Metrika, Volume 32, 1985, page 109–123.

One-Sided and Multi-Sided Rank Tests Against Trend in the Case of Concordant Observers

By *T.J. Terpstra*, Enschede[1])

Summary: Optimal rank tests are derived for testing homogeneity of k populations observed by m independent and *concordant* observers against oriented respectively non-oriented contiguity alternatives with respect to a location parameter.

By means of the theorems obtained appropriate rank tests are proposed for testing against a one-sided respectively a multi-sided trend.

1. Introduction

Observers are called *concordant* if an observation x made by some observer on an object implies for another observer α the observation

$$x_\alpha = \phi_\alpha (x), \; x \in R_1, \; \phi_\alpha \in G, \tag{1.1}$$

on the same object, G denoting the group of continuous and strictly increasing functions ϕ onto R_1.

We consider k variables $X_1, \ldots, X_k$ with distribution functions $F_1, \ldots, F_k$. The hypothesis H_0 states that

$$H_0 : F_1 = \cdots = F_k; \tag{1.2}$$

the alternative hypothesis H_1 (one-sided trend) that

$$H_1 : F_1 (x) \geqslant F_2 (x) \geqslant \cdots \geqslant F_k (x), \quad \forall x, \; F_1 \neq F_k \tag{1.3}$$

and the alternative hypothesis H_2 (multi-sided trend) that

$$H_2 : F_{i_1} (x) \geqslant F_{i_2} (x) \geqslant \cdots \geqslant F_{i_k} (x), \quad \forall x, \; F_{i_1} \neq F_{i_k}, \text{ for some permutation } (i_1, \ldots, i_k) \text{ of } (1, 2, \ldots, k). \tag{1.4}$$

---

[1]) *T.J. Terpstra*, Department of Applied Mathematics, Twente University of Technology, Enschede, The Netherlands.

For m observers with unknown ϕ_α, $\alpha \leqslant m$ the observations taken from $X_{\alpha,i} = \phi_\alpha (X_i)$ are denoted by $x_{\alpha,i;l}$, $l \leqslant n_{\alpha,i}$, $i = 1, \ldots, k$. We define

$$n_\alpha := \sum_{i=1}^{k} n_{\alpha,i},$$
$$n := \sum_{\alpha=1}^{m} n_\alpha. \tag{1.5}$$

It is assumed that all n observations are completely independent. By means of these observations we want to test H_0 against H_1 respectively H_2.

As the functions ϕ_α are unknown the m sets of k samples cannot be transformed into one set of k samples from the variables $X_1, \ldots, X_k$. The problems stated above are invariant under the group of transformations $x^*_{\alpha,i;l} = \phi^*_\alpha (x_{\alpha,i;l})$, $l \leqslant n_{\alpha,i}$, $i = 1, \ldots, k$.

Maximal invariants for these groups of transformations are the m vectors of n_α ranks which are obtained by arranging for each α the corresponding n_α observations according to increasing magnitude.

In the sequel we denote by $\mathcal{R}_{\alpha,i}$ the vector of $n_{\alpha,i}$ ordered ranks of the sample $(x_{\alpha,i;l}, l = 1, \ldots, n_{\alpha,i})$, $i = 1, \ldots, k$.

It is assumed that the densities $f_i(x) := F_i'(x)$, $i \leqslant k$ are absolutely continuous, thus for each α the n_α observations are different with probability one. Invariant tests with respect to the transformations ϕ^*_α are based on statistics which are functions of the m vectors of ranks. These statistics have the convenient property that their probability distributions under H_0 do not depend on the common d.f. F.

In this paper rank tests are derived for the foregoing two problems.

For the special case $m = 1$ and ϕ_1 the identity function appropriate rank tests already exist.

The hypothesis H_0 can be tested against H_1 by means of a critical region consisting of large values of the statistic [cf. *Hájek/Šidák*].

$$T(\vec{R}) := \sum_{i=1}^{k} i \left(R_i - \frac{1}{2} n_i (n+1)\right), \tag{1.6}$$

where

$$R_i := \sum_{l \in \mathcal{R}_i} l, \quad i = 1, \ldots, k. \tag{1.7}$$

The hypothesis H_0 can be tested against H_2 by means of a critical region consisting of large values of the statistic [cf. *Kruskal/Wallis*]

$$H(\vec{R}) := \frac{12}{n(n+1)} \sum_i \frac{1}{n_i} \left(R_i - \frac{1}{2} n_i (n+1)\right)^2. \tag{1.8}$$

These statistics are under H_0 and appropriate conditions for large n asymptotically normally distributed respectively distributed according a χ^2-distribution with $k-1$ degrees of freedom.

If F is the logistic distribution function and $f_i(x) = f(x - \theta_i)$ with $\theta_i - \theta_{i-1} = (i-1)\,\Delta$, $i = 2, \ldots, k$, the first test is the locally most powerful rank test for small positive Δ.

