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

Citation for published version (APA):
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/vdiewl

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 structure, 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), Bagliovo 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 Chakrabarti (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

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

where $a : \{1, \ldots, N\} \rightarrow \mathbb{R}$ is said to be a rank score function, $m + n = N$, $Z_{\ell} = 1$ if the $\ell$th 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 $H_0$

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

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 $\ell$th rank score for test statistic $T_N$ by $a_{T_N}(\ell)$. Substitution 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 $\mu_X$ and $\mu_Y$ 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:

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

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

where $\Phi^{-1}$ denotes the inverse of the standard normal distribution function.

2.3.2 Two-sample scale tests

We wish to test the null hypothesis

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

where $\sigma_X$ and $\sigma_Y$ 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:

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

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

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

where $\Phi^{-1}$ 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 $s$th and lower $r$th 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 incorporating knowledge of any characteristic of the underlying distribution functions (e.g. asymmetry) into the test statistic. Let $S = \lceil N_s \rceil$ and $R = \lceil N_r \rceil$, where $\lceil N_q \rceil$ denotes the largest integer not exceeding the number $q N$, $q = r, s$. 

3
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.

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

(8)

and

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

(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 $Q_{15} = 7174454$. 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 $N \leq 20$ 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 $cx^by^k$, 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: \( \{ \frac{81}{4}, \frac{49}{4}, \frac{25}{4}, \frac{9}{4}, \frac{1}{4}, \frac{1}{4}, \frac{1}{4}, \frac{1}{4}, \frac{1}{4}, \frac{0}{4} \} \).

**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 \( \Pr(T_N = t) = c \), \( l \leq t \leq u \).

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

\[
\{ \frac{81}{4}, \frac{49}{4}, \frac{25}{4}, \frac{9}{4}, \frac{1}{4}, \frac{1}{4}, \frac{1}{4}, \frac{1}{4}, \frac{1}{4}, \frac{0}{4} \}
\]

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 \( \frac{3}{4} - \frac{1}{4} \)). 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:

\[
P = 1 + \left( 2x^{\frac{4}{4}} + 2x^{\frac{4}{4}} + 2x^{\frac{4}{4}} \right) y + \left( x^{\frac{4}{4}} + 4x^{\frac{4}{4}} + x^{\frac{4}{4}} + 4x^{\frac{4}{4}} + 4x^{\frac{4}{4}} + x^{\frac{4}{4}} \right) y^2
+ \left( 2x^{\frac{4}{4}} + 2x^{\frac{4}{4}} + 2x^{\frac{4}{4}} + 8x^{\frac{4}{4}} + 2x^{\frac{4}{4}} + 2x^{\frac{4}{4}} + 2x^{\frac{4}{4}} \right) y^3
+ \left( x^{37} + 4x^{45} + 4x^{51} + x^{53} + 4x^{55} + x^{65} \right) y^4 + \left( 2x^{\frac{4}{4}} + 2x^{\frac{4}{4}} + 2x^{\frac{4}{4}} \right) y^5 + x^{\frac{11}{4}} y^6
\]

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 '\( x^{\frac{4}{4}} y^5 \)' in \( P \) will not not contribute to this coefficient. Therefore, we may remove '1' and '\( x^{\frac{4}{4}} y^5 \)' from \( P \).

\[
P = \left( 2x^{\frac{4}{4}} + 2x^{\frac{4}{4}} + 2x^{\frac{4}{4}} \right) y + \left( x^{\frac{4}{4}} + 4x^{\frac{4}{4}} + x^{\frac{4}{4}} + 4x^{\frac{4}{4}} + 4x^{\frac{4}{4}} + x^{\frac{4}{4}} \right) y^2
+ \left( 2x^{\frac{4}{4}} + 2x^{\frac{4}{4}} + 2x^{\frac{4}{4}} + 8x^{\frac{4}{4}} + 2x^{\frac{4}{4}} + 2x^{\frac{4}{4}} + 2x^{\frac{4}{4}} \right) y^3
+ \left( x^{37} + 4x^{45} + 4x^{51} + x^{53} + 4x^{55} + x^{65} \right) y^4 + \left( 2x^{\frac{4}{4}} + 2x^{\frac{4}{4}} + 2x^{\frac{4}{4}} \right) y^5.
\]
Let us now consider the term of degree 4 in \( P \), denoted by \( P[y^4] \). Thus,

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

Let \( cx^k \) be a term of the coefficient of \( y^t \). Because \( m = 5 \), precisely one more score has to be added to \( b \). The maximum \( M^+ \) of the scores that are left is \( \frac{5}{4} \), the minimum \( M^- \) is \( \frac{3}{4} \). 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 \( cx^k \) 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 \( x^{37} + 4x^{45} + 4x^{51} + x^{53} \), which results in:

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

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 \( c(\frac{4}{4})/\binom{N}{u} = 4/252 = 0.016 \), where \( \binom{4}{4} \) 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:

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

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, \( \frac{2}{4} \), and multiply \( P \) by \( 1 + x^\frac{2}{4} y \). 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, \( \alpha \in [0,1] \), \( v \): integer, \([l,u]\): interval that contains the right critical value, rank scores: reals.

