Fast computation of the exact null distribution of Spearman's ρ and Page's L statistic for samples with and without ties

M.A. van de Wiel and A. Di Bucchianico
Department of Mathematics and Computing Science
Eindhoven University of Technology
P. O. Box 513
5600 MB Eindhoven
The Netherlands
markvdw@win.tue.nl   sandro@win.tue.nl

Abstract

We present a new algorithm for computing the exact null distribution of the Spearman rank correlation statistic ρ, which also works in the case of ties. The algorithm is based on symmetries in the representation of the probability generating function as a permanent with monomial entries. We present new critical values for sample sizes $19 \leq n \leq 22$. Finally, we show how to derive the exact null distribution of Page's L statistic from the null distribution of ρ.

Keywords   Spearman rank correlation test statistic, Page's L test statistic, computer algebra, probability generating function, ties, permanent, symmetric group, permutation.

AMS classification   62E15, 62E30, 62G10, 62J10, 62Q05, 65U05

Running title   Exact distribution of Spearman's ρ and Page's L.

1   Introduction

Spearman's rank correlation test statistic ρ is a widely used nonparametric test statistic for testing correlation. To perform the test we need the distribution of ρ under the null hypothesis of no rank correlation. Computing the exact null distribution of ρ is a very hard problem, since ρ induces a fine partition on the $n!$ permutations of $\{1, \ldots, n\}$. On the other hand, this indicates that ρ attains many confidence levels. With respect to this aspect, ρ performs better than its competitor Kendall's τ.

Existing methods for computing the null distribution of ρ are based on counting permutations (in a clever way), which is very time-consuming. The exact null distribution of ρ when ties are not present can be found in Owen (1962) for $2 \leq n \leq 11$, in de Jonge and van Montfort (1972) for $n = 12$, in Otten (1973) for $13 \leq n \leq 16$ (critical values) and in Franklin (1988a) for $13 \leq n \leq 18$ (complete distribution). Closed formulas for the first eight moments are given in David et al. (1951); with the help of a computer algebra system, it is easy to compute higher moments with their method. Approximations to the null distribution of ρ when

ties are not present were studied extensively. David et al. (1951) derive exact expressions for cumulants of ρ which they use for approximations based on Edgeworth series. Franklin (1988b) compares seven approximations (not including Edgeworth series) and concludes that the Pearson Type II approximation is the most accurate approximation. It is shown in Best and Roberts (1975) that approximations based on Edgeworth series are more accurate than Pearson Type II approximations. For further discussion of approximations and critical values based on Edgeworth expansions, we refer to Ramsay (1989).

In this article we compute the null distribution of ρ using the representation of its probability generating function as a permanent (a 'signless determinant') with monomial entries, following an idea that goes back to Olds (1938) and Kendall et al. (1939). This method also applies in case of ties. For sample sizes up to $n = 10$, direct evaluation of the permanent in the computer algebra system Maple is feasible. An alternative way of computing permanents is using Ryser's algorithm based on inclusion-exclusion. Unfortunately, both methods are too time-consuming for larger sample sizes. However, the representation of the probability generating function as a permanent reveals many symmetries which we use to achieve a considerable reduction in the number of necessary computations. In this way, we were able to compute the exact null distribution of ρ for $n \leq 22$ for the case without ties. Our methods also allow fast computation of tail probabilities. Computing exact tail probabilities is especially important when ties are present, because the dependence of the null distribution on the ties structure inhibits a compact presentation of tables with critical values.

In Section 7 we show another advantage of our approach. We show that the probability generating function of Page's L statistic for ordered alternatives in a randomised block design equals the probability generating function of Spearman's ρ raised to the number of blocks.

This paper is organised as follows. In Section 2 we relate Spearman's ρ to statistics on the symmetric group. The representation of the probability generating function of ρ as a permanent is introduced in Section 3. Various properties of this permanent are derived in Section 4. The implementation of our results of Section 4 is discussed in Section 5. The computation in the case of ties is presented in Section 6. In Section 7, we show our results on Page's L statistic. Finally, Section 8 contains a table with new critical values of Spearman's ρ for $19 \leq n \leq 22$.