Moreover, it is asymptotically a maximin most powerful test for H_0 against a certain class of sequences of "oriented" alternatives which are continguous to H_0 [cf. *Hâjek/Šidâk*]. For further results concerning efficiency and optimality properties of a general class of k-sample rank tests against trend, to which class the foregoing test belongs, cf. *Terpstra.*

The second test is asymptotically a maximin most powerful test for H_0 against the corresponding class of sequences of "non-oriented" alternatives which are contiguous to H_0.

For the generalized problems of testing H_0 against H_1 respectively H_2 by means of the m vectors of ranks for the observations of m concordant observers analogous results are obtained.

In section 2 the locally most powerful rank test is derived against a given regression-alternative for a location parameter.

In section 3 it is shown that a sequence of these tests is asymptotically maximin most powerful against a class of sequences of "oriented" location-alternatives which are contiguous to H_0.

For the generalized problem of testing H_0 against H_2 a rank-test is derived which is asymptotically maximin most powerful against a class of sequences of "non-oriented" location-alternatives which are contiguous to H_0 (section 4).

Using certain efficiency properties of these tests appropriate tests are obtained for testing H_0 against the hypothesis H_1 of a one-sided trend given by (1.3) respectively for testing H_0 against the hypothesis H_2 of a multi-sided trend given by (1.4) (cf. section 5).

As already indicated extensive use will be made of the book "Theory of Rank Tests" by *Hâjek/Šidâk*.

In the sequel references to this book will be abbreviated by *Hâjek/Šidâk*, simultaneously using their method of reffering.

Finally we remark that analogous results can be obtained if we consider alternatives with a scale-parameter [cf. *Hâjek/Šidâk*].

2. The Locally Most Powerful Rank Test Against a Regression-Alternative

We assume that $f_i(x) = f(x - \theta_i)$, $i \leqslant k$ and consider the problem of deriving a locally most powerful rank-test for H_0 against the simple regression-alternative

$$h_1 : \vec{\theta} = \theta\,\vec{\iota} + \Delta\,\vec{c}, \quad \Delta > 0, \tag{2.1}$$

where $\vec{\iota}\,' := (1, \ldots, 1)$, Δ a small positive value and $\vec{c}$ the so called regression vector.

Defining the statistic

$$\begin{cases} R := (R_1, \ldots, R_m), \\ \text{where} \\ R_\alpha := (R_{\alpha,1}, \ldots, R_{\alpha,k}), \end{cases} \tag{2.2}$$

R being a sufficient rank statistic with respect to f_i, $i = 1, \ldots, k$, the critical region for the most powerful rank-test for H_0 against H_1 consists of those R for which $P_1(R) / P_0(R)$ is large, P_0 and P_1 being the probability measures under H_0 respectively H_1.

Now

$$P(R) = \prod_\alpha P(R_\alpha) \tag{2.3}$$

and

$$P_0(R_\alpha) = (n_\alpha!)^{-1} \prod_i n_{\alpha,i}!. \tag{2.4}$$

To obtain $P_1(R)$ we introduce following vectors.

By $\vec{X}_\alpha$ we denote the vector of n_α observations taken by observer α from the variables $X_{\alpha,i}$, $i \leqslant k$ and by $\vec{R}_\alpha$ the corresponding vector of n_α ranks.

By $\vec{X}_{\alpha;0}$ we denote the corresponding vector if k samples of sizes $n_{\alpha,i}$, $i \leqslant k$ are taken of the variables X_i, $i \leqslant k$. By $\vec{R}_{\alpha;0}$ we denote the corresponding vector of n_α ranks.

Then we have for any set S generated by the n_α ranks in the n_α-dimensional Euclidean sample space (cf. (1.1))

$$\begin{aligned} P(\vec{R}_\alpha \in S) &= P(R(\vec{X}_\alpha) \in S) = P(R(\phi_\alpha^{-1}(\vec{X}_\alpha)) \in S) = \\ &= P(R(\vec{X}_{\alpha;0}) \in S) = P(\vec{R}_{\alpha;0} \in S). \end{aligned} \tag{2.5}$$

Thus we obtain

$$P_1(R_\alpha) = \prod_i n_{\alpha,i}!\, P_1(\vec{R}_\alpha), \tag{2.6}$$

where [cf. *Lehmann*, p. 254]

$$P_1(\vec{R}_\alpha) = (n_\alpha!)^{-1}\, E\left(\prod_i \prod_{l \in R_{\alpha,i}} \frac{f(X_\theta^{(l)} - (\theta + \Delta c_i))}{f(X_\theta^{(l)} - \theta)} \right), \tag{2.7}$$

$X_\theta^{(1)}, \ldots, X_\theta^{(n_\alpha)}$ being the order statistics for a random sample of size n_α from the density $f(x - \theta)$.