**Output:** one-sided right \( \alpha \) 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 \( P = \prod_{i=1}^{n} (1 + a(x^i)y) \).
3. \( j \rightarrow v, q \rightarrow 0 \).
4. Let \( k \) be the exponent of \( y \) in an arbitrary term of \( P \), then \( P \rightarrow P - I \), 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 \( P \), \( k \rightarrow r \).
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 \( P \rightarrow P - J_1 y^k \), 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 $b + M^− ≥ u$ and add the coefficients of these terms.

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

10. $q → q + t$.

11. $P → P - J_2y^k$.

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

13. $j → j + 1$.

14. If $j ≤ N$, $P → P (1 + x^{a(j)}y)$ and go to step 4.

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

16. Let $b(i)$ be the exponent of $x$ in the $i$th term of $\frac{P}{y^m}$. Furthermore, let $c(i)$ be the coefficient in the $i$th term of $\frac{P}{y^m}$, where $P$ is given in increasing order of $b(i)$ and let $d(i) = \sum_{j≥i} c(j)$. Now determine the first term $i^*$ for which $d(i^*) ≤ (\frac{N}{m})(α - \Pr(T_N ≥ u))$. Then the one-sided right 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 initialization 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 $m - (N - j) ≤ k ≤ m$.

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 \( P[y^n] \) 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

\[
d(i) \leq \binom{N}{m} (\alpha - \Pr(T_N \geq u)) \iff \sum_{j \geq i} d(j)/\binom{N}{m} + \Pr(T_N \geq u) \leq \alpha
\]

\[
\iff \Pr(u > T_N \geq b(i)) + \Pr(T_N \geq u) \leq \alpha \iff \Pr(T_N \geq b(i)) \leq \alpha.
\]

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

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

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

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

The test statistic \( U_c \) is given by

\[
U_c = M(m-r_m,n-r_n) + r_m(n-r_n),
\]

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

$$H_{U_c}(x|v) = \sum_{u=0}^{\infty} \Pr(U_c = u|v_m + v_n = r) x^u.$$  \hspace{1cm} (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-Whitney statistic for sample sizes $a$ and $b$. Then, under $H_0 : F = G$, with $\bar{v}_m = m - r_m$ and $\bar{v}_n = n - r_n$,

$$\Pr(U_c = u|\bar{v}_m = a, \bar{v}_n = b) = \Pr(M(a, b) = u + (m - a)b).$$  \hspace{1cm} (12)

Proof: see Appendix A.

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

$$\Psi(P, k) = P[[y^k]] = \sum_{\ell} c(\ell) x^\ell (y^k).$$  \hspace{1cm} (13)

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

$$H_{U_c}(x|v) = \sum_{i=0}^{m} \frac{1}{i! (m-n)} x^i (1 + x^i y)^{i} \Psi \left( \frac{m+n-r}{m-n} \prod_{\ell=1}^{m} \left( 1 + x^\ell y \right), m - i \right),$$  \hspace{1cm} (14)

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

Proof: Using Lemma 4.2 we have

$$\Pr(U_c = u|v_m + v_n = r) = \sum_{i=0}^{m} \Pr(U_c = u|v_m + v_n = r, r_m = i) \Pr(r_m = i|v_m + v_n = r).$$  \hspace{1cm} (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

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

$$= \frac{(n)_i (m+n)_r-i}{(m+n)_r} \left( 1 - F(T) \right)^{n-r+i} \left( 1 - F(T) \right)^{i} = \frac{(m+n)_r}{(m+n)_r - i},$$  \hspace{1cm} (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: \( M(m, n) = W(m, n) - \frac{m}{2}m(m + 1) \) to write

\[
\Pr(U_c = u|r_m + r_n = r, r_m = i) = \Pr(U_c = u|r_n = m-i, r_n = n-r + i) = \Pr(M(m - i, n - r + i) = u + g(i)) = \frac{1}{\binom{m+n-r}{m-i}} h(u - g(i), m - i, n - r + i),
\]