2 The Spearman test

The Spearman test is a two-sample rank correlation test. Denote the samples by $X_1, \ldots, X_n$ and $Y_1, \ldots, Y_n$, respectively. Rank both samples separately from small to large. The Spearman rank correlation statistic ρ is then defined as the sum of the squared differences between rank scores of the pairs $(X_1, Y_1), \ldots, (X_n, Y_n)$. Under the null hypothesis of no rank correlation, all joint rank configurations of the two samples are equiprobable. Hence, computing the null distribution of ρ boils down to enumerating the values of the following statistic (in the combinatorial sense) on $\mathcal{S}_n \times \mathcal{S}_n$, where $\mathcal{S}_n$ denotes the symmetric group of n elements, i.e. the group of permutations on $\{1, 2, \ldots, n\}$:

$$S_2 : (\sigma, \tau) \longmapsto \sum_{j=1}^{n} (\sigma(j) - \tau(j))^2. \tag{1}$$

Define the following equivalence relation on $\mathcal{S}_n \times \mathcal{S}_n : (\sigma, \tau) \sim (\rho, \varsigma)$ if there exists a $\nu \in \mathcal{S}_n$ such that $\rho = \sigma \circ \nu$ and $\zeta = \tau \circ \nu$. Note that for all $\nu \in \mathcal{S}_n$, we have $S_2(\sigma \circ \nu, \tau \circ \nu) = S_2(\sigma, \tau)$

since this is nothing but rearranging the sum in (1). Hence, S_2 is constant on the equivalence classes of $\sim$. It is clear that the elements (σ, Id) with $\sigma \in \mathcal{S}_n$ form a set of representatives for the equivalence classes of $\sim$ and that each equivalence class has exactly $n!$ elements. Now define a statistic S_1 on $\mathcal{S}_n$ as follows:

$$S_1 : \sigma \longmapsto \sum_{j=1}^{n} (\sigma(j) - j)^2. \tag{2}$$

It follows from the discussion above that the distributions of the statistics S_1 and S_2 coincide. Hence, we may restrict ourselves to studying S_1. The statistic S_1 can be written as follows:

$$S_1 = 2 \left(\sum_{j=1}^{n} j^2 - \sum_{j=1}^{n} j\, \sigma(j) \right) = \frac{1}{3} n(n+1)(2n+1) - 2 \sum_{j=1}^{n} j\, \sigma(j). \tag{3}$$

Hence, S_1 is statistically equivalent to $\widetilde{S_1} := \sum_{j=1}^{n} j\, \sigma(j)$. The distribution of $\widetilde{S_1}$ also appears in the context of linear rank statistics with equidistant regression scores. Let R_i be the rank of X_i in the pooled sample if $i \le n$ and the rank of Y_{i-n} if $i > n$. A linear rank statistic with regression scores is a statistic of the form $T = \sum_{i=1}^{2n} a(i)\, R_i$. If the regression scores $a(i)$ are equidistant, i.e. $a(i) = ci + d$ for some constants c and d, then up to an affine transformation, the distribution of T equals the distribution of $\widetilde{S_1}$. Hence, computing the distribution of S_1 also yields the distribution of T when the regression scores are equidistant.

3 The probability generating function

Let A be a $n * n$ square matrix with entries a_{ij}. The permanent of A is defined as

$$\mathrm{per}(A) = \sum_{\sigma \in \mathcal{S}_n} \prod_{j=1}^{n} a_{\sigma(j), j}. \tag{4}$$

The definition of the permanent resembles that of the determinant of a matrix. However, unlike for determinants, the signs of the permutations are ignored. Endow the symmetric group $\mathcal{S}_n$ with the normalized counting measure, i.e. each permutation on $\{1, \ldots, n\}$ has probability $1/n!$. Then the null distribution of ρ and the distribution of S_1 as a random variable on $\mathcal{S}_n$ coincide. Following Olds (1938) and Kendall et al. (1939), we express the probability generating function of S_1 and $\widetilde{S_1}$ in terms of permanents:

$$\sum_{k=0}^{\infty} \Pr(S_1 = k)\, x^k = \frac{1}{n!}\, \mathrm{per}(P) \quad \text{and} \quad \sum_{k=0}^{\infty} \Pr(\widetilde{S_1} = k)\, x^k = \frac{1}{n!}\, \mathrm{per}(\widetilde{P}), \tag{5}$$

where

$$P_{ij} = x^{(i-j)^2} \quad \text{and} \quad \widetilde{P}_{ij} = x^{ij}, \quad i, j = 1, \ldots, n. \tag{6}$$

Unfortunately, the permanent does not share many of the nice properties of the determinant. E.g., there does not exist an analogue of Gaussian elimination. However, the permanent is invariant under transpositions of rows and columns. For other properties of permanents we refer to Minc (1978). Several ways of computing permanents are compared in Chapter 7

of Minc (1978). The best method seems to be an algorithm based on inclusion-exclusion
due to Ryser (see Ryser (1963)). His algorithm, together with some improvements, is also
described in Nijenhuis and Wilf (1975). The Ryser algorithm was designed for matrices
with scalar entries, but it also works, although less efficient, for our matrices P and $\widetilde{P}$ with
monomial entries. The computer algebra system Maple has a built-in function `permanent`,
which computes permanents using minor expansions. We experimented both with the built-in
function `permanent` in Maple and a self-written implementation of the Ryser algorithm.