Now, for an absolutely continuous density f with [*Hâjek/Šidâk*, p. 67 and Lemma 4.8a]

$$\int_{-\infty}^{+\infty} |f'(x)|\, dx < \infty \tag{2.8}$$

we obtain

$$P_1(\vec{R}_\alpha) = (n_\alpha!)^{-1} (1 + \Delta \sum_i c_i S_{\alpha,i}(\gamma_f, R_\alpha) + o(\Delta)), \tag{2.9}$$

where

$$S_{\alpha,i}(\gamma_f, R_\alpha) := \sum_{l \in R_{\alpha,i}} a_n(\gamma_f, l), \tag{2.10}$$

in which the scorefunction γ_f is defined by

$$\gamma_f(u) := \frac{-f'(F^{-1}(u))}{f(F^{-1}(u))},\ 0 < u < 1 \tag{2.11}$$

and the scores by

$$a_n(\gamma_f, l) := E\,\gamma_f(U_n^{(l)}),\ \ l \leqslant n, \tag{2.12}$$

$U_n^{(1)}, \ldots, U_n^{(n)}$ beint the order statistics of a random sample of size n from the uniform distribution on [0, 1].

By means of the foregoing results we finally obtain

$$\frac{P_1(R)}{P_0(R)} = 1 + \Delta \sum_i c_i S_i(\gamma_f, R) + o(\Delta), \tag{2.13}$$

in which

$$S_i(\gamma_f, R) := \sum_\alpha S_{\alpha,i}(\gamma_f, R_\alpha). \tag{2.14}$$

Thus, for sufficiently small Δ, the most powerful rank-test against H_1 is given by a critical region consisting of large values of the statistic

$$T_{\vec{c}}(\vec{S}(\gamma_f)) := \vec{c}\,'\vec{S}(\gamma_f), \tag{2.15}$$

in which

$$\vec{S}\,'(\gamma_f) := (S_1(\gamma_f, R), \ldots, S_k(\gamma_f, R)). \tag{2.16}$$

Thus we obtain

Theorem 2.1

If the density f is absolutely continuous and satisfies (2.8), then the test whisch rejects H_0 for large values of the statistic $T_{\vec{c}}(\vec{S}(\gamma_f))$ is for small positive Δ the locally most powerful rank-test against the hypothesis

$$K_{f,\Delta\vec{c}} := \left\{(f, \vec{\theta}) \mid \vec{\theta} = \theta\,\vec{1} + \Delta\,\vec{c}, \ -\infty < \theta < +\infty\right\}. \tag{2.17}$$

3. Asymptotic Properties of the One-Sided Rank Test Against Location Alternatives

We consider a probability distribution $f(x;\theta) = f(x-\theta)$ with finite Fisher-information

$$\left\{\begin{array}{l} 0 < I_f < \infty, \\ \text{where} \\ I_f := \int_0^1 \gamma_f^2(u)\,du. \end{array}\right. \tag{3.1}$$

Then condition (2.8) also holds which implies that [cf. *Hâjek/Šidâk*, p. 66]

$$\sum_{l=1}^{n} a_n(\gamma_f, l) = n \int_0^1 \gamma_f(u)\,du = 0. \tag{3.2}$$

Remark

It follows that $T_{\vec{c}}(\vec{S}(\gamma_f))$ is independent of a shift in the vector $\vec{c}$. In the sequel it will be assumed that $c_1 + \cdots + c_k = 0$. Using the equality

$$T_{\vec{c}}(\vec{S}(\gamma_f)) = \sum_{\alpha} T_{\vec{c}}(\vec{S}_{\alpha}(\gamma_f)), \tag{3.3}$$

we immediately obtain from Theorem V. 1.5. a in *Hâjek/Šidâk*

Theorem 3.1

If for $\nu = 1, 2, \ldots, m$ and k do not depend on ν and

$$\min_{\alpha} \frac{\sum_i n_{\alpha,i;\nu}(c_{i;\nu} - \bar{c}_{\alpha;\nu})^2}{\max_i (c_{i;\nu} - \bar{c}_{\alpha;\nu})^2} \to \infty, \tag{3.4}$$

where

$$\bar{c}_{\alpha} := (n_{\alpha})^{-1} \sum_i n_{\alpha,i}\, c_i, \tag{3.5}$$

then under H_0 and condition (3.1) the statistic $T_{\vec{c}}(\vec{S}(\gamma_f))$ is for $\nu \to \infty$ asymtotically normal (μ_c, σ_c^2) with

$$\begin{cases} \mu_c = 0, \\ \sigma_c^2 = I_f \sum_\alpha \sum_i n_{\alpha,i} (c_i - \bar{c}_\alpha)^2. \end{cases} \tag{3.6}$$

We now consider the problem of determining the limit distribution of $T_{\vec{c}}(\vec{S}(\gamma_f))$ under the alternative hypothesis that $X_{i,\nu}$, $i = 1, 2, \ldots, k$ is distributed according to the density $g(x - d_{i,\nu})$ with $\{\vec{d}_\nu\}$ satisfying for each α the condition that

$$\begin{cases} \sum_i n_{\alpha,i} (d_{i;\nu} - \bar{d}_{\alpha;\nu})^2 \to \delta_\alpha^2 > 0, \\ \max_i (d_{i;\nu} - \bar{d}_{\alpha;\nu})^2 \to 0. \end{cases} \tag{3.7}$$

Then for each α the sequence of densities induced by $\{\vec{X}_{\alpha;0;\nu}\}$ (cf. section 2) is contiguous to the sequence of densities induced by $\{\vec{X}_{\alpha;0;\nu}\}$ under the supposition that each $X_{i,\nu}$ is distributed according to the same density $g(x - \bar{d}_{\alpha;\nu})$. The latter supposition satisfies the hypothesis $H_{0,\nu}$.