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

\[
H_{U_c}(x|r) = \sum_{u=0}^{\infty} \sum_{i=0}^{m} \binom{m}{i} \binom{n}{r-i} \binom{m+n-r}{m-i} \sum_{u=0}^{\infty} h(w, m - i, n - r + i)x^u = \sum_{i=0}^{m} \frac{1}{i!} \frac{1}{(m+n)} \sum_{u=0}^{\infty} h(w, m - i, n - r + i)x^u,
\]

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

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

\[
\sum_{u=0}^{\infty} h(w, m - i, n - r + i)x^u = \Psi \left( \prod_{i=1}^{m+n-r} (1 + x^i) \right),
\]

(19)

After substituting (19) into (18) we obtain the desired result.

\[
\square
\]

**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.

\[
\sum_{u=0}^{\infty} \Pr(M(a, b) = k)x^k = \frac{1}{\binom{a+b}{a}} \prod_{i=1}^{a+b} (1 - x^i).
\]

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 Römel (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:

\[ Y_{ij} = \mu + \beta_i + \gamma_j + \epsilon_{ij}, \]

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

\[ H_0 : \gamma_1 = \ldots = \gamma_t = 0 \]

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 $i$th block belonging to the $j$th 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 $S = \sum_{i=1}^{b} T_j^2$. 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

\[ a_i : \{1, \ldots, t\} \to \mathbb{R}. \]  

(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

\[ HR_i(\vec{x}) = \sum_{\vec{v}} \Pr(R_{i1} = v_1, \ldots, R_{it} = v_t) x_1^{v_1} \ldots x_t^{v_t}, \]

(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 $S_t$ be the symmetric group of all these permutations of $\{1, \ldots, t\}$. Under $H_0 : \gamma_1 = \ldots = \gamma_t = 0$ we have

\[ HR_i(\vec{x}) = \frac{1}{t!} \sum_{\sigma \in S_t} x_1^{a_i(\sigma(1))} \ldots x_t^{a_i(\sigma(t))}. \]

(24)

Proof: We know that $\Pr(R_{i1} = v_1, \ldots, R_{it} = v_t) = 0$ if $\{v_1, \ldots, v_t\} \neq \{a_1(1), \ldots, a_t(t)\}$. Under $H_0$, we have $\Pr(R_{ij} = a_i(\ell)) = \Pr(R_{ik} = a_k(\ell))$ for all $j \neq k$, for all $\ell$ and for all $i$. Therefore, we have $\Pr(R_{i1} = a_i(\sigma(1)), \ldots, R_{it} = a_t(\sigma(t))) = \frac{1}{t!}$ for all $\sigma \in S_t$ and the result follows. \qed

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

11
**Definition 5.3** The probability generating function of the row vector of treatment totals \( T = (T_1, \ldots, T_t) \) is defined by

\[
H_T(\bar{x}) = \sum_{\bar{u}=0}^{\infty} \Pr(T_1 = u_1, \ldots, T_t = u_t) x_1^{u_1} \cdots x_t^{u_t},
\]

where \( \bar{u} = (u_1, \ldots, u_t) \) and \( \bar{x}, \bar{0} \) and \( \infty \) 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

\[
H_T(\bar{x}) = \prod_{i=1}^{b} H_{R_i}(\bar{x}),
\]

where \( \bar{x} \) as in Definition 5.1.

Proof: Let \( \bar{u} = (u_1, \ldots, u_t), \bar{v} = (v_1, \ldots, v_t), \bar{0} = (0, \ldots, 0) \) and \( \infty = (\infty, \ldots, \infty) \), which are vectors of length \( t \) and let \( T_i^j = \sum_{s=1}^{j} R_{ij}, 1 \leq s \leq b \). 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 \( T_i^{s-1} \) and \( R_{ij} \) are mutually independent too.

\[
H_T(\bar{x}) = \sum_{\bar{u}=0}^{\infty} \Pr(T_1 = u_1, \ldots, T_t = u_t) x_1^{u_1} \cdots x_t^{u_t}
= \sum_{\bar{u}=0}^{\infty} \sum_{\bar{v}=0}^{\infty} \Pr(T_1^{b-1} = u_1 - v_1, \ldots, T_t^{b-1} = u_t - v_t) \\
Pr(R_{b1} = v_1, \ldots, R_{bt} = v_t) x_1^{u_1 - v_1} \cdots x_t^{u_t - v_t}
= \sum_{\bar{v}=0}^{\infty} \sum_{\bar{u}=0}^{\infty} \Pr(T_1^{b-1} = u_1 - v_1, \ldots, T_t^{b-1} = u_t - v_t) \\
Pr(R_{b1} = v_1, \ldots, R_{bt} = v_t) x_1^{v_1} \cdots x_t^{v_t}
= \sum_{\bar{v}=0}^{\infty} \Pr(R_{b1} = v_1, \ldots, R_{bt} = v_t) x_1^{v_1} \cdots x_t^{v_t}
= H_{R_1}(\bar{x}) H_{T^{b-1}}(\bar{x}).
\]

Repeating this argument, we obtain (26).