Our experiments show that both the built-in Maple function and our implementation of the
Ryser algorithm only work for moderate sample sizes. Above $n = 10$, both computation time
and memory use increase dramatically. E.g., computation time for $n = 14$ was 10 times larger
than for $n = 13$ and memory usage increased from 2 MB to 12 MB. Therefore, we need other
methods for computing the permanent of P or $\widetilde{P}$. In the next sections we will derive some
theorems on the structure of P. With some small modifications, these theorems also hold for
$\widetilde{P}$.

4   Symmetries and expansions

In this section we explain the main ideas that we used to compute the permanent given in
(5). We first need some notation.

Definition 4.1 *If n is a positive integer, then we define the following:*

- $n_c = \lceil n/2 \rceil, n_f = \lfloor n/2 \rfloor$.

- $U = \{1, \ldots, n_c\}, L = \{n_c + 1, \ldots, n\}$.

- $\mathcal{S}$ is the set of all ordered n_c-tuples with elements from $\{1, \ldots, n\}$.

- *if $S = (s_1, \ldots, s_{n_c})$ is an element of $\mathcal{S}$, then $\Lambda(S)$ is the ordered n_f-tuple, the elements
  of which form the set $\{1, \ldots, n\} \backslash \{s_1, \ldots, s_{n_c}\}$.*

- *the map τ acting on $\{1, \ldots, n\}$ is defined by $\tau(x) = n + 1 - x$.*

- *if $S = (s_1, \ldots, s_{n_c})$ is an element of $\mathcal{S}$, then $T(S)$ is the ordered n_c-tuple, the elements
  of which form the set $\{\tau(s_1), \ldots, \tau(s_{n_c})\}$.*

- *for a matrix $P = (P_{i,j})_{i,j}$ and ordered tuples $A = (A_1, \ldots, A_k)$ and $B = (B_1, \ldots, B_k)$,
  $P(A|B)$ denotes the submatrix of P with entries P_{A_i, B_j}. In particular, we will write
  $P_{i,j}(A|B)$ for P_{A_i, B_j}. Note that $P(U|S)$ and $P(L|S)$ denote the upper and lower half,
  respectively, of the submatrix of P consisting of the columns with indices from S.*

The first idea is to use the analogue for permanents of the Laplace expansion of determinants. E.g., if $n = 4$, then we may expand the permanent as follows:

$$\text{per} \begin{pmatrix} 1 & x & x^4 & x^9 \\ x & 1 & x & x^4 \\ x^4 & x & 1 & x \\ x^9 & x^4 & x & 1 \end{pmatrix} = \text{per} \begin{pmatrix} 1 & x \\ x & 1 \end{pmatrix} \text{per} \begin{pmatrix} 1 & x \\ x & 1 \end{pmatrix} + \text{per} \begin{pmatrix} 1 & x^4 \\ x & x \end{pmatrix} \text{per} \begin{pmatrix} x & x \\ x^4 & 1 \end{pmatrix}$$

$$+ \text{per} \begin{pmatrix} 1 & x^9 \\ x & x^4 \end{pmatrix} \text{per} \begin{pmatrix} x & 1 \\ x^4 & x \end{pmatrix} + \text{per} \begin{pmatrix} x & x^4 \\ 1 & x \end{pmatrix} \text{per} \begin{pmatrix} x^4 & x \\ x^9 & 1 \end{pmatrix}$$

$$+ \text{per} \begin{pmatrix} x & x^9 \\ 1 & x^4 \end{pmatrix} \text{per} \begin{pmatrix} x^4 & 1 \\ x^9 & x \end{pmatrix} + \text{per} \begin{pmatrix} x^4 & x^9 \\ x & x^4 \end{pmatrix} \text{per} \begin{pmatrix} x^4 & x \\ x^9 & x^4 \end{pmatrix} .$$

Theorem 4.2 (Laplace Expansion)

$$per(P) = \sum_{S \in \mathcal{S}} per\Big(P(U|S) \Big) \, per\Big(P(L|\Lambda(S)) \Big), \tag{7}$$

with P as in (6), per(A) as in (4), and U, L, $\mathcal{S}$, S and $P(A|B)$ as in Definition 4.1.

In Franklin (1988a), essentially the same idea is used for decomposing permutations. However, the idea is more transparent in our generating function setting.