Remark

As condition (3.7) is invariant for a shift in the location-vector $\vec{d}$ we introduce for each $\vec{d}$ with $d_1 + \cdots + d_k = 0$ the class $K_{g;\delta_1^2,\ldots,\delta_m^2}(\{\vec{d}_\nu\})$ of contiguity alternatives $\{\vec{\theta}_\nu\}$ with $\theta_{i,\nu} = d_{i,\nu} + \theta_\nu$, θ_ν being arbitrary. For these alternatives the vector $\vec{\theta}'$, with $\theta_i' := \theta_i' - k^{-1}(\theta_1 + \cdots + \theta_k)$, $i = 1, 2, \ldots, k$, is equal to the known vector $\vec{d}$.

Theorem 3.2

If for $\nu = 1, 2, \ldots, m$ and k do not depend on ν and (3.1) holds for f and g, then under condition (3.4) and under the alternative $K_{g;\delta_1^2,\ldots,\delta_m^2}(\{\vec{d}\})$ given by (3.7) the variable $T_{\vec{c}}(\vec{S}(\gamma_f))$ is asymptotically normal (μ_{cd}, σ_c^2) with σ_c^2 given by (3.6) and

$$\mu_{cd} = (\gamma_f, \gamma_g)\, \vec{d}'\, C(\vec{n}_1, \ldots, \vec{n}_m)\, \vec{c}, \tag{3.8}$$

where

$$(\gamma_f, \gamma_g) := \int_0^1 \gamma_f(u)\, \gamma_g(u)\, du, \tag{3.9}$$

$$C(\vec{n}_1, \ldots, \vec{n}_m) := \sum_\alpha C(\vec{n}_\alpha) \tag{3.10}$$

and

$$C(\vec{n}_\alpha) = n_\alpha^{-1}\begin{pmatrix} n_{\alpha 1}(n_\alpha - n_{\alpha 1}) - n_{\alpha 1} n_{\alpha 2} \cdots - n_{\alpha 1} n_{\alpha k} \\ \hline -n_{\alpha k} n_{\alpha 1} - n_{\alpha k} n_{\alpha 2} \cdots n_{\alpha k}(n_\alpha - n_{\alpha k}) \end{pmatrix}. \tag{3.11}$$

Proof

The proof is quite analogous to the proof of Theorem VI. 2.4 in *Hájek/Šidák*. It follows by applying LeCAM's third lemma. We obtain

$$\mu_{dc} = \mathrm{cov}_0\,(T_{\vec{c}}, T_{\vec{d}}) = E_0\,(\vec{d}'\vec{S}_g\,\vec{S}_f'\vec{c}) = \vec{d}'\,\mathrm{cov}_0\,(\vec{S}_g, \vec{S}_f)\,\vec{c}, \tag{3.12}$$

where

$$\mathrm{cov}_0\,(\vec{S}_g, \vec{S}_f) = \sum_\alpha \mathrm{cov}_0\,(\vec{S}_{\alpha,g}, \vec{S}_{\alpha,f}) \tag{3.13}$$

and [cf. *Hájek/Šidák*, p. 217]

$$\mathrm{cov}_0\,(\vec{S}_{\alpha,g}, \vec{S}_{\alpha,f}) = (\gamma_f, \gamma_g)\,C(\vec{n}_\alpha). \tag{3.14}$$

It immediately follows that a sequence of tests based on large values of the variable $T_{\vec{c}}(\vec{S}(\gamma_f))$ is (among all tests based on a variable of this structure) asymptotically most powerful against the corresponding class of contiguity alternatives $K_{f;\delta_1^2,\ldots,\delta_m^2}(\{c\})$ defined by (3.7). As

$$P_{H_0}\,[T_{\vec{c}} \geqslant \xi_{1-\alpha}\,\sigma_c] \to \alpha, \tag{3.15}$$

we obtain for the asymptotic power

$$P_{(f,\vec{c});\delta_1^2,\ldots,\delta_m^2}\,[T_{\vec{c}} \geqslant \xi_{1-\alpha}\,\sigma_c] \to 1 - \phi\,(\xi_{1-\alpha} - (I_f \sum_\alpha \delta_\alpha^2)^{1/2}), \tag{3.16}$$

where $\xi_\alpha = \phi^{-1}(\alpha)$ and ϕ the standard normal distribution function. It follows that the efficiency of the test based on large values of $T_{\vec{c}}(\vec{S}(\gamma_f))$ for testing H_0 against $K_{g;\delta_1^2,\ldots,\delta_m^2}(\{d\})$ is, under the condition that

