Ansari (1957). Two-way rank sum tests for variance. Technical report, Virginia Polytechnic Institute Technical Report to Office of Ordnance Research and National Science Foundation.

Friedman, M. (1937). The use of ranks to avoid the assumption of normality implicit in the analysis of variance. *Journal of the American Statistical Association*, **32**, 675–701.

Gastwirth, J.L. (1965). Percentile modifications of two sample rank tests. *Journal of the American Statistical Association*, **63**, 692–706.

Gibbons, J.D., and S. Chakraborti (1992). *Nonparametric statistical inference*. Marcel Dekker.

Hájek, J. (1969). *Nonparametric statistics*. Holden-Day.

Halperin, M. (1960). Extension of the Wilcoxon-Mann-Whitney test to samples censored at the same fixed point. *Journal of the American Statistical Association*, **55**, 125–138.

Hollander, M., and D.A. Wolfe (1973). *Nonparametrical statistical methods*. John Wiley & Sons.

Kendall, M., and A. Stuart (1977). *The advanced theory of statistics, Volume 2*. Charles Griffin & Company ltd.

Kendall, W.S. (1993). Computer algebra in statistics. *Stat. Neerlandica*, **47**, 9–25.

Klotz, J.H. (1962). Nonparametric tests for scale. *Annals of Mathematical Statistics*, **33**, 495–512.

Laubscher, N.F., F.E. Steffens, and E.M. DeLange (1968). Exact critical values for Mood's distribution-free test statistic for dispersion and its normal approximation. *Technometrics*, **10**, 497–508.

Lehmann, E.L. (1975). *Nonparametrics: statistical Methods Based on ranks*. Holden-Day.

Mehta, C.R., N.R. Patel, and L.J. Wei (1987). Constructing exact significance tests with restricted randomization rules. *Biometrika*, **75(2)**, 295–302.

Mood, A.M. (1954). On the asymptotic efficiency of certain nonparametric two-sample tests. *Annals of Mathematical Statistics*, **25**, 514–522.

Odeh, R.E. (1977). Extended tables of the distribution of Friedman's S-statistic in the two-way layout. *Communications in Statistics - Simulation and Computation*, **B6**, 29–48.

Streitberg, B., and J. Röhmel (1986). Exact distributions for permutation and rank tests: an introduction to some recently published algorithms. *Statistical Software Newsletter*, **12**, 10–17.

Van de Wiel, M.A. (1996). *Exact distributions of nonparametric statistics using computer algebra*. Master's thesis, Eindhoven University of Technology, The Netherlands.

Van de Wiel, M.A., A. Di Bucchianico, and P. Van der Laan (1997). Symbolic computation and exact distributions of nonparametric test statistics. Technical Report COSOR 97-17, Eindhoven University of Technology, The Netherlands.

Van der Laan, P., and B. Van Putten (1987). Extensive tables of exact critical values of the nonparametric two-sample Halperin for censored samples, and comparison with the Normal approximation. *Communications in Statistics - Simulation and Computation*, **16(1)**, 37–53.

Van der Waerden, B.L. (1957). *Mathematische Statistik*. Springer-Verlag.

Wilcoxon, F. (1945). Individual comparisons by ranking methods. *Biometrics*, **1**, 80–83.

n	0.1	0.05	0.025	0.01	0.005	n	0.1	0.05	0.025	0.01	0.005
4	33	37	37	*	*	18	2322.50	2428.50	2516.50	2618.50	2686.50
5	59.25	65.25	67.25	71.25	*	19	2718.75	2838.75	2942.75	3058.75	3134.75
6	97.50	103.50	109.50	115.50	125.50	20	3159	3295	3411	3545	3633
7	151.75	161.75	169.75	179.75	185.75	21	3643.25	3797.25	3927.25	4077.25	4179.25
8	222	236	248	262	268	22	4173.50	4345.50	4493.50	4663.50	4775.50
9	312.25	330.25	346.25	364.25	374.25	23	4751.75	4945.75	5111.75	5299.75	5425.75
10	422.50	446.50	466.50	488.50	502.50	24	5382	5596	5782	5992	6134
11	556.75	586.75	612.75	640.75	660.75	25	6064.25	6302.25	6508.25	6740.25	6898.25
12	715	753	785	821	845	26	6802.50	7066.50	7290.50	7548.50	7722.50
13	901.25	949.25	987.25	1031.25	1061.25	27	7596.75	7886.75	8134.75	8418.75	8608.75
14	1119.50	1173.50	1221.50	1275.50	1309.50	28	8453	8769	9041	9353	9561
15	1365.75	1433.75	1489.75	1551.75	1593.75	29	9367.25	9713.25	10011.25	10351.25	10579.25
16	1648	1726	1792	1868	1918	30	10347.50	10723.50	11047.50	11417.50	11667.50
17	1966.25	2058.25	2134.25	2222.25	2280.25	31	11391.75	11799.75	12151.75	12553.75	12825.75

