coin: A Computational Framework for Conditional Inference

Torsten Hothorn[1], Kurt Hornik[2], Mark van de Wiel[3] and Achim Zeileis[2]

[1]Institut für Medizininformatik, Biometrie und Epidemiologie
Friedrich-Alexander-Universität Erlangen-Nürnberg
Waldstraße 6, D-91054 Erlangen, Germany
`Torsten.Hothorn@R-project.org`

[2]Department für Statistik und Mathematik, Wirtschaftsuniversität Wien
Augasse 2-6, A-1090 Wien, Austria
`Kurt.Hornik@R-project.org`
`Achim.Zeileis@R-project.org`

[3]Department of Mathematics, Vrije Universiteit
De Boelelaan 1081a, 1081 HV Amsterdam, The Netherlands
`mark.vdwiel@vumc.nl`

1 Introduction

The **coin** package implements a unified approach for conditional inference procedures commonly known as *permutation tests*. The theoretical basis of design and implementation is the unified framework for permutation tests given by Strasser and Weber (1999). For a very flexible formulation of multivariate linear statistics, Strasser and Weber (1999) derived the conditional expectation and covariance of the conditional (permutation) distribution as well as the multivariate limiting distribution. For a more detailed overview see Hothorn *et al.* (2006).

Conditional counterparts of a large amount of classical (unconditional) test procedures for continuous, categorical and censored data are part of this framework, for example the Cochran-Mantel-Haenszel test for independence in general contingency tables, linear association tests for ordered categorical data, linear rank tests and multivariate permutation tests.

The conceptual framework of permutation tests by Strasser and Weber (1999) for arbitrary problems is available via the generic `independence_test`. Because convenience functions for the most prominent problems are available, users will

not have to use this extremely flexible procedure. Currently, the conditional variants of the following test procedures are available:

`oneway_test`	two- and K-sample permutation test
`wilcox_test`	Wilcoxon-Mann-Whitney rank sum test
`normal_test`	van der Waerden normal quantile test
`median_test`	Median test
`kruskal_test`	Kruskal-Wallis test
`ansari_test`	Ansari-Bradley test
`fligner_test`	Fligner-Killeen test
`chisq_test`	Pearson's χ^2 test
`cmh_test`	Cochran-Mantel-Haenszel test
`lbl_test`	linear-by-linear association test
`surv_test`	two- and K-sample logrank test
`maxstat_test`	maximally selected statistics
`spearman_test`	Spearman's test
`friedman_test`	Friedman test
`wilcoxsign_test`	Wilcoxon-Signed-Rank test
`mh_test`	marginal homogeneity test (Maxwell-Stuart).

Those convenience functions essentially perform a certain transformation of the data, e.g., a rank transformation, and call `independence_test` for the computation of linear statistics, expectation and covariance and the test statistic as well as their null distribution. The exact null distribution can be approximated either by the asymptotic distribution or via conditional Monte-Carlo for all test procedures, the exact null distribution is available for special cases. Moreover, all test procedures allow for the specification of blocks for stratification.

2 Permutation Tests

In the following we assume that we are provided with n observations

$$(\mathbf{Y}_i, \mathbf{X}_i, w_i, b_i), \quad i = 1, \ldots, n.$$

The variables $\mathbf{Y}$ and $\mathbf{X}$ from sample spaces $\mathcal{Y}$ and $\mathcal{X}$ may be measured at arbitrary scales and may be multivariate as well. In addition to those measurements, case weights w and a factor b coding blocks may be available. For the sake of simplicity, we assume $w_i = 1$ and $b_i = 0$ for all observations $i = 1, \ldots, n$ for the moment.

We are interested in testing the null hypothesis of independence of $\mathbf{Y}$ and $\mathbf{X}$

$$H_0 : D(\mathbf{Y}|\mathbf{X}) = D(\mathbf{Y})$$

against arbitrary alternatives. Strasser and Weber (1999) suggest to derive scalar test statistics for testing H_0 from multivariate linear statistics of the form

$$\mathbf{T} = \text{vec}\left(\sum_{i=1}^{n} w_i g(\mathbf{X}_i) h(\mathbf{Y}_i, (\mathbf{Y}_1, \ldots, \mathbf{Y}_n))^\top\right) \in \mathbb{R}^{pq}. \tag{1}$$

Here, $g : \mathcal{X} \to \mathbb{R}^p$ is a transformation of the $\mathbf{X}$ measurements and the *influence function* $h : \mathcal{Y} \times \mathcal{Y}^n \to \mathbb{R}^q$ depends on the responses $(\mathbf{Y}_1, \ldots, \mathbf{Y}_n)$ in a permutation symmetric way. We will give specific examples how to choose g and h later on.