Expansion of (26) yields the simultaneous distribution of \( T_1, \ldots, T_t \). Since we are interested in the distribution of \( S = \sum_{j=1}^{t} T_j^* \), it suffices to replace all terms of the form \( k x_1^{u_1} \cdots x_t^{u_t} \) by \( k x^U \) with \( U = \sum_{j=1}^{t} u_j^* \). 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 \( H_{T_i}^T(\bar{x}) \):
**Definition 5.5** Let $T_j^s = \sum_{i=1}^s R_{ij}$, $1 \leq s \leq b$, let $T^s = (T_1^s, \ldots, T_t^s)$, let $\bar{u} = (u_1, \ldots, u_t)$ and let $d(\bar{u})$ be the number of distinct permutations of $\bar{u}$, then

$$H_{T^s}(\bar{x}) = \sum_{u_1 \geq \ldots \geq u_t} d(\bar{u}) \Pr(T_1^s = u_1, \ldots, T_t^s = u_t) x_1^{u_1} \ldots x_t^{u_t}, \quad (27)$$

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

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

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

$$H_{T^s}(\bar{x}) = I \left( H_{T^s -}(\bar{x}) H_{R_s}(\bar{x}) \right), \quad (28)$$

where $I$ is the linear operator defined by $I(cx_1^{u_1} \ldots x_t^{u_t}) = cx_1^{\max(u_1, \ldots, u_t)} \ldots x_t^{\min(u_1, \ldots, u_t)}$.

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

$$H_{T^s -}(\bar{x}) H_{R_s}(\bar{x}) = \sum_{u_1 \geq \ldots \geq u_t} d(\bar{u}) \Pr(T_1^{s-1} = u_1, \ldots, T_t^{s-1} = u_t) x_1^{u_1} \ldots x_t^{u_t}$$

$$= \sum_{u_1 \geq \ldots \geq u_t} \Pr(R_s = v_1, \ldots, R_t = v_t) x_1^{v_1} \ldots x_t^{v_t}$$

$$= \sum_{\bar{v}} \sum_{u_1 \geq \ldots \geq u_t} d(\bar{u}) \Pr(T_1^{s} = u_1 + v_1, \ldots, T_t^{s} = u_t + v_t) x_1^{u_1 + v_1} \ldots x_t^{u_t + v_t}.$$

We let $(\tilde{u}_1, \ldots, \tilde{u}_t)$ be a tuple for which $\tilde{u}_1 \geq \ldots \geq \tilde{u}_t$. Applying $I$ to the entire polynomial has the effect that those terms $cx_1^{u_1 + v_1} \ldots x_t^{u_t + v_t}$, 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 $\tilde{u}$ equal to $\bar{u}$. Hence,

$$I \left( H_{T^s -}(\bar{x}) H_{R_s}(\bar{x}) \right) = \sum_{\tilde{u}_1 \geq \ldots \geq \tilde{u}_t} d(\tilde{u}) \Pr(T_1^{s} = \tilde{u}_1, \ldots, T_t^{s} = \tilde{u}_t) x_1^{\tilde{u}_1} \ldots x_t^{\tilde{u}_t}$$

$$= H_{T^s}(\bar{x}). \quad \Box$$

**Example**

Let $H_{T^s}(\bar{x}) = \frac{1}{6}(x_1^2 x_2^2 x_3 + x_1^3 x_2 x_3 + x_1 x_2 x_3)$ and $H_{R_{ij}}(\bar{x}) = \frac{1}{6}(x_1^3 x_2 x_3 + x_1 x_2 x_3)$, then

$$H_{T^s}(\bar{x}) = I \left( \frac{1}{6}(x_1^3 x_2 x_3 + x_1^2 x_2 x_3 + x_1 x_2^2 x_3 + x_1^3 x_2^2 x_3 + x_1 x_2 x_3^2 + x_1^2 x_2 x_3^2) + x_1 x_2 x_3 + x_1^3 x_2 x_3 + x_1^2 x_2^2 x_3 + x_1^2 x_2 x_3^2 \right)$$

$$= \frac{1}{6}(x_1^2 x_2^2 x_3 + x_1^2 x_2 x_3 + 3 x_1 x_2 x_3^2 + x_1^3 x_2 x_3).$$

