Kybernetika

František Rublík

Estimates of the covariance matrix of vectors of u-statistics and confidence regions for vectors of Kendall's tau

Kybernetika, Vol. 52 (2016), No. 2, 280–293

Persistent URL: http://dml.cz/dmlcz/145775

KYBERNETIKA — VOLUME *52* (2016), NUMBER 2, PAGES 280–293

ESTIMATES OF THE COVARIANCE MATRIX OF VECTORS OF U-STATISTICS AND CONFIDENCE REGIONS FOR VECTORS OF KENDALL'S TAU

František Rublík

Consistent estimators of the asymptotic covariance matrix of vectors of U-statistics are used in constructing asymptotic confidence regions for vectors of Kendall's correlation coefficients corresponding to various pairs of components of a random vector. The regions are products of intervals computed by means of a critical value from multivariate normal distribution. The regularity of the asymptotic covariance matrix of the vector of Kendall's sample coefficients is proved in the case of sampling from continuous multivariate distribution under mild conditions. The results are applied also to confidence intervals for the coefficient of agreement. The coverage and length of the obtained (multivariate) product of intervals are illustrated by simulation.

Keywords: consistent estimate of asymptotic covariance matrix, U-statistics, vector of Kendall's coefficients, coefficient of agreement, confidence interval

Classification: 62G05, 62G15

1. INTRODUCTION

Kendall's tau is a common measure of dependence in a pair of random variables. It is commonly used to check for independence and when the estimate of Kendall's tau is significantly different from 0, a confidence interval with appropriate coverage can be constructed, owing to the asymptotic normality of this U-statistic. This paper considers the construction of rectangular asymptotic confidence regions for vectors of Kendall's tau measuring the dependence between various pairs of components of a multivariate random vector. As explained in Section 3, the regions can be used for construction of multiple comparisons procedures for the hypothesis of independence, because the pairs of coordinates for which the corresponding subintervals do not contain zero can be declared as the ones violating the assumption of independence. In general, the resulting rule can serve for detecting the violators of the quasi-independence. In accordance with [4] quasi-independence means that for each of the chosen pairs of coordinates of the random vector the corresponding pair of coordinate variables is a pair of independent random variables. The results of the paper include also an assertion on regularity of the asymptotic covariance matrix of the vector of sample Kendall's coefficients, and the construction of the asymptotic confidence intervals for the coefficient of agreement,

DOI: 10.14736/kyb-2016-2-0280

which is an extension of assertions on the asymptotic distribution of this coefficient presented in [3] and [2].

Throughout the paper we assume that $\{\mathbf{X}_j\}_{j=1}^{\infty}$ are independent identically distributed copies of a random vector $\mathbf{X}$ taking values in the space $\mathbb{R}^d$ of column vectors. Further we assume that $\boldsymbol{\Psi}(\mathbf{x}_1,\dots,\mathbf{x}_m) : \mathbb{R}^d \times \dots \times \mathbb{R}^d \to \mathbb{R}^s$ is a measurable symmetric function of its arguments $\mathbf{x}_i \in \mathbb{R}^d$ such that

$$E(\|\boldsymbol{\Psi}(\mathbf{X}_1,\dots,\mathbf{X}_m)\|^2) < +\infty. \tag{1.1}$$

The corresponding multivariate U-statistic is defined by the formula

$$\mathbf{U}_n = \mathbf{U}(\mathbf{X}_1,\dots,\mathbf{X}_n) = \frac{1}{\binom{n}{m}} \sum_{r \in C(m,n)} \boldsymbol{\Psi}(\mathbf{X}_{r_1},\dots,\mathbf{X}_{r_m}), \tag{1.2}$$

where $C(m,n)$ denotes the collection of all subsets of $\{1,\dots,n\}$ consisting of m elements.

The topic of Section 2 is estimation of the asymptotic covariance matrix of U-statistics. Results on Kendall's correlation coefficient and on the coefficient of agreement can be found in Section 3, and simulation results are in Section 4.