The distribution of $\mathbf{T}$ depends on the joint distribution of $\mathbf{Y}$ and $\mathbf{X}$, which is unknown under almost all practical circumstances. At least under the null hypothesis one can dispose of this dependency by fixing $\mathbf{X}_1, \ldots, \mathbf{X}_n$ and conditioning on all possible permutations S of the responses $\mathbf{Y}_1, \ldots, \mathbf{Y}_n$. This principle leads to test procedures known as *permutation tests*.

The conditional expectation $\mu \in \mathbb{R}^{pq}$ and covariance $\Sigma \in \mathbb{R}^{pq \times pq}$ of $\mathbf{T}$ under H_0 given all permutations $\sigma \in S$ of the responses are derived by Strasser and Weber (1999):

$$\begin{aligned}
\mu &= \mathbb{E}(\mathbf{T}|S) = \mathrm{vec}\left(\left(\sum_{i=1}^n w_i g(\mathbf{X}_i)\right) \mathbb{E}(h|S)^\top\right), \\
\Sigma &= \mathbb{V}(\mathbf{T}|S) \\
&= \frac{\mathbf{w}_\cdot}{\mathbf{w}_\cdot - 1} \mathbb{V}(h|S) \otimes \left(\sum_i w_i g(\mathbf{X}_i) \otimes w_i g(\mathbf{X}_i)^\top\right) \\
&- \frac{1}{\mathbf{w}_\cdot - 1} \mathbb{V}(h|S) \otimes \left(\sum_i w_i g(\mathbf{X}_i)\right) \otimes \left(\sum_i w_i g(\mathbf{X}_i)\right)^\top
\end{aligned} \tag{2}$$

where $\mathbf{w}_\cdot = \sum_{i=1}^n w_i$ denotes the sum of the case weights, and $\otimes$ is the Kronecker product. The conditional expectation of the influence function is

$$\mathbb{E}(h|S) = \mathbf{w}_\cdot^{-1} \sum_i w_i h(\mathbf{Y}_i, (\mathbf{Y}_1, \ldots, \mathbf{Y}_n)) \in \mathbb{R}^q$$

with corresponding $q \times q$ covariance matrix

$$\begin{aligned}
\mathbb{V}(h|S) = \mathbf{w}_\cdot^{-1} \sum_i w_i &\left(h(\mathbf{Y}_i, (\mathbf{Y}_1, \ldots, \mathbf{Y}_n)) - \mathbb{E}(h|S)\right) \\
&\left(h(\mathbf{Y}_i, (\mathbf{Y}_1, \ldots, \mathbf{Y}_n)) - \mathbb{E}(h|S)\right)^\top.
\end{aligned}$$

Having the conditional expectation and covariance at hand we are able to standardize a linear statistic $\mathbf{T} \in \mathbb{R}^{pq}$ of the form (1). Univariate test statistics c mapping an observed linear statistic $\mathbf{t} \in \mathbb{R}^{pq}$ into the real line can be of arbitrary form. An obvious choice is the maximum of the absolute values of the standardized linear statistic

$$c_{\max}(\mathbf{t}, \mu, \Sigma) = \max \left| \frac{\mathbf{t} - \mu}{\mathrm{diag}(\Sigma)^{1/2}} \right|$$

utilizing the conditional expectation μ and covariance matrix Σ. The application of a quadratic form $c_{\mathrm{quad}}(\mathbf{t}, \mu, \Sigma) = (\mathbf{t} - \mu)\Sigma^+(\mathbf{t} - \mu)^\top$ is one alternative, although computationally more expensive because the Moore-Penrose inverse Σ^+ of Σ is involved.