Using $H_{T^s}(\bar{x})$ and $I$ instead of $H_T(\bar{x})$ we gain time and memory, because $H_{T^s}(\bar{x})$ contains much less terms than $H_T(\bar{x})$. Since the values of the test statistic is the same for all permutations of $(u_1, \ldots, u_t)$, $H_{T^s}(\bar{x})$ suffices for computing the null distribution of the test statistic.
Table 1: Computing times in seconds on a SUNSparc 5 for the Friedman test

<table>
<thead>
<tr>
<th>$b$</th>
<th>$t = 3$</th>
<th>$t = 4$</th>
<th>$t = 5$</th>
</tr>
</thead>
<tbody>
<tr>
<td>2</td>
<td>0.12</td>
<td>0.68</td>
<td>11.00</td>
</tr>
<tr>
<td>4</td>
<td>0.47</td>
<td>6.53</td>
<td>168.22</td>
</tr>
<tr>
<td>7</td>
<td>1.53</td>
<td>33.22</td>
<td></td>
</tr>
<tr>
<td>10</td>
<td>3.43</td>
<td>99.30</td>
<td></td>
</tr>
<tr>
<td>20</td>
<td>19.17</td>
<td></td>
<td></td>
</tr>
<tr>
<td>30</td>
<td>56.57</td>
<td></td>
<td></td>
</tr>
<tr>
<td>40</td>
<td>124.60</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

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:

$$Q = \frac{12}{bt(t+1)} \sum_{j=1}^{r} \left( T_j - \frac{b(t+1)}{2} \right)^2$$

$$= \frac{12}{bt(t+1)} \sum_{j=1}^{r} \left( T_j^2 + \frac{b^2(t+1)^2}{4} - bt(t+1)T_j \right)$$

$$= c_1 S + 3bt(t+1) - \frac{12}{t} \sum_{j=1}^{r} T_j$$

$$= c_1 S + 3bt(t+1) - \frac{12 b t(t+1)}{2}$$

$$= c_1 S + c_2,$$

with $c_1 = \frac{12}{bt(t+1)}$ and $c_2 = -3bt(t+1)$. 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:

$$H_{T_F}(\bar{x}) = \prod_{i=1}^{b} H_{R_{i}}(\bar{x}) = (H_{R_i}(\bar{x}))^b = \left( \frac{1}{t!} \sum_{p=1}^{t} x^{\sigma_p(1)}_1 \ldots x^{\sigma_p(t)}_t \right)^b.$$

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 $T_j = \sum_{i=1}^{b} R_{ij}$, $B_i = \sum_{j=1}^{r} R_{ij}$ 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
We know who many equals either 0 or 1. Let us first define Cochran’s test statistic $C$.

\[
C = t(t - 1) \frac{\sum_{j=1}^{t} (T_j - \bar{E})^2}{\sum_{i=1}^{b} B_i(t - B_i)}
\]

\[
= t(t - 1) \frac{\sum_{j=1}^{t} (T_j - \bar{E})^2}{\sum_{j=1}^{t} j(t - j)b_j}
\]

\[
= t(t - 1) \frac{\sum_{j=1}^{t} (T_j - \bar{E})^2}{t \sum_{j=1}^{t} jb_j - \sum_{j=1}^{t} j^2b_j}
\]

\[
= (t - 1) \frac{\sum_{j=1}^{t} T_j^2 - \bar{E}^2}{tE - \sum_{j=1}^{t} j^2b_j}
\]

\[
= c_1S + c_2,
\]

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

So $C$ is linear in $S$ and we therefore use Theorem 5.4 with scores 0 and 1. We observe that $H_R(\bar{x})$ 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

$$
\xi = (\xi_{z_1}, \ldots, \xi_{z_t}) \equiv \{Z_1 < \ldots < Z_t \leq T\} \cap \{Z_{i+1} > T\}, \ldots, \{Z_{i+m} > T\},
$$

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 $\xi$, we know how many $X$’s and $Y$’s are smaller or equal to $T$ and we know how they are configured.

<table>
<thead>
<tr>
<th>$b^*$</th>
<th>$t = 3$</th>
<th>$t = 4$</th>
<th>$t = 5$</th>
<th>$t = 6$</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>0.10</td>
<td>0.65</td>
<td>3.62</td>
<td>28.00</td>
</tr>
<tr>
<td>2</td>
<td>0.25</td>
<td>1.85</td>
<td>19.87</td>
<td>246.37</td>
</tr>
<tr>
<td>3</td>
<td>0.38</td>
<td>3.90</td>
<td>55.75</td>
<td></td>
</tr>
<tr>
<td>4</td>
<td>0.62</td>
<td>8.08</td>
<td>133.23</td>
<td></td>
</tr>
<tr>
<td>10</td>
<td>4.48</td>
<td>130.88</td>
<td></td>
<td></td>
</tr>
<tr>
<td>20</td>
<td>27.20</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>30</td>
<td>84.52</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