The Laplace expansion reduces the number of operations (additions and multiplications) needed to compute the permanent, since it efficiently combines permutations σ_1 and σ_2 if, given a fixed tuple S,

$$\sum_{j \in S} (\sigma_1(j) - j)^2 = \sum_{j \in S} (\sigma_2(j) - j)^2 \ \text{ or } \ \sum_{j \in \Lambda(S)} (\sigma_1(j) - j)^2 = \sum_{j \in \Lambda(S)} (\sigma_2(j) - j)^2 .$$

E.g., if $n = 6$ and $S = (1, 2, 3)$, then $P(U|S)$ is the left-upper 3x3 submatrix of matrix (8). Let $\widetilde{\sigma}$ be an arbitrary permutation of $(4, 5, 6)$ and, using the cycle notation for a permutation of $(1, 2, 3)$, consider $\sigma_1 = (123)\widetilde{\sigma}$ and $\sigma_2 = (132)\widetilde{\sigma}$. Then,

$$\sum_{j \in S} (\sigma_1(j) - j)^2 = \sum_{j \in S} (\sigma_2(j) - j)^2 = 6 .$$

Therefore, these permutations are combined and together they contribute $2x^6$ to $\text{per}\big(P(U|S) \big)$. It is not easy to calculate how much faster this expansion is, but the larger n is, the more efficient this expansion is compared to direct evaluation of the permanent.

The symmetries of the distributions of S_1 and $\widetilde{S_1}$ carry over to symmetries of the matrices defined in (6). The following example shows that because of these symmetries, it is not necessary to evaluate all permanents of submatrices in the Laplace expansion.

Example, $n = 6$:

5

$$P = \begin{pmatrix} 1 & x & x^4 & x^9 & x^{16} & x^{25} \\ x & 1 & x & x^4 & x^9 & x^{16} \\ x^4 & x & 1 & x & x^4 & x^9 \\ x^9 & x^4 & x & 1 & x & x^4 \\ x^{16} & x^9 & x^4 & x & 1 & x \\ x^{25} & x^{16} & x^9 & x^4 & x & 1 \end{pmatrix} \qquad (8)$$

Suppose $S = (1, 4, 6)$. Then $\Lambda(S) = (2, 3, 5)$, $T(S) = (1, 3, 6)$, $T(\Lambda(S)) = (2, 4, 5)$. Therefore,

$$P(U|S) = \begin{pmatrix} 1 & x^9 & x^{25} \\ x & x^4 & x^{16} \\ x^4 & x & x^9 \end{pmatrix} \qquad P(L|T(S)) = \begin{pmatrix} x^9 & x & x^4 \\ x^{16} & x^4 & x \\ x^{25} & x^9 & 1 \end{pmatrix}$$

$$P(U|T(\Lambda(S))) = \begin{pmatrix} x & x^9 & x^{16} \\ 1 & x^4 & x^9 \\ x & x & x^4 \end{pmatrix} \qquad P(L|\Lambda(S)) = \begin{pmatrix} x^4 & x & x \\ x^9 & x^4 & 1 \\ x^{16} & x^9 & x \end{pmatrix} \; .$$

This illustrates that $P(L|T(S))$ can be obtained from $P(U|S)$ by swapping the first and the third row and the first and the third column. Therefore, their permanents are equal. This also holds for the two other matrices.

Theorem 4.3 *If n is even and S is an arbitrary ordered $n/2-$tuple with elements from $\{1, \ldots, n\}$, then*

$$per\Big(P(U|S)\Big) = per\Big(P(L|T(S))\Big), \qquad (9)$$

with P as in (6), $per(A)$ as in (4) and $U, L, T, P(A|B)$ as in Definition 4.1.

Proof: First note that $(T(S))_k = n + 1 - S_{|S|+1-k} = n + 1 - S_{n/2+1-k}$, because n is even. From (6) we know that $P_{i,j} = P_{n+1-i,n+1-j}$. Therefore, with $k = n/2 + 1 - j$,

$$P_{i,j}(U|S) = P_{U_i,S_j} = P_{i,S_j} = P_{n+1-i,n+1-S_j}$$
$$= P_{L_{n/2+1-i},(T(S))_{n/2+1-j}} = P_{n/2+1-i,n/2+1-j}(L, T(S)).$$

So we may obtain the matrix $P(L|T(S))$ from the matrix $P(U|S)$ by swapping the ith and $(n/2+1-i)$th row and the jth and $(n/2+1-j)$th column for $i, j = 1, \ldots, n/2$. Since swapping of rows and swapping of columns leaves the permanent invariant, the permanents of $P(L|T(S))$ and $P(U|S)$ are equal. We replace S by $T(\Lambda(S))$ and we note that $\Lambda(\Lambda(S)) = S$, where Λ is as in Definition 4.1. The result is that the permanents of $P(L|\Lambda(S))$ and $P(U|T(\Lambda(S)))$ are also equal and our proof is complete. $\square$

It trivially follows from (2), that the null distribution of ρ is symmetric around its mean. In order to exploit this symmetry, we now define a linear operator Ψ_M that operates on polynomials.

Definition 4.4 *The operator* Ψ_M *is defined by*

$$\Psi_M(x^b) := x^{M-b}, \tag{10}$$

where M equals twice the mean of ρ, i.e., $M = n(n^2 - 1)/3$. The definition of Ψ_M is extended by linearity to arbitrary polynomials.