The definition of one- and two-sided p-values used for the computations in the **coin** package is

$$\begin{aligned} P(c(\mathbf{T}, \mu, \Sigma) &\leq c(\mathbf{t}, \mu, \Sigma)) \quad \text{(less)} \\ P(c(\mathbf{T}, \mu, \Sigma) &\geq c(\mathbf{t}, \mu, \Sigma)) \quad \text{(greater)} \\ P(|c(\mathbf{T}, \mu, \Sigma)| &\leq |c(\mathbf{t}, \mu, \Sigma)|) \quad \text{(two-sided)}. \end{aligned}$$

Note that for quadratic forms only two-sided p-values are available and that in the one-sided case maximum type test statistics are replaced by

$$\min\left(\frac{\mathbf{t} - \mu}{\text{diag}(\Sigma)^{1/2}}\right) \quad \text{(less) and} \quad \max\left(\frac{\mathbf{t} - \mu}{\text{diag}(\Sigma)^{1/2}}\right) \quad \text{(greater)}.$$

The conditional distribution and thus the p-value of the statistics $c(\mathbf{t}, \mu, \Sigma)$ can be computed in several different ways. For some special forms of the linear statistic, the exact distribution of the test statistic is trackable. For two-sample problems, the shift-algorithm by Streitberg and Röhmel (1986) and Streitberg and Röhmel (1987) and the split-up algorithm by van de Wiel (2001) are implemented as part of the package. Conditional Monte-Carlo procedures can be used to approximate the exact distribution. Strasser and Weber (1999) proved (Theorem 2.3) that the conditional distribution of linear statistics $\mathbf{T}$ with conditional expectation μ and covariance Σ tends to a multivariate normal distribution with parameters μ and Σ as $n, \mathbf{w}. \to \infty$. Thus, the asymptotic conditional distribution of test statistics of the form $c_{\max}$ is normal and can be computed directly in the univariate case ($pq = 1$) or approximated by means of quasi-randomized Monte-Carlo procedures in the multivariate setting (Genz, 1992). For quadratic forms c_{quad} which follow a χ^2 distribution with degrees of freedom given by the rank of Σ (see Johnson and Kotz, 1970, Chapter 29), exact probabilities can be computed efficiently.

3 Illustrations and Applications

The main workhorse `independence_test` essentially allows for the specification of $\mathbf{Y}, \mathbf{X}$ and b through a formula interface of the form `y ~ x | b`, weights can be defined by a formula with one variable on the right hand side only. Four additional arguments are available for the specification of the transformation g (`xtrans`), the influence function h (`ytrans`), the form of the test statistic c (`teststat`) and the null distribution (`distribution`).

Independent K-Sample Problems. When we want to compare the distribution of an univariate qualitative response $\mathbf{Y}$ in K groups given by a factor $\mathbf{X}$ at K levels, the transformation g is the dummy matrix coding the groups and h is either the identity transformation or a some form of rank transformation.

For example, the Kruskal-Wallis test may be computed as follows (example taken from Hollander and Wolfe, 1999, Table 6.3, page 200):

```
R> library("coin")
R> YOY <- data.frame(
+      length = c(46, 28, 46, 37, 32, 41, 42, 45, 38, 44,
+                 42, 60, 32, 42, 45, 58, 27, 51, 42, 52,
+                 38, 33, 26, 25, 28, 28, 26, 27, 27, 27,
+                 31, 30, 27, 29, 30, 25, 25, 24, 27, 30),
+      site = factor(c(rep("I", 10), rep("II", 10),
+                      rep("III", 10), rep("IV", 10))))
R> it <- independence_test(length ~ site, data = YOY,
+      ytrafo = function(data) trafo(data, numeric_trafo = rank),
+      teststat = "quadtype")
R> it

        Asymptotic General Independence Test

data:  length by site (I, II, III, IV)
chi-squared = 22.8524, df = 3, p-value = 4.335e-05
```

The linear statistic $\mathbf{T}$ is the sum of the ranks in each group and can be extracted via

```
R> statistic(it, "linear")

I   278
II  307
III 119
IV  116
```

Note that `statistic(..., "linear")` currently returns the linear statistic in matrix form, i.e.

$$\sum_{i=1}^{n} w_i g(\mathbf{X}_i) h(\mathbf{Y}_i, (\mathbf{Y}_1, \ldots, \mathbf{Y}_n))^\top \in \mathbb{R}^{p \times q}.$$

The conditional expectation and covariance are available from

```
R> expectation(it)

  I  II III  IV
205 205 205 205

R> covariance(it)

            I        II       III        IV
I   1019.0385 -339.6795 -339.6795 -339.6795
II  -339.6795 1019.0385 -339.6795 -339.6795
III -339.6795 -339.6795 1019.0385 -339.6795
IV  -339.6795 -339.6795 -339.6795 1019.0385
```

and the standardized linear statistic $(\mathbf{T} - \mu)\text{diag}(\Sigma)^{-1/2}$ is

```
R> statistic(it, "standardized")

I    2.286797
II   3.195250
III -2.694035
IV  -2.788013
```