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

Let $\bar{r}_m(\xi)$ and $\bar{r}_n(\xi)$ be the number of 0’s and 1’s in $\xi$, respectively, then we introduce two classes
\[
A_{a,b} = \{ \xi : \bar{r}_m(\xi) = a, \bar{r}_n(\xi) = b \}
\] and
\[
A = \bigcup_{a=1}^{m} \bigcup_{b=1}^{n} A_{a,b}
\]
Now we formally define the statistic $U_c$ and the Mann-Whitney statistic $M_{a,b}$:
\[
U_c : A \rightarrow N, \quad U_c(\xi) = \sum_{j < k} I[\xi_j = 1, \xi_k = 0] + (m - \bar{r}_m(\xi))\bar{r}_n(\xi)
\]
\[
M_{a,b} : A_{a,b} \rightarrow N, \quad M_{a,b}(\xi) = \sum_{j < k} I[\xi_j = 1, \xi_k = 0],
\]
where $I_E$ is the indicator function of the event $E$. We define $\mu : A \rightarrow [0,1]$ as the normalized counting measure. Then,
\[
\Pr(U_c = u | \bar{r}_m = a, \bar{r}_n = b) = \frac{\Pr(U_c = u, \bar{r}_m = a, \bar{r}_n = b)}{\Pr(\bar{r}_m = a, \bar{r}_n = b)}
\]
and
\[
\Pr(U_c = u, \bar{r}_m = a, \bar{r}_n = b) = \sum_{\xi \in A : U_c(\xi) = u} \mu(\xi, \bar{r}_m = a, \bar{r}_n = b)
\]
\[
= \sum_{\xi \in A_{a,b} : M_{a,1}(\xi) = u - (m - a) \cdot b} \mu(\xi, \bar{r}_m = a, \bar{r}_n = b),
\]
because $\mu(\xi, \bar{r}_m = a, \bar{r}_n = b) = 0$ if $\xi \notin A_{a,b}$. 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 $\xi \in A_{a,b}$. 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 $\xi \in A_{a,b}$,
\[
\mu(\xi, \bar{r}_m = a, \bar{r}_n = b) = \binom{m}{a}\binom{n}{b} \Pr(w_1 < \ldots < w_B \leq T, w_{B+1} > T, \ldots, w_N > T)
\]
\[
= \binom{m}{a}\binom{n}{b} \int_{-\infty}^{T} \int_{-\infty}^{w_2} \ldots \int_{-\infty}^{w_z} dF(w_1) \ldots dF(w_B) \left(1 - F(T)\right)^{N-B}
\]
\[
= \binom{m}{a}\binom{n}{b} \frac{1}{B!} \left(1 - F(T)\right)^{N-B}.
\]
Substitution of (34) into (33) results in:
\[
\Pr(U_c = u, \bar{r}_m = a, \bar{r}_n = b) = \sum_{\xi \in A_{a,b} : M_{a,1}(\xi) = u - (m - a) \cdot b} \frac{\binom{m}{a}\binom{n}{b}}{B!} \left(1 - F(T)\right)^{N-B}.
\]
We return to equality (32) and note that
\[
\Pr(\bar{r}_m = a, \bar{r}_n = b) = \Pr(\bar{r}_m = a) \Pr(\bar{r}_n = b) = \binom{m}{a}\binom{n}{b} \left(1 - F(T)\right)^{B} \left(1 - F(T)\right)^{N-B}.
\]
Substitution of (35) and (36) into (32) completes the proof:
\[
\Pr(U_c = u | \bar{r}_m = a, \bar{r}_n = b) = \sum_{\xi \in A_{a,b} : M_{a,1}(\xi) = u - (m - a) \cdot b} \frac{1}{B!} = \Pr(M_{a,b} = u - (m - a)b).
\]
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 $\alpha$ 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