Example, $n = 3$:

$$P = \begin{pmatrix} 1 & x & x^4 \\ x & 1 & x \\ x^4 & x & 1 \end{pmatrix}$$

Let $S = (1,2)$, then $\Lambda(S) = (3)$, $T(S) = (2,3)$, $T(\Lambda(S)) = (1)$. So,

$$P(U|S) = \begin{pmatrix} 1 & x \\ x & 1 \end{pmatrix} \quad P(L|\Lambda(S)) = (1) \quad P(L|T(S)) = \begin{pmatrix} x & x^4 \\ 1 & x \end{pmatrix} \quad P(U|T(\Lambda(S))) = (x^4) \ .$$

We have $n = 3$ and therefore $M = 8$. Hence,

$$\Psi_8\left(\mathrm{per}\Big(P(U|S)\Big)\,\mathrm{per}\Big(P(L|\Lambda(S))\Big)\right) = \Psi_8(1 + x^2) = x^6 + x^8$$

$$= \mathrm{per}\Big(P(U|T(S))\Big)\,\mathrm{per}\Big(P(L|T(\Lambda(S)))\Big).$$

Theorem 4.5 *Let Ψ_M be as in Definition 4.4, then*

$$\Psi_M\left(per\Big(P(U|S)\Big)\,per\Big(P(L|\Lambda(S))\Big)\right) = per\Big(P(U|T(S))\Big)\,per\Big(P(L|T(\Lambda(S)))\Big), \tag{11}$$

with P as in (6), $per(A)$ as in (4), and U, L, S , Λ, T and $P(A|B)$ as in Definition 4.1.

Proof: It suffices to find a bijection of $\mathcal{S}_{n_c} \times \mathcal{S}_{n_f} : (\sigma_1, \sigma_2) \mapsto (\sigma_1^*, \sigma_2^*)$ such that

$$\prod_{j=1}^{n_c} P_{\sigma_1(j),j}(U|S) \prod_{j=1}^{n_f} P_{\sigma_2(j),j}(L|\Lambda(S)) \prod_{j=1}^{n_c} P_{\sigma_1^*(j),j}(U|T(S)) \prod_{j=1}^{n_f} P_{\sigma_2^*(j),j}(L|T(\Lambda(S))) = x^M, \tag{12}$$

where $M = n(n^2 - 1)/3$. We accomplish this by establishing bijective correspondences in the following commutative diagram:

$$\begin{array}{ccc} (\sigma_1,\sigma_2) & \longrightarrow & \sigma \\ \downarrow & & \downarrow \\ (\sigma_1^*,\sigma_2^*) & \longleftarrow & \sigma^* \end{array},$$

where $\sigma \in \mathcal{S}_n$ is such that $\sigma(S) \subset U$ and $\sigma^* \in \mathcal{S}_n$ is such that $\sigma^*(T(S)) \subset U$. We define

- $\sigma(S_j) = \sigma_1(j)$, $j = 1, \ldots, n_c$

- $\sigma(\Lambda(S)_j) = \sigma_2(j) + n_c$, $j = 1, \ldots, n_f$

7

- $\sigma_1^*(j) = \sigma^*(\tau(S_j))$, $j = 1, \ldots, n_c$

- $\sigma_2^*(j) = \sigma^*(\tau(\Lambda(S)_j)) - n_c$, $j = 1, \ldots, n_f$.

Since the elements of S and $\Lambda(S)$ together form the set $\{1, \ldots, n\}$, the relation (12) is easily seen to be equivalent to

$$\prod_{j=1}^{n} P_{\sigma(j),j} \prod_{j=1}^{n} P_{\sigma^*(j),j} = x^M. \tag{13}$$

Hence, it suffices to find a bijective correspondence between σ and σ^* such that (13) holds. Define

$$\sigma^*(j) = \sigma(n + 1 - j), \tag{14}$$

i.e., $\sigma^* = \sigma \circ \tau$. Obviously, this is a bijection on $\mathcal{S}_n$ such that $\sigma^*(T(S)) = \sigma(T^2(S)) = \sigma(S) \subset U$. Substitution of (14) into (13), yields that (13) is equivalent to

$$\sum_{j=1}^{n} (\sigma(j) - j)^2 + \sum_{j=1}^{n} (\sigma(n + 1 - j) - j)^2 = M. \tag{15}$$

However, this holds for *any* $\sigma \in \mathcal{S}$ by a straightforward computation, using the well-known identities $(a - b)^2 + (a + b)^2 = 2(a^2 + b^2)$ and $\sum_{j=1}^{n} j^2 = n(n + 1)(2n + 1)/6$. $\qquad \square$