Since a quadratic form of the test statistic was requested via `teststat = "quadtype"`, the test statistic is

```
R> statistic(it)

[1] 22.85242
```

By default, the asymptotic distribution of the test statistic is computed, the p-value is

```
R> pvalue(it)

[1] 4.334659e-05
```

Life is much simpler with convenience functions very similar to those available in package **stats** for a long time. The exact null distribution of the Kruskal-Wallis test can be approximated by 9999 Monte-Carlo replications via

```
R> kw <- kruskal_test(length ~ site, data = YOY,
+                     distribution = approximate(B = 9999))
R> kw

        Approximative Kruskal-Wallis Test

data:  length by site (I, II, III, IV)
chi-squared = 22.8524, p-value < 2.2e-16
```

with p-value (and 99% confidence interval) of

```
R> pvalue(kw)

[1] 0
99 percent confidence interval:
 0.0000000000 0.0005297444
```

Of course it is possible to choose a $c_{\max}$ type test statistic instead of a quadratic form.

Independence in Contingency Tables. Independence in general two- or three-dimensional contingency tables can be tested by the Cochran-Mantel-Haenszel test. Here, both g and h are dummy matrices (example data from Agresti, 2002, Table 7.8, page 288):

```
R> data("jobsatisfaction", package = "coin")
R> it <- cmh_test(jobsatisfaction)
R> it

        Asymptotic Generalized Cochran-Mantel-Haenszel Test

data:  Job.Satisfaction by
         Income (<5000, 5000-15000, 15000-25000, >25000)
         stratified by Gender
chi-squared = 10.2001, df = 9, p-value = 0.3345
```

The standardized contingency table allowing for an inspection of the deviation from the null hypothesis of independence of income and jobsatisfaction (stratified by gender) is

```
R> statistic(it, "standardized")

            Very Dissatisfied A Little Satisfied
<5000               1.3112789         0.69201053
5000-15000          0.6481783         0.83462550
15000-25000        -1.0958361        -1.50130926
>25000             -1.0377629        -0.08983052
            Moderately Satisfied Very Satisfied
<5000                 -0.2478705     -0.9293458
5000-15000             0.5175755     -1.6257547
15000-25000            0.2361231      1.4614123
>25000                -0.5946119      1.2031648
```

Ordered Alternatives. Of course, both job satisfaction and income are ordered variables. When $\mathbf{Y}$ is measured at J levels and $\mathbf{X}$ at K levels, $\mathbf{Y}$ and $\mathbf{X}$ are associated with score vectors $\xi \in \mathbb{R}^J$ and $\gamma \in \mathbb{R}^K$, respectively. The linear statistic is now a linear combination of the linear statistic $\mathbf{T}$ of the form

$$\mathbf{MT} \quad = \quad \text{vec}\left(\sum_{i=1}^{n} w_i \gamma^\top g(\mathbf{X}_i) \left(\xi^\top h(\mathbf{Y}_i, (\mathbf{Y}_1, \ldots, \mathbf{Y}_n)\right)^\top\right) \in \mathbb{R} \text{ with } \mathbf{M} = \xi \otimes \gamma.$$

By default, scores are $\xi = 1, \ldots, J$ and $\gamma = 1, \ldots, K$.

```
R> lbl_test(jobsatisfaction)

        Asymptotic Linear-by-Linear Association Test

data:  Job.Satisfaction (ordered) by
```

```
        Income (<5000 < 5000-15000 < 15000-25000 < >25000)
        stratified by Gender
chi-squared = 6.6235, df = 1, p-value = 0.01006
```

The scores ξ and γ can be specified to the linear-by-linear association test via a list those names correspond to the variable names

```
R> lbl_test(jobsatisfaction,
+      scores = list(Job.Satisfaction = c(1, 3, 4, 5),
+                    Income = c(3, 10, 20, 35)))

        Asymptotic Linear-by-Linear Association Test

data:  Job.Satisfaction (ordered) by
        Income (<5000 < 5000-15000 < 15000-25000 < >25000)
        stratified by Gender
chi-squared = 6.1563, df = 1, p-value = 0.01309
```