Table 3: Right critical values for the Mood scale test, $n = \frac{1}{2}N$

n	0.1	0.05	0.025	0.01	0.005	n	0.1	0.05	0.025	0.01	0.005
4	11	12	12	*	*	23	294.5	302.5	309.5	317.5	323.5
5	16.5	17.5	17.5	18.5	*	24	320	329	336	345	351
6	23	24	25	26	27	25	346.5	355.5	363.5	372.5	379.5
7	30.5	31.5	33.5	34.5	35.5	26	374	384	392	402	409
8	39	41	42	44	45	27	402.5	413.5	421.5	432.5	439.5
9	48.5	50.5	52.5	54.5	55.5	28	432	443	453	463	471
10	60	62	64	66	67	29	462.5	474.5	484.5	495.5	503.5
11	71.5	74.5	76.5	78.5	80.5	30	494	507	517	529	537
12	84	87	90	93	95	31	527.5	539.5	550.5	563.5	572.5
13	98.5	101.5	104.5	107.5	109.5	32	561	574	586	599	608
14	113	117	120	124	126	33	595.5	609.5	621.5	635.5	645.5
15	129.5	133.5	136.5	141.5	143.5	34	631	646	659	673	683
16	146	151	155	159	163	35	668.5	683.5	696.5	711.5	722.5
17	164.5	169.5	173.5	178.5	182.5	36	706	722	736	752	762
18	183	189	194	199	203	37	744.5	761.5	775.5	792.5	803.5
19	203.5	209.5	214.5	220.5	224.5	38	785	802	817	834	846
20	225	231	237	244	248	39	825.5	843.5	859.5	877.5	889.5
21	247.5	254.5	260.5	267.5	271.5	40	868	886	903	921	934
22	270	278	285	292	297						

Table 4: Right critical values for the Freund-Ansari-Bradley test, $n = \frac{1}{2}N$

n	0.05	0.02	0.01
5	2.612	3.082	3.311
6	2.877	3.295	3.685
7	3.116	3.631	3.968
8	3.399	3.925	4.269
9	3.638	4.243	4.624
10	3.872	4.517	4.934
11	4.094	4.784	5.228
12	4.305	5.037	5.514
13	4.509	5.281	5.787
14	4.705	5.516	6.048
15	4.894	5.742	6.300
16	5.077	5.960	6.543

Table 5: Right critical values for the Van der Waerden test, $n = \frac{1}{2}N$

n	0.05	0.02	0.01
4	4.149	4.149	*
5	5.122	5.229	5.582
6	6.155	6.445	6.777
7	7.285	7.612	8.078
8	8.448	8.903	9.305
9	9.616	10.002	10.594
10	10.797	11.338	11.881
11	11.965	12.555	13.162
12	13.127	13.757	14.421
13	14.292	14.966	15.686
14	15.450	16.167	16.938
15	16.607	17.365	18.186
16	17.761	18.559	19.428
17	18.912	19.750	20.665
18	20.061	20.937	21.897
19	21.208	22.121	23.125
20	22.352	23.301	24.348

Table 6: Right critical values for the Klotz test, $n = \frac{1}{2}N$

t	b	0.1	0.05	0.01	0.001
3	2	*	*	*	*
	3	18	18	*	*
	4	24	26	32	*
	5	26	32	42	50
	6	32	42	54	72
	7	38	50	62	86
	8	42	50	72	98
	9	50	56	86	114
	10	50	62	96	126
	11	56	72	104	146
	12	62	78	114	152
	13	62	86	122	162
	14	72	86	128	186
	15	74	96	134	194
	16	78	104	150	216
	17	86	104	158	222
	18	86	114	162	234
	19	96	122	182	254
	20	98	126	186	266
	21	104	128	194	278
	22	104	134	200	296
	23	114	146	216	302
	24	122	150	222	314
	25	122	152	224	338

t	b	0.1	0.05	0.01	0.001
4	2	20	20	*	*
	3	33	37	45	*
	4	42	52	64	74
	5	53	65	83	105
	6	64	76	102	128
	7	75	91	123	157
	8	84	102	140	184
	9	93	115	161	211
	10	106	128	178	242
	11	115	141	197	267
	12	126	154	216	296
	13	133	169	235	323
	14	148	180	254	352
	15	157	193	273	377
	16	168	208	292	404
	17	179	221	313	433
	18	190	232	328	458
	19	201	249	349	489
	20	208	260	370	512

t	b	0.1	0.05	0.01	0.001
5	2	36	38	40	*
	3	56	64	76	86
	4	76	88	112	132
	5	96	112	146	180
	6	116	136	178	228
	7	136	160	212	274
	8	154	184	246	320
	9	174	208	280	368
	10	194	232	312	414
	11	214	256	346	460
	12	232	280	378	506

Table 7: Right critical values for the sum of squares of treatment totals (S) in the Friedman test, b = number of blocks, t = number of treatments.