$$\frac{\sum_\alpha \sum_i n_{\alpha,i;\nu}\,(c_{i;\nu} - \bar{c}_{\alpha;\nu})\,(d_{i;\nu} - \bar{d}_{\alpha;\nu})}{\left\{\sum_\alpha \sum_i n_{\alpha,i;\nu}\,(c_{i;\nu} - \bar{c}_{\alpha;\nu})^2 \sum_\alpha \sum_i (d_{i;\nu} - \bar{d}_{\alpha;\nu})^2\right\}^{1/2}} \to \rho \geqslant 0, \tag{3.17}$$

equal to $\rho^2\rho_{f,g}^2$, where

$$\rho_{f,g} := \frac{\int_0^1 \gamma_f(u)\,\gamma_g(u)\,du}{\left\{\int_0^1 \gamma_f^2(u)\,du \int_0^1 \gamma_g^2(u)\,du\right\}^{1/2}} \geqslant 0 \tag{3.18}$$

[cf. *Hájek/Šidák*, VII.2].

For each ν a maximin most powerful test exists for testing H_0 against the class of contiguity-alternatives $K_{f;\delta_1^2,\ldots,\delta_m^2}(\{\vec{c}\})$. Denoting the maximin power of this test by $\beta_\gamma(\alpha; H_0, K_{f;\delta_1^2,\ldots,\delta_m^2}(\{\vec{c}\}))$ the following theorem can be shown [cf. *Hájek/Šidák*, Theorem VII.1.3].

Theorem 3.3

If for $\nu = 1, 2, \ldots, m$ and k do not depend on ν, then under condition (3.1)

$$\beta_\nu(\alpha; H_0, K_{f;\delta_1^2,\ldots,\delta_m^2}(\{\vec{c}\})) \to 1 - \phi(\xi_{1-\alpha} - (I_f \sum_\alpha \delta_\alpha^2)^{1/2}). \tag{3.19}$$

The maximin-power is asymptotically reached by the rank test based on large values of the rank statistic $T_{\vec{c}}(\vec{S}(\gamma_f))$.

4. A Multi-Sided Rank Test Against Location Alternatives and its Asymptotic Properties

In the foregoing section it has been shown that the sequence of tests based on large values of $T_{\vec{c}}(\vec{S}(\gamma_f))$ is asymptotically maximin for testing H_0 against the class of contiguity alternatives $K_{f;\delta_1^2,\ldots,\delta_m^2}(\{\vec{c}\})$ with a *known* sequence of vectors $\vec{c}\,(c_1 + \cdots + c_k = 0)$.

We now consider the class $K_{f;\delta_1^2,\ldots,\delta_m^2}$ of alternatives $\{\vec{\theta}\}$ also satisfying condition (3.7), but for this class the vectors $\vec{\theta}'$, with $\theta_i' := \theta_i - k^{-1}(\theta_1 + \cdots + \theta_k)$, are *unknown*.

To obtain an asymptotically maximin rank test for H_0 against this alternative is seems promising to investigate first the limit distribution of the vector $\vec{S}_\nu(\gamma_f)$ under H_0 and under $K_{f;\delta_1^2,\ldots,\delta_m^2}$.

Analogously to Theorem V.2.2 and Theorem VI.3.1 in *Hájek/Šidák* we can prove

Theorem 4.1

If for $\nu \to \infty$ m and k do not depend on ν and $0 < I_f < \infty$, then under $H_{0,\nu}$ and for $\min_{\alpha,i} n_{\alpha,i;\nu} \to \infty$ the vector $\vec{S}_\nu$ is asymptotically multinormal $(\vec{0}, I_f C(\vec{n}_{1;\nu}, \ldots, \vec{n}_{m;\nu}))$, where I_f and $C(\vec{n}_1, \ldots, \vec{n}_m)$ are given by (3.1) respectively (3.10) and (3.11).

Theorem 4.2

If for $\nu \to \infty$ m and k do not depend on ν and $0 < I_f < \infty$, then under the alternative $K_{f;\delta_1^2, \ldots, \delta_m^2}$ defined by (3.7) and for $\min\limits_{\alpha,i} n_{\alpha,i;\nu} \to \infty$ the vector $\vec{S}_\nu(\gamma_f)$ is asymptotically multinormal

$$(\vec{\mu}_{f;\nu}(\vec{\theta}_\nu), I_f\, C(\vec{n}_{1;\nu}, \ldots, \vec{n}_{m;\nu})), \text{ where (cf. (3.8))}$$

$$\vec{\mu}_f(\vec{\theta}) = I_f \;C(\vec{n}_1, \ldots, \vec{n}_m)\,\vec{\theta}. \tag{4.1}$$

The asymptotic analogue of the foregoing problem is thus that of testing for a vector $\vec{Z} = (Z_1, \ldots, Z_k)'$ with a $(k-1)$-dimensional normal distribution with mean $\vec{\eta} = E\vec{Z}$ and given covariance matrix A the hypothesis H that $\vec{\eta} = \vec{0}$. This problem is equivalent with a linear hypothesis problem [cf. *Lehmann*, section 7.12] and for this problem a UMP invariant test exists given by the rejection region $\sum\limits_{i=1}^{k-1} \sum\limits_{j=1}^{k-1} b_{ij}\, z_i\, z_j \geqslant C$, in which b_{ij} are the elements of the inverse matrix of the non-singular covariance-matrix of the variables $Z_1, \ldots, Z_{k-1}$.