2. ESTIMATION OF THE ASYMPTOTIC COVARIANCE MATRIX OF A U-STATISTIC

Obviously, the mean vector $\boldsymbol{\mu} = E(\boldsymbol{\Psi}(\mathbf{X}_1,\dots,\mathbf{X}_m))$ exists. The results of Theorem 7.1 of [6] (cf. also Theorem 2 on p. 76 of [10]) imply that the U-statistic (1.2) is asymptotically normal. More precisely, let for $c = 1,\dots,m$

$$\boldsymbol{\zeta}_c = E\Big(\boldsymbol{\Psi}_c(\mathbf{X}_1,\dots,\mathbf{X}_c)\boldsymbol{\Psi}_c(\mathbf{X}_1,\dots,\mathbf{X}_c)^\top\Big) - \boldsymbol{\mu}\boldsymbol{\mu}^\top, \tag{2.1}$$

where $\boldsymbol{\Psi}_c(\mathbf{x_1},\dots,\mathbf{x}_c) = E\Big(\boldsymbol{\Psi}(\mathbf{x_1},\dots,\mathbf{x}_c,\mathbf{X}_{c+1},\dots,\mathbf{X}_m)\Big)$ if the coordinates of this integral are real numbers, otherwise $\boldsymbol{\Psi}_c(\mathbf{x_1},\dots,\mathbf{x}_c) = \mathbf{0}$. Then

$$\sqrt{n}(\mathbf{U}_n - \boldsymbol{\mu}) \longrightarrow N_s(\mathbf{0}, \mathbf{V}_{\boldsymbol{\Psi}}) \tag{2.2}$$

in distribution as $n \to \infty$, where

$$\mathbf{V}_{\boldsymbol{\Psi}} = m^2 \boldsymbol{\zeta}_1. \tag{2.3}$$

Theorem 2.1. Let

$$C^*(m-1,n,i) = \Big\{r^* \in C(m-1,n);\ r^* = \{r_1^*,\dots,r_{m-1}^*\},\ i \notin \{r_1^*,\dots,r_{m-1}^*\}\Big\} \tag{2.4}$$

denote the collection of all subsets of size $m-1$ of elements from $\{1,\dots,n\}$ not containing the number i. Put

$$\boldsymbol{\Gamma}_i^{(n)} = \frac{1}{\binom{n-1}{m-1}} \sum_{r^* \in C(m-1,n,i)} \Psi(\mathbf{X}_i, \mathbf{X}_{r_1^*},\dots,\mathbf{X}_{r_{m-1}^*}), \quad \boldsymbol{\Gamma}_n = \frac{1}{n}\sum_{i=1}^{n} \boldsymbol{\Gamma}_i^{(n)} \boldsymbol{\Gamma}_i^{(n)\top}. \tag{2.5}$$

Then $\mathbf{\Gamma}_n = E(\mathbf{\Psi}_1(\mathbf{X}_1)\mathbf{\Psi}_1(\mathbf{X}_1)^\top) + o_P(1)$ and the matrix

$$\tilde{\mathbf{V}}_n = m^2(\mathbf{\Gamma}_n - \mathbf{U}_n\mathbf{U}_n^\top) \tag{2.6}$$

is a consistent estimate of the asymptotic covariance matrix $\mathbf{V}_{\mathbf{\Psi}}$.

P r o o f. This theorem is a straightforward consequence of the results of [14]. Indeed, fix an arbitrary $\mathbf{t} \in \mathbb{R}^s$ and note that $\mathbf{t}^\top\mathbf{U}_n$ is a univariate U-statistic with kernel $\mathbf{t}^\top\mathbf{\Psi} : \mathbb{R}^d \to \mathbb{R}$. By Proposition 1 in [14] , $\mathbf{t}^\top\tilde{\mathbf{V}}_n\mathbf{t}$ is a consistent estimator of the asymptotic variance $\mathbf{t}^\top\mathbf{V}_{\mathbf{\Psi}}\mathbf{t}$ of $\mathbf{t}^\top\mathbf{U}_n$. Hence by taking successively $\mathbf{t} = \mathbf{e}_i$ and $\mathbf{t} = \mathbf{e}_i - \mathbf{e}_j$, where $\mathbf{e}_i \in \mathbb{R}^s$ has in its ith coordinate 1 and 0 elsewhere, it is easy to see that $\tilde{\mathbf{V}}_n$ is a consistent estimate of $\mathbf{V}_{\mathbf{\Psi}}$, as claimed. □