5 Computation of the probability generating function

In order to apply Theorem 4.2, we need a rule that, given the tuple S, generates the next tuple, $\widetilde{S}$, such that, starting with $(1, \ldots, n_c)$, all elements of $\mathcal{S}$ are considered once. We use the following rule: let s_ℓ be the largest element in S for which $s_\ell + 1 \leq n$ and $s_\ell + 1 \notin S$. We denote the first $\ell - 1$ elements of S by $s_1, \ldots s_{\ell-1}$ and we denote the number of elements larger or equal to s_ℓ by $c(\ell)$. If s_ℓ exists, then we define the next tuple $\widetilde{S} = (s_1, \ldots, s_{\ell-1}, s_\ell + 1, \ldots, s_\ell + c(\ell))$. If no such s_ℓ exists, then $S = (n - n_c + 1, \ldots, n)$ and we are done. Two examples for $n = 8$:

$$S = \{1, 3, 4, 6\} \rightarrow \widetilde{S} = \{1, 3, 4, 7\} \tag{16}$$
$$S = \{2, 3, 7, 8\} \rightarrow \widetilde{S} = \{2, 4, 5, 6\}. \tag{17}$$

In order to avoid unnecessary evaluations of permanents, we use Theorem 4.3 (if n is even) and Theorem 4.5. Let $S \prec \widetilde{S}$ denote the event that S is considered before $\widetilde{S}$, which is the case if $s_i < \tilde{s}_i$, where i is the smallest element of S such that $s_i \neq \tilde{s}_i$. Using this notation we present the algorithm for computing $\mathcal{P} = \mathrm{per}(P)$ if n is even:

1. $\mathcal{P} = 0$, $S = (1, \ldots, n_c)$

2. compute $T(S)$ and $T(\Lambda(S))$

3. if $T(S) \prec S$ or $T(\Lambda(S)) \prec S \rightarrow \mathcal{P}_1 = 0$, else $\mathcal{Q} = \mathrm{per}\Big(P(U|S)\Big) \mathrm{per}\Big(P(L|\Lambda(S))\Big)$

4. (a) if $T(S) = S$ and $S \prec T(\Lambda(S)) \rightarrow \mathcal{P}_1 = 2\mathcal{Q}$

8

(b) if $S \prec T(S)$ and $T(\Lambda(S)) = S \rightarrow \mathcal{P}_1 = \mathcal{Q} + \Psi_M(\mathcal{Q})$

(c) if $S \prec T(S)$ and $S \prec T(\Lambda(S)) \rightarrow \mathcal{P}_1 = 2(\mathcal{Q} + \Psi_M(\mathcal{Q}))$

5. $\mathcal{P} = \mathcal{P} + \mathcal{P}_1$

6. if $s_1 < n - n_c + 1 \rightarrow$ stop, else $\ S = (s_1, \ldots, s_{\ell-1}, s_\ell + 1, \ldots, s_\ell + c(\ell))$ and go to 2.

Application of Ψ_M is justified by Theorem 4.5 and multiplication by two is justified by Theorem 4.3. Both theorems reduce the computing time by nearly a factor 2. If n is odd, we skip the comparisons of S with $T(\Lambda(S))$, since we can not apply Theorem 4.3 in this case. All we need to store is $\mathcal{P}$, so we do not use much memory.

The probability generating function is an elegant tool for computing the exact distribution of ρ. However, direct computing with polynomials, as we do in the theorems mentioned in the previous section, involves the use of a computer algebra package, e.g. Maple. Professor Andries Brouwer provided us with a much faster implementation in C, where polynomials are represented by arrays. He also showed us that with his implementation, the Laplace expansion of Theorem 4.2 is efficient for small sizes as well. He therefore used the the Laplace expansion more than once for each step. This is not the case for our implementation in Maple.

6   Ties

When ties are present, it is customary to assign average rank scores to the observations. Let $(Z_1, \ldots, Z_n)$ be the X-sample or the Y-sample. We denote by k_Z the number of different values $Z_1' < \ldots < Z_k'$, $k \leq n$. The size of the jth tie is denoted by $\xi_{Z,j}$ for $j = 1, \ldots, k$ and $Z = X$ or Y. The average scores for the ties are defined by:

$$\widetilde{a}_Z(j) = \sum_{i=1}^{j-1} \xi_{Z,i} + \frac{1}{2}\left(\xi_{Z,j} + 1\right), \tag{18}$$

for $j = 1, \ldots, k$ (the sum is empty for $j = 1$). Then the rank scores are

$$a_Z(1) = \ldots = a_Z(\xi_{Z,1}) = \widetilde{a}_Z(1), \ldots, a_Z(n + 1 - \xi_{Z,k}) = \ldots = a_Z(n) = \widetilde{a}_Z(k), Z = X \text{ or } Y.$$

Conditional on the tie structure, Spearman's ρ for tied observations is now defined as:

$$\overline{\rho} = \sum_{\ell=1}^{n} \left(a_Y(\sigma(\ell)) - a_X(\ell)\right)^2 \tag{19}$$