This quadratic test-statistic has under H a χ^2_{k-1}-distribution and under the alternative $\vec{\eta}$ a non-central χ^2-distribution with $k-1$ degrees of freedom and non-centrality parameter $\lambda^2 = \sum\limits_{i=1}^{k-1} \sum\limits_{j=1}^{k-1} b_{ij}\, \eta_i\, \eta_j$.

This invariant test has also the character of a maximin test [cf. *Lehmann*, p. 338]. Thus it is obvious to introduce the rank-statistic

$$H_{\vec{n}_1, \ldots, \vec{n}_m}(\vec{S}(\gamma_f)) := I_f^{-1}\, \vec{S}'_{k-1}(\gamma_f)\, C^{-1}_{k-1,k-1}(\vec{n}_1, \ldots, \vec{n}_m)\, \vec{S}_{k-1}(\gamma_f), \tag{4.2}$$

in which $\vec{S}'_{k-1} := (S_1, \ldots, S_{k-1})$ and $C_{k-1,k-1}$ the matrix which is obtained from $C(\vec{n}_1, \ldots, \vec{n}_m)$ by deleting the k-th row and k-th column. It may be expected that this rank-statistic has under H_0 asymptotically a χ^2-distribution with $k-1$ degrees of freedom and under the alternative $K_{f;\delta_1^2, \ldots, \delta_m^2}$ a non-central χ^2-distribution with $k-1$ degrees of freedom. We first derive a symmetric expression for the rank-statistic by making use of the following consideration.

The vector $\vec{S}$ is a $(k-1)$-dimensional vector and the rank of $C(\vec{n}_1, \ldots, \vec{n}_m)$ is also equal to $k-1$.

Then, apart from a translation vector, a unique solution $\vec{S}^*$ exists for the equation

$$\vec{S} = C(\vec{n}_1, \ldots, \vec{n}_m)\,\vec{S}^*. \tag{4.3}$$

Then we obtain

$$H_{\vec{n}_1,\ldots,\vec{n}_m}(\vec{S}(\gamma_f)) = I_f^{-1}\,\vec{S}^{*\prime}\,C_{k,k-1}\,C_{k-1,k-1}^{-1}\,C_{k-1,k}\,\vec{S}^{*} =$$

$$= I_f^{-1}\,\vec{S}^{*\prime}\,C\,\vec{S}^{*} = I_f^{-1}\sum_{\alpha}\vec{S}^{*\prime}\,C_\alpha\,\vec{S}^{*}, \tag{4.4}$$

where

$$\left\{\begin{array}{l} \vec{S}^{*\prime}\,C_\alpha\,\vec{S}^{*} = \sum_i n_{\alpha,i}\,(S_i^* - \bar{S}_\alpha^*)^2 \\ \text{and} \\ \bar{S}_\alpha^* := n_\alpha^{-1}\sum_i n_{\alpha,i}\,S_i^*. \end{array}\right. \tag{4.5}$$

For the special case $m = 1$ it can be shown that $S_i^* - \bar{S}^* = n_i^{-1}\,S_i$ and we obtain

$$H_{\vec{n}}(\vec{S}(\gamma_f)) = I_f^{-1}\sum_i \frac{S_i^2(\gamma_f)}{n_i}, \tag{4.6}$$

the known *Kruskal-Wallis* statistic for the k-sample problem. This also suggests to introduce the vectors

$$\vec{S}_\alpha^* := C_\alpha\,\vec{S}^*,\ \alpha \leqslant m. \tag{4.7}$$

Then we obtain

$$H_{\vec{n}_1,\ldots,\vec{n}_m}(\vec{S}(\gamma_f)) = \sum_\alpha H_{\vec{n}_\alpha}(\vec{S}_\alpha^*(\gamma_f)), \tag{4.8}$$

in which $H_{\vec{n}_\alpha}(\vec{S}_\alpha^*(\gamma_f))$ is given by (4.6).

From the definition of $H_{\vec{n}_1,\ldots,\vec{n}_m}(\vec{S}(\gamma_f))$ we immediately obtain [*Hájek/Šidák*, the last part of the proof of Theorem V.2.2 and also Corollary VI.3.1].