Incomplete Randomised Blocks. Rayner and Best (2001), Chapter 7, discuss the application of Durbin's test to data from sensoric experiments, where incomplete block designs are common. As an example, data from taste-testing on ten dried eggs where mean scores for off-flavour from seven judges are given and one wants to assess whether there is any difference in the scores between the ten egg samples. The sittings are a block variable which can be added to the formula via '|'.

```
R> egg_data <- data.frame(
+      scores = c(9.7, 8.7, 5.4, 5.0, 9.6, 8.8, 5.6, 3.6, 9.0,
+                 7.3, 3.8, 4.3, 9.3, 8.7, 6.8, 3.8, 10.0, 7.5,
+                 4.2, 2.8, 9.6, 5.1, 4.6, 3.6, 9.8, 7.4, 4.4,
+                 3.8, 9.4, 6.3, 5.1, 2.0, 9.4, 9.3, 8.2, 3.3,
+                 8.7, 9.0, 6.0, 3.3, 9.7, 6.7, 6.6, 2.8, 9.3,
+                 8.1, 3.7, 2.6, 9.8, 7.3, 5.4, 4.0, 9.0, 8.3,
+                 4.8,3.8,9.3,8.3,6.3,3.8),
+      sitting = factor(rep(c(1:15), rep(4,15))),
+      product = factor(c(1, 2, 4, 5, 2, 3, 6, 10, 2, 4, 6, 7,
+                         1, 3, 5, 7, 1, 4, 8, 10, 2, 7, 8, 9,
+                         2, 5, 8, 10, 5, 7, 9, 10, 1, 2, 3, 9,
+                         4, 5, 6, 9, 1, 6, 7, 10, 3, 4, 9, 10,
+                         1, 6, 8, 9, 3, 4, 7, 8, 3, 5, 6, 8)))
R> yt <- function(data) trafo(data, numeric_trafo = rank,
+                              block = egg_data$sitting)
R> independence_test(scores ~ product | sitting,
+                    data = egg_data, teststat = "quadtype",
+                    ytrafo = yt)

        Asymptotic General Independence Test
```

```
data:  scores by
         product (1, 2, 3, 4, 5, 6, 7, 8, 9, 10)
         stratified by sitting
chi-squared = 39.12, df = 9, p-value = 1.096e-05
```

and the Monte-Carlo p-value can be computed via

```
R> pvalue(independence_test(scores ~ product | sitting,
+      data = egg_data, teststat = "quadtype", ytrafo = yt,
+      distribution = approximate(B = 19999)))

[1] 0
99 percent confidence interval:
 0.000000000 0.000264894
```

If we assume that the products are ordered, the Page test is appropriate and can be computed as follows

```
R> independence_test(scores ~ product | sitting, data = egg_data,
+                    scores = list(product = 1:10),
+                    ytrafo = yt)

        Asymptotic General Independence Test

data:  scores by
         product (1 < 2 < 3 < 4 < 5 < 6 < 7 < 8 < 9 < 10)
         stratified by sitting
Z = -6.2166, p-value = 5.081e-10
alternative hypothesis: two.sided
```

Multiple Tests. One may be interested in testing multiple hypotheses simultaneously, either by using a linear combination of the linear statistic $\mathbf{KT}$, or by specifying multivariate variables $\mathbf{Y}$ and / or $\mathbf{X}$. For example, all pair comparisons may be implemented via

```
R> if (require("multcomp")) {
+      xt <- function(data) trafo(data, factor_trafo = function(x)
+          model.matrix(~x - 1) %*% t(contrMat(table(x), "Tukey")))
+      it <- independence_test(length ~ site, data = YOY, xtrafo = xt,
+          teststat = "max", distribution = approximate(B = 9999))
+      print(pvalue(it))
+      print(pvalue(it, method = "single-step"))
+  }

[1] 0.00010001
99 percent confidence interval:
 5.013042e-07 7.428484e-04
```

```
II - I   0.64726473
III - I  0.03680368
IV - I   0.02110211
III - II 0.00010001
IV - II  0.00010001
IV - III 0.99799980
```

When either g or h are multivariate, single-step adjusted p-values based on maximum-type statistics are computed as described in Westfall and Young (1993), algorithm 2.5 (page 47) and equation (2.8), page 50. Note that for the example shown above only the *minimum* p-value is adjusted appropriately because the subset pivotality condition is violated, i.e., the distribution of the test statistics under the complete null-hypothesis of no treatment effect of `site` is the basis of all adjustments instead of the corresponding partial null-hypothesis.