Note that when there are no ties, $\overline{\rho}$ reduces to ρ. Given a certain tie structure, we apply (5) with $P_{ij} = \left(a_Y(i) - a_X(j)\right)^2$, $i, j = 1, \ldots, n$ to find the conditional probability generating function of $\overline{\rho}$. Theorem 4.2 holds for any permanent, so we may also apply it here. However, we are able to make it more suitable for computation when ties occur, because then some

9

columns of P coincide. In order to reformulate (7) we need the following notation:

$$
\begin{aligned}
T_j &= \sum_{i=1}^{j-1} \xi_{Z,i} + 1 \\
\Gamma(j) &= \begin{cases} \emptyset & \text{if } t_j = 0, \\ \{T_j, \ldots, T_j + t_j - 1\} & \text{if } 1 \le t_j \le \tau_{Z,j} \end{cases} \\
T &= \bigcup_{j=1}^{k} \Gamma(j) \\
T \in \mathcal{T} &\Leftrightarrow \sum_{j=1}^{k} t_k = n_c = \lceil n/2 \rceil.
\end{aligned}
$$

Using the above notation and Definition 4.1, we now rewrite (7) as

$$
\text{per}(P) = \sum_{T \in \mathcal{T}} \prod_{j=1}^{k} \binom{\xi_{Z,j}}{t_j} \, \text{per}\Big(P(U|T) \Big) \, \text{per}\Big(P(L|\Lambda(T)) \Big). \tag{20}
$$

Note that the presence of ties has several effects on computing the conditional null distribution of $\overline{\rho}$. Since $\mathcal{T}$ is smaller than $\mathcal{S}$, we need to compute less permanents and to perform less multiplications, which in itself is favourable. However, we can not apply Theorems 4.3 and 4.5, because, in general, P is not symmetric when ties are present. The structure of the ties effects the number of values that $\overline{\rho}$ attains. The following examples illustrate this. If there are no ties, then the number of different values that ρ attains equals $1 + n(n^2 - 1)/6$. In particular, if $n = 6$, then ρ attains 36 values. In the extreme case that at least one of the samples has one tie of size 6, $\overline{\rho}$ attains one value only. If the available rank scores are $\{1\frac{1}{2}, 1\frac{1}{2}, 3, 4, 5, 6\}$ and $\{1, 2, 3, 4, 5\frac{1}{2}, 5\frac{1}{2}\}$, then the number of values that $\overline{\rho}$ attains goes up to 81. If $\overline{\rho}$ attains few values, then computation time goes down, because this speeds up the time-consuming multiplications that take place after the expansion.

Joeri Roels developed a fast implementation of the above way of computing the conditional null distribution in both Pascal and C. In order to increase speed, it is important to multiply all scores by 2, in order to get rid of the fraction caused by the averaged rank scores.

7 Page's L-statistic

Page's test (see Page (1963)) is a test for ordered alternatives in a randomised block design. Within each block rank scores $1, \ldots, t$ are assigned to t treatments. Let β_j denote the block effect of the jth treatment. The null hypothesis is

$$
H_0 : \beta_1 = \ldots = \beta_t
$$

and the alternative hypothesis is

$$
H_1 : \beta_1 \le \ldots \le \beta_t,
$$

with at least one strict inequality. Let b be the number of blocks. The statistic L is defined as:

$$L = \sum_{j=1}^{t} j \, R_j,$$

where R_j is the sum of the b ranks assigned to treatment j. The following theorem shows why we can use the theorems in the previous sections for computing the exact null distribution of L.

Theorem 7.1 *Under H_0, we have*

$$G_{b,t}(x) = \sum_{k=0}^{\infty} \Pr(L = k) \, x^k = \left(\frac{1}{t!} \, per(\widetilde{P}) \right)^b, \tag{21}$$

where $\widetilde{P}$ is defined in (6).

Proof: Under H_0, there is no block effect and hence the rank configurations of all blocks are identical, as well as independent. Hence,