Theorem 4.3

If for $\nu \to \infty$ m and k do not depend on ν, $0 < I_f < \infty$ and $\min_{\alpha,i} n_{\alpha,i;\nu} \to \infty$, then under the hypothesis $H_{0;\nu}$ the variable $H_{\vec{n}_{1;\nu},\ldots,\vec{n}_{m;\nu}}(\vec{S}_\nu(\gamma_f))$ has asymptotically a χ^2-distribution with $k-1$ degrees of freedom; under the alternative $K_{f;\delta_1^2,\ldots,\delta_m^2}$ defined by (3.7) the variable $H_{\vec{n}_{1;\nu},\ldots,\vec{n}_{m;\nu}}(\vec{S}_\nu(\gamma_f))$ has asymptotically a non-central χ^2-distribution with $k-1$ degrees of freedom and non-centrality parameter

$$\delta^2 = \lim_{\nu\to\infty} H_{\vec{n}_{1;\nu},\ldots,\vec{n}_{m;\nu}} (\vec{\mu}_{f;\nu}(\vec{\theta}_\nu)) = I_f \sum_\alpha \delta^2_\alpha . \tag{4.9}$$

If we denote the $(1-\alpha)$-quantile of the χ^2-distribution with $k-1$ degrees of freedom by $\chi^2_{1-\alpha,k-1}$ amd by $F_{k-1}(.,\delta^2)$ the distribution function of the non-central χ^2 with $k-1$ degrees of freedom and non-centrality parameter δ^2, then it follows from Theorem 4.3 that

$$P_{H_0}[H_{\vec{n}_1,\ldots,\vec{n}_m}(\vec{S}(\gamma_f)) \geqslant \chi^2_{1-\alpha,k-1}] \to \alpha \tag{4.10}$$

and

$$P_{(f,\vec{\theta});\delta_1^2,\ldots,\delta_m^2}[H_{\vec{n}_1,\ldots,\vec{n}_m}(\vec{S}(\gamma_f)) \geqslant \chi^2_{1-\alpha,k-1}] \to$$
$$\to 1 - F_{k-1}(\chi^2_{1-\alpha,k-1}, I_f \sum_\alpha \delta^2_\alpha). \tag{4.11}$$

Then we can prove analogously to Theorem VII.1.4 in *Håjek/Šidåk.*

Theorem 4.4

If for $\nu\to\infty$ m and k do not depend on ν and $0 < I_f < \infty$ then for $\min_{\alpha,i} n_{\alpha,i;\nu} \to \infty$

$$\beta_\nu(\alpha; H_0, K_{f;\delta_1^2,\ldots,\delta_m^2}) \to 1 - F_{k-1}(\chi^2_{1-\alpha,k-1}, I_f \sum_\alpha \delta^2_\alpha). \tag{4.12}$$

This power is asymptotically reached by the rank test based on large values of $H_{\vec{n}_1,\ldots,\vec{n}_m}(\vec{S}(\gamma_f))$.

It can be shown [cf. *Håjek/Šidåk*, Theorem VII.2.3] that the efficiency of the rank test based on $H_{\vec{n}_1,\ldots,\vec{n}_m}(\vec{S}(\gamma_f))$ for testing H_0 against the alternative $K_{g;\delta_1^2,\ldots,\delta_m^2}$ is equal to $\rho^2_{f,g}$ $(\rho_{f,g} \geqslant 0)$, where $\rho_{f,g}$ is defined by (3.18).

It is interesting to investigate the influence of the fact of more than one observer on the power of the test. Thus we compare the power of the test for the design $\vec{n}_1,\ldots,\vec{n}_m$ with the power of the test for the design with one observer and $\vec{n}_1$ taken equal to

$$\vec{n}^* := \sum_\alpha \vec{n}_\alpha . \tag{4.13}$$

From the equality

$$\left\{ \begin{array}{l} \sum_\alpha \sum_i n_{\alpha,i}(\theta_i - \bar{\theta}_\alpha)^2 = \sum_i n_i^*(\theta_i - \bar{\theta})^2 - \sum_\alpha n_\alpha(\bar{\theta}_\alpha - \bar{\theta})^2, \\ \text{where} \\ \bar{\theta} = n^{-1}\sum_\alpha n_\alpha \bar{\theta}_\alpha, \end{array} \right. \tag{4.14}$$

it follows that with respect to contiguity alternatives the power of the m-observer-test is at most equal to the power of the one-observer-test, this equality being reached if

$$\sum_{\alpha} n_{\alpha;\nu} (\bar{\theta}_{\alpha;\nu} - \bar{\theta}_{\nu})^2 \to 0. \tag{4.15}$$

5. Testing Against a One-Sided and a Multi-Sided Trend

First we consider the problem of testing H_0 against H_1 (cf. (1.2) and (1.3)).

For the special case that under H_1 $F_i(x) = F(x - c_i), i = 1, 2, \ldots, k$, it has been shown in section 3 that the test with a critical region consisting of large values of the statistic $T_{\vec{c}}(\vec{S}(\gamma_f))$ is an asymptotically optimal rank test.