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

<table>
<thead>
<tr>
<th>$n$</th>
<th>0.1</th>
<th>0.05</th>
<th>0.025</th>
<th>0.01</th>
<th>0.005</th>
</tr>
</thead>
<tbody>
<tr>
<td>4</td>
<td>33</td>
<td>37</td>
<td>37</td>
<td>*</td>
<td>*</td>
</tr>
<tr>
<td>5</td>
<td>59.25</td>
<td>65.25</td>
<td>67.25</td>
<td>71.25</td>
<td>*</td>
</tr>
<tr>
<td>6</td>
<td>97.50</td>
<td>103.50</td>
<td>109.50</td>
<td>115.50</td>
<td>125.50</td>
</tr>
<tr>
<td>7</td>
<td>151.75</td>
<td>161.75</td>
<td>169.75</td>
<td>179.75</td>
<td>185.75</td>
</tr>
<tr>
<td>8</td>
<td>222</td>
<td>236</td>
<td>248</td>
<td>262</td>
<td>268</td>
</tr>
<tr>
<td>9</td>
<td>312.25</td>
<td>330.25</td>
<td>346.25</td>
<td>364.25</td>
<td>374.25</td>
</tr>
<tr>
<td>10</td>
<td>422.50</td>
<td>446.50</td>
<td>466.50</td>
<td>488.50</td>
<td>502.50</td>
</tr>
<tr>
<td>11</td>
<td>556.75</td>
<td>586.75</td>
<td>612.75</td>
<td>640.75</td>
<td>660.75</td>
</tr>
<tr>
<td>12</td>
<td>715</td>
<td>753</td>
<td>785</td>
<td>821</td>
<td>845</td>
</tr>
<tr>
<td>13</td>
<td>901.25</td>
<td>949.25</td>
<td>987.25</td>
<td>1031.25</td>
<td>1061.25</td>
</tr>
<tr>
<td>14</td>
<td>1119.50</td>
<td>1173.50</td>
<td>1221.50</td>
<td>1275.50</td>
<td>1309.50</td>
</tr>
<tr>
<td>15</td>
<td>1365.75</td>
<td>1433.75</td>
<td>1489.75</td>
<td>1551.75</td>
<td>1593.75</td>
</tr>
<tr>
<td>16</td>
<td>1648</td>
<td>1726</td>
<td>1792</td>
<td>1868</td>
<td>1918</td>
</tr>
<tr>
<td>17</td>
<td>1966.25</td>
<td>2058.25</td>
<td>2134.25</td>
<td>2222.25</td>
<td>2280.25</td>
</tr>
</tbody>
</table>

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

<table>
<thead>
<tr>
<th>$n$</th>
<th>0.1</th>
<th>0.05</th>
<th>0.025</th>
<th>0.01</th>
<th>0.005</th>
</tr>
</thead>
<tbody>
<tr>
<td>4</td>
<td>11</td>
<td>12</td>
<td>12</td>
<td>*</td>
<td>*</td>
</tr>
<tr>
<td>5</td>
<td>16.5</td>
<td>17.5</td>
<td>17.5</td>
<td>18.5</td>
<td>*</td>
</tr>
<tr>
<td>6</td>
<td>23</td>
<td>24</td>
<td>25</td>
<td>26</td>
<td>27</td>
</tr>
<tr>
<td>7</td>
<td>30.5</td>
<td>31.5</td>
<td>33.5</td>
<td>34.5</td>
<td>35.5</td>
</tr>
<tr>
<td>8</td>
<td>39</td>
<td>41</td>
<td>42</td>
<td>44</td>
<td>45</td>
</tr>
<tr>
<td>9</td>
<td>48.5</td>
<td>50.5</td>
<td>52.5</td>
<td>54.5</td>
<td>55.5</td>
</tr>
<tr>
<td>10</td>
<td>60</td>
<td>62</td>
<td>64</td>
<td>66</td>
<td>67</td>
</tr>
<tr>
<td>11</td>
<td>71.5</td>
<td>74.5</td>
<td>76.5</td>
<td>78.5</td>
<td>80.5</td>
</tr>
<tr>
<td>12</td>
<td>84</td>
<td>87</td>
<td>90</td>
<td>93</td>
<td>95</td>
</tr>
<tr>
<td>13</td>
<td>98.5</td>
<td>101.5</td>
<td>104.5</td>
<td>107.5</td>
<td>109.5</td>
</tr>
<tr>
<td>14</td>
<td>113</td>
<td>117</td>
<td>120</td>
<td>124</td>
<td>126</td>
</tr>
<tr>
<td>15</td>
<td>129.5</td>
<td>133.5</td>
<td>136.5</td>
<td>141.5</td>
<td>143.5</td>
</tr>
<tr>
<td>16</td>
<td>146</td>
<td>151</td>
<td>155</td>
<td>159</td>
<td>163</td>
</tr>
<tr>
<td>17</td>
<td>164.5</td>
<td>169.5</td>
<td>173.5</td>
<td>178.5</td>
<td>182.5</td>
</tr>
<tr>
<td>18</td>
<td>183</td>
<td>189</td>
<td>194</td>
<td>199</td>
<td>203</td>
</tr>
<tr>
<td>19</td>
<td>203.5</td>
<td>209.5</td>
<td>214.5</td>
<td>220.5</td>
<td>224.5</td>
</tr>
<tr>
<td>20</td>
<td>225</td>
<td>231</td>
<td>237</td>
<td>244</td>
<td>248</td>
</tr>
<tr>
<td>21</td>
<td>247.5</td>
<td>254.5</td>
<td>260.5</td>
<td>267.5</td>
<td>271.5</td>
</tr>
<tr>
<td>22</td>
<td>270</td>
<td>278</td>
<td>285</td>
<td>292</td>
<td>297</td>
</tr>
</tbody>
</table>