$$
\begin{aligned}
G_{b,t}(x) &:= \sum_{k=0}^{\infty} \Pr(L = k) \, x^k \\
&= \sum_{k=0}^{\infty} \Pr\left(\sum_{j=1}^{t} j \, R_j = k \right) x^k \\
&= \frac{1}{(t!)^b} \sum_{k=0}^{\infty} \# \left\{ (\sigma_1, \ldots, \sigma_b) \in \bigotimes_{u=1}^{b} \mathcal{S}_t \mid \sum_{j=1}^{t} j \sum_{i=1}^{b} \sigma_i(j) = k \right\} x^k \\
&= \frac{1}{(t!)^b} \sum_{k=0}^{\infty} \# \left\{ (\sigma_1, \ldots, \sigma_b) \in \bigotimes_{u=1}^{b} \mathcal{S}_t \mid \sum_{i=1}^{b} \sum_{j=1}^{t} j \, \sigma_i(j) = k \right\} x^k \\
&= \frac{1}{(t!)^b} \sum_{\ell=0}^{\infty} \sum_{k=\ell}^{\infty} \# \left\{ \sigma_b \in \mathcal{S}_t \mid \sum_{j=1}^{t} j \, \sigma_b(j) = \ell \right\} x^\ell \, . \\
&\quad \# \left\{ (\sigma_1, \ldots, \sigma_{b-1}) \in \bigotimes_{u=1}^{b-1} \mathcal{S}_t \mid \sum_{i=1}^{b-1} \sum_{j=1}^{t} j \, \sigma_i(j) = k - \ell \right\} x^{k-\ell} \\
&= \frac{1}{(t!)^b} \sum_{\ell=0}^{\infty} \# \left\{ \sigma_b \in \mathcal{S}_t \mid \sum_{j=1}^{t} j \, \sigma_b(j) = \ell \right\} x^\ell \, . \\
&\quad \sum_{m=0}^{\infty} \# \left\{ (\sigma_1, \ldots, \sigma_{b-1}) \in \bigotimes_{u=1}^{b-1} \mathcal{S}_t \mid \sum_{i=1}^{b-1} \sum_{j=1}^{t} j \, \sigma_i(j) = m \right\} x^m \\
&= \frac{1}{t!} \, \mathrm{per}\left(\widetilde{P} \right) G_{b-1,t}.
\end{aligned}
$$

The last equality holds, because we recognize the probability generating function of $\widetilde{S_1}$ (see (3) and (5)). Iteration of this argument completes the proof. □

11

8   Exact critical values for n = 19, ... , 22

In this section we present new tables with exact critical values of Spearman's ρ for sample sizes $19 \leq n \leq 22$. We stress the fact that for the case $n = 22$ one has to consider $\prod_{i=19}^{22} i = 175560$

	0.0005	0.001	0.0025	0.005	0.01	0.025	0.05	0.10	0.15	0.20
19	332	370	426	474	530	616	694	788	854	906
20	408	450	516	572	636	736	824	932	1006	1066
21	494	544	618	684	756	868	970	1090	1174	1242
22	592	650	734	808	890	1018	1132	1268	1362	1436

Table 1: Exact left critical values of Spearman's ρ for $n = 19 \ldots 22$.

times more permutations than for the largest case known in the literature up to now, $n = 18$. The results are given in Table 1. We did not print the entire null distribution, because even when we use the symmetry around the mean of ρ, this would involve printing about 6,000 values of the statistic together with their corresponding tail probabilities. The entire distribution can be obtained from the authors on request.

References

Best, D.J., and D.E. Roberts (1975). Algorithm AS 89: The upper tail probabilities of Spearman's rho. *Applied Statistics*, **24**, 377–379.

David, S.T., M.G. Kendall, and A. Stuart (1951). Some questions of distribution in the theory of rank correlation. *Biometrika*, **38**, 131–140.

Jonge, C. de, and A.J. van Montfort (1972). The null distribution of spearman's s when $n = 12$. *Statistica Neerlandica*, **26**, 15–17.

Franklin, L.A. (1988a). The complete exact null-distribution of Spearman's Rho for $n = 12(1)18$. *Journal of Statistical Computation and Simulation*, **29**, 255–269.

Franklin, L.A. (1988b). A note on approximations and convergence in distribution for Spearman's rank correlation coefficient. *Communications in Statistics-Theory and Methods*, **17**, 55–59.

Kendall, M.G., S.F.H. Kendall, and B. Babington Smith (1939). The distribution of Spearman's coefficient of rank correlation in a universe in which all rankings occur an equal number of times. *Biometrika*, **30**, 251–273.

Minc, H. (1978). *Permanents*. Addison-Wesley, Reading.

Nijenhuis, A., and H.S. Wilf (1975). *Combinatorial algorithms*. London Academic Press.

Olds, E.G. (1938). Distributions of sums of squares of rank differences for small numbers of individuals. *Annals of Mathematical Statistics*, **9**, 133–148.

Otten, A. (1973). The null distribution of Spearman's s when when $n = 13(1)16$. *Statistica Neerlandica*, **27**, 19–20.

Owen, D.B. (1962). *Handbook of statistical tables*. Addison-Wesley.

Page, E.B. (1963). Oredered hypothesis for multiple treatments: a significance test for linear ranks. *Journal of the American Statistical Association*, **58**, 216–230.

Ramsay, P.H. (1989). Critical values for Spearman's rank order correlation. *Journal of Educational Statistics*, **14**, 245–254.

Ryser, H.J. (1963). *Combinatorial mathematics*. The Mathematical Association of America.