If we do not know anything about $F_1, \ldots, F_k$ and $\vec{c}$ then we may prefer the simplicity of the statistic to the efficiency of the corresponding test and consider the statistic

$$T(\vec{R}_1, \ldots, \vec{R}_m) = \sum_i i\, S_i(\vec{R}_1, \ldots, \vec{R}_m), \tag{5.1}$$

where

$$S_i(\vec{R}_1, \ldots, \vec{R}_m) = \sum_{\alpha} \frac{1}{n_\alpha + 1} \{2R_{\alpha,i} - n_{\alpha,i}(n_\alpha + 1)\} \tag{5.2}$$

and

$$R_{\alpha,i} = \sum_{l \in \mathcal{R}_{\alpha,i}} l, \tag{5.3}$$

$\mathcal{R}_{\alpha,i}$ denoting the vector of $n_{\alpha,i}$ ordered ranks for the sample $X_{\alpha,i;l}, l \leqslant n_{\alpha,i}$, if all observations in row α are arranged according to increasing magnitude.

The hypothesis H_0 will be rejected against H_1 for large values of $T(\vec{R}_1, \ldots, \vec{R}_m)$.

Under H_0 and the conditions of Theorem 3.1 the variable T is asymptotically normally distributed with mean and variance given by [*Hájek/Šidák*, p. 61]

$$\begin{cases} E(T(\vec{R}_1, \ldots, \vec{R}_m) \mid H_0) = 0, \\ \operatorname{var}(T(\vec{R}_1, \ldots, \vec{R}_m) \mid H_0) = \frac{1}{3} \sum_{\alpha} \frac{n_\alpha}{n_\alpha + 1} \sum_i n_{\alpha,i} (i - n_\alpha^{-1} \sum_j j\, n_{\alpha,j})^2 . \end{cases} \tag{5.4}$$

In the same way, for testing H_0 against the hypothesis H_2 of a multisided trend (cf. (1.4)), an appropriate test is obtained by using the statistic $H_{\vec{n}_1, \ldots, \vec{n}_m}(\vec{S}(\gamma_f))$

defined by (4.4) and (4.5), taking for $\vec{S}$ the vector defined by (5.2) and (5.3).

Then we obtain the statistic

$$H_{\vec{n}_1,\ldots,\vec{n}_m}(\vec{R}_1,\ldots,\vec{R}_m) = 3 \sum_\alpha \sum_i n_{\alpha,i} (S_i^* - \bar{S}_\alpha^*)^2, \tag{5.5}$$

where $\vec{S}^*$ is a solution of (4.3), in which $\vec{S}$ is given by (5.2) and (5.3) and $\bar{S}_\alpha^*$ is defined by (4.5).

Another expression follows from (4.8) and (4.6), obtaining

$$H_{\vec{n}_1,\ldots,\vec{n}_m}(\vec{R}_1,\ldots,\vec{R}_m) = 3 \sum_\alpha \sum_i \frac{S_{\alpha,i}^{*\,2}}{n_{\alpha,i}}, \tag{5.6}$$

in which $S_{\alpha,i}^*$ follows from (4.3) and (4.7).

The hypothesis H_0 will be rejected against H_2 for large values of $H(\vec{R}_1,\ldots,\vec{R}_m)$.

Under the hypothesis H_0 and for $\min_{\alpha,i} n_{\alpha,i} \to \infty$ (cf. Theorem 4.3) the variable $H(\vec{R}_1,\ldots,\vec{R}_m)$ has asymptotically a χ^2-distribution with $k-1$ degrees of freedom.

Finally we remark that the foregoing property also holds for $m \to \infty$ and

$$n_{\alpha,i} = n_i, \quad \alpha \leqslant m, \quad i \leqslant k. \tag{5.7}$$

Then the Central Limit Theorem for multi-dimensional probability distributions may be applied to the vector $\vec{S} = \sum_\alpha \vec{S}_\alpha$, from which it follows that the statistic

$$H_{m;n_1,\ldots,n_k}(\vec{R}_1,\ldots,\vec{R}_m) = \frac{3}{m(n+1)^2} \sum_i \frac{(2R_i - m(n+1)n_i)^2}{n_i}, \tag{5.8}$$

in which

$$R_i = \sum_\alpha R_{\alpha,i}, \tag{5.9}$$

has under H_0 and for $m \to \infty$ asymptotically a χ^2-distribution with $k-1$ degrees of freedom.

Foregoing statistic is for the special case that $n_1 = \cdots = n_k = 1$ equal to *Friedman's* statistic for the problem of m rankings [cf. *Friedman; Hájek/Šidák*, p. 117].

References

Friedman, M.: The use of ranks to avoid the assumption of normality implicit in the analysis of variance. JASA **32**, 1937, 675–701.

Hájek, J., and *Z. Šidák*: Theory of Rank Tests. Prague 1967.

Kruskal, W.H., and *W.A. Wallis*: Use of ranks in one-criterion variance analysis. JASA **47**, 1952, 583–621. Errata: JASA **48**, 1953, p. 910.

Lehmann, E.L.: Testing Statistical Hypotheses. New York 1959.

Terpstra, T.J.: Efficiency and optimality properties of a class of k-sample rank tests against trend. Metrika **27**, 1980, 225–235.

Received December 14, 1981
(Revised version October 11, 1982)