The entries in the tables are critical values for the respective tests at different significance levels.
<table>
<thead>
<tr>
<th></th>
<th>0.05</th>
<th>0.02</th>
<th>0.01</th>
</tr>
</thead>
<tbody>
<tr>
<td>5</td>
<td>2.612</td>
<td>3.082</td>
<td>3.311</td>
</tr>
<tr>
<td>6</td>
<td>2.877</td>
<td>3.295</td>
<td>3.685</td>
</tr>
<tr>
<td>7</td>
<td>3.116</td>
<td>3.631</td>
<td>3.968</td>
</tr>
<tr>
<td>8</td>
<td>3.389</td>
<td>3.925</td>
<td>4.269</td>
</tr>
<tr>
<td>9</td>
<td>3.638</td>
<td>4.243</td>
<td>4.624</td>
</tr>
<tr>
<td>10</td>
<td>3.872</td>
<td>4.517</td>
<td>4.934</td>
</tr>
<tr>
<td>11</td>
<td>4.094</td>
<td>4.784</td>
<td>5.228</td>
</tr>
<tr>
<td>12</td>
<td>4.305</td>
<td>5.037</td>
<td>5.514</td>
</tr>
<tr>
<td>13</td>
<td>4.509</td>
<td>5.281</td>
<td>5.787</td>
</tr>
<tr>
<td>14</td>
<td>4.705</td>
<td>5.516</td>
<td>6.048</td>
</tr>
<tr>
<td>15</td>
<td>4.894</td>
<td>5.742</td>
<td>6.300</td>
</tr>
<tr>
<td>16</td>
<td>5.077</td>
<td>5.960</td>
<td>6.543</td>
</tr>
</tbody>
</table>

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

<table>
<thead>
<tr>
<th></th>
<th>0.05</th>
<th>0.02</th>
<th>0.01</th>
<th>0.001</th>
</tr>
</thead>
<tbody>
<tr>
<td>5</td>
<td>4.149</td>
<td>4.149</td>
<td>*</td>
<td></td>
</tr>
<tr>
<td>6</td>
<td>5.122</td>
<td>5.229</td>
<td>5.582</td>
<td></td>
</tr>
<tr>
<td>7</td>
<td>6.155</td>
<td>6.445</td>
<td>6.777</td>
<td></td>
</tr>
<tr>
<td>8</td>
<td>7.285</td>
<td>7.612</td>
<td>8.078</td>
<td></td>
</tr>
<tr>
<td>9</td>
<td>8.448</td>
<td>8.903</td>
<td>9.305</td>
<td></td>
</tr>
<tr>
<td>11</td>
<td>10.797</td>
<td>11.338</td>
<td>11.881</td>
<td></td>
</tr>
<tr>
<td>12</td>
<td>11.965</td>
<td>12.555</td>
<td>13.162</td>
<td></td>
</tr>
<tr>
<td>13</td>
<td>13.127</td>
<td>13.757</td>
<td>14.421</td>
<td></td>
</tr>
<tr>
<td>15</td>
<td>15.450</td>
<td>16.167</td>
<td>16.938</td>
<td></td>
</tr>
<tr>
<td>16</td>
<td>16.607</td>
<td>17.365</td>
<td>18.186</td>
<td></td>
</tr>
<tr>
<td>17</td>
<td>17.761</td>
<td>18.559</td>
<td>19.428</td>
<td></td>
</tr>
<tr>
<td>18</td>
<td>18.912</td>
<td>19.750</td>
<td>20.665</td>
<td></td>
</tr>
<tr>
<td>19</td>
<td>20.061</td>
<td>20.937</td>
<td>21.897</td>
<td></td>
</tr>
<tr>
<td>20</td>
<td>21.208</td>
<td>22.121</td>
<td>23.125</td>
<td></td>
</tr>
<tr>
<td>21</td>
<td>22.352</td>
<td>23.301</td>
<td>24.348</td>
<td></td>
</tr>
</tbody>
</table>

Table 6: Right critical values for the Klotz test, \( n = \frac{1}{b}N \)

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

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.