Another important application are simultaneous tests for many response variables. This problem frequently occurs in microarray expression studies and we shall have a look at an artificial example: 100 variables (from a normal distribution) are to be tested in a one-way classification with $n = 40$ observations. Only the first variable shows a difference and we are interested in both a global test and the adjusted p-values. Here, the example is formulated within the **Biobase** (Gentleman and Carey, 2005) framework (example currently not run because of dependencies problems):

```
R> if (require("Biobase")) {
+
+      p <- 100
+      pd <- new("phenoData",
+                pData = data.frame(group = gl(2, 20)),
+                varLabels = list(group = c("1", "2")))
+      exprs <- matrix(rnorm(p * 40), nrow = p)
+      exprs[1,1:20] <- exprs[1,1:20] + 1.5
+      ex <- new("exprSet", exprs = exprs,
+                           phenoData = pd)
+
+      it <- independence_test(group ~ ., data = ex,
+          distribution = approximate(B = 1000))
+
+      print(pvalue(it))
+      print(which(pvalue(it, method = "step-down") < 0.05))
+
+  }
```

4 Quality Assurance

The test procedures implemented in package **coin** are continuously checked against results obtained by the corresponding implementations in package **stats** (where available). In addition, the test statistics and exact, approximative and asymptotic p-values for data examples given in the **StatXact** -6 user manual (Cytel Inc., 2003) are compared with the results reported in the **StatXact** 6 manual. Step-down multiple adjusted p-values have been checked against results reported by `mt.maxT` from package **multtest** (Pollard *et al.*, 2008). For details on the test procedures we refer to the R transcript files in directory `coin/tests`.

5 Acknowledgements

We would like to thank Helmut Strasser for discussions on the theoretical framework. Henric Nilsson provided clarification and examples for the Maxwell-Stuart test.

References

Agresti A (2002). *Categorical Data Analysis*. John Wiley & Sons, Hoboken, New Jersey, 2nd edition.

Cytel Inc (2003). **StatXact** *6: Statistical Software for Exact Nonparametric Inference*. Cytel Software Corporation, Cambridge, MA. URL http://www.cytel.com/.

Gentleman R, Carey V (2005). **Biobase***: Base Functions for Bioconductor*. R package version 1.5.12, URL http://www.bioconductor.org/.

Genz A (1992). "Numerical Computation of Multivariate Normal Probabilities." *Journal of Computational and Graphical Statistics*, **1**, 141–149.

Hollander M, Wolfe DA (1999). *Nonparametric Statistical Inference*. John Wiley & Sons, New York, 2nd edition.

Hothorn T, Hornik K, van de Wiel MA, Zeileis A (2006). "A Lego System for Conditional Inference." *The American Statistician*, **60**(3), 257–263. doi: 10.1198/000313006X118430.

Johnson NL, Kotz S (1970). *Distributions in Statistics: Continuous Univariate Distributions 2*. John Wiley & Sons, New York.

Pollard KS, Ge Y, Dudoit S (2008). **multtest***: Resampling-Based Multiple Hypothesis Testing*. R package version 1.21.1, URL http://CRAN.R-project.org/package=multtest.

Rayner JCW, Best DJ (2001). *A Contingency Table Approach to Nonparametric Testing*. Chapman & Hall, New York.

Strasser H, Weber C (1999). "On the Asymptotic Theory of Permutation Statistics." *Mathematical Methods of Statistics*, **8**, 220–250. Preprint available from http://epub.wu-wien.ac.at/dyn/openURL?id=oai:epub.wu-wien.ac.at:epub-wu-01_94c.

Streitberg B, Röhmel J (1986). "Exact Distributions for Permutation and Rank Tests: An Introduction to Some Recently Published Algorithms." *Statistical Software Newsletter*, **12**(1), 10–17. ISSN 1609-3631.

Streitberg B, Röhmel J (1987). "Exakte Verteilungen für Rang- und Randomisierungstests im allgemeinen c-Stichprobenfall." *EDV in Medizin und Biologie*, **18**(1), 12–19.

van de Wiel MA (2001). "The Split-Up Algorithm: A Fast Symbolic Method for Computing p Values of Rank Statistics." *Computational Statistics*, **16**, 519–538.

Westfall PH, Young SS (1993). *Resampling-Based Multiple Testing*. John Wiley & Sons, New York.