Now consider the special case where $m = 2$ and $\mathbf{\Psi}(\mathbf{x}_i, \mathbf{x}_j) = \mathbf{0}$ whenever $\mathbf{x}_i = \mathbf{x}_j$. Making use of the transformation $\mathbf{U}_n \to \mathbf{t}^\top\mathbf{U}_n$ and the results of [6] one obtains that the covariance matrix

$$\operatorname{cov}(\sqrt{n}\mathbf{U}_n) = 4\frac{n-2}{n-1}\boldsymbol{\zeta}_1 + \frac{2}{n-1}\boldsymbol{\zeta}_2. \tag{2.7}$$

This suggests that for $n > 2$ an alternative consistent estimate of $\mathbf{V}_{\mathbf{\Psi}}$ is given by

$$\hat{\mathbf{V}}_n = 4\frac{n-2}{n-1}\hat{\boldsymbol{\zeta}}_1 + \frac{2}{n-1}\hat{\boldsymbol{\zeta}}_2, \tag{2.8}$$

$$\hat{\boldsymbol{\zeta}}_1 = \mathbf{\Gamma}_n - \mathbf{U}_n\mathbf{U}_n^\top, \quad \hat{\boldsymbol{\zeta}}_2 = \frac{1}{\binom{n}{2}} \sum_{r\in C(2,n)} \Big(\mathbf{\Psi}(\mathbf{X}_{r_1}, \mathbf{X}_{r_2}) - \mathbf{U}_n\Big)\Big(\mathbf{\Psi}(\mathbf{X}_{r_1}, \mathbf{X}_{r_2}) - \mathbf{U}_n\Big)^\top. \tag{2.9}$$

3. CONFIDENCE REGIONS FOR KENDALL'S COEFFICIENTS

Throughout this section we consider fixed, mutually distinct pairs of coordinate indexes $(\ell_1, u_1), \cdots, (\ell_s, u_s)$ of the d-dimensional random vector $\mathbf{X}$ (i. e., $|\ell_i - \ell_j| + |u_i - u_j| > 0$ whenever $i \neq j$) such that for $v = 1, \dots, s$

$$1 \le \ell_v < u_v \le d. \tag{3.1}$$

For $\mathbf{x} = (x_1, \dots, x_d)' \in \mathbb{R}^d$ let

$$\mathbf{x}(j) = \pi_j(\mathbf{x}) = x_j \tag{3.2}$$

denote the jth coordinate of the vector. Suppose that it is desired to construct a confidence region for the vector $(\tau(\ell_1, u_1), \dots, \tau(\ell_s, u_s))^\top$, where for each $v \in \{1, \dots, s\}$, $\tau(\ell_v, u_v)$ denotes the theoretical value of Kendall's tau in the pair $(\pi_{\ell_v}(\mathbf{X}), \pi_{u_v}(\mathbf{X}))$. Suppose that $\mathbf{x}_1, \dots, \mathbf{x}_n$ belong to $\mathbb{R}^d$ and ℓ, u are two integers such that $1 \le \ell < u \le d$. A standard consistent estimate of $\tau(\ell, u)$ is then given by

$$\begin{aligned} \tau_n(\ell, u) &= \tau_n(\ell, u, \mathbf{x}_1, \dots, \mathbf{x}_n) \\ &= \frac{2}{n(n-1)} \sum_{1\le i<j\le n} \operatorname{sign}\Big(\mathbf{x}_j(\ell) - \mathbf{x}_i(\ell)\Big)\operatorname{sign}\Big(\mathbf{x}_j(u) - \mathbf{x}_i(u)\Big), \end{aligned} \tag{3.3}$$

which is called the sample Kendall's correlation coefficient of the pair (ℓ, u). When $\mathbf{X}$ is a continuous random variable, the asymptotic distribution of $\tau_n(\ell, u)$ is known, as well as its finite sample behavior under the null hypothesis of independence between components $\pi_\ell(\mathbf{X})$ and $\pi_u(\mathbf{X})$ (cf. [5]).

In constructing a confidence region for the vector $\boldsymbol{\tau} = (\tau(\ell_1, u_1), \ldots, \tau(\ell_s, u_s))^\top$ we shall use the vector of sample coefficients

$$\boldsymbol{\tau}_n = \boldsymbol{\tau}_n(\mathbf{x}_1, \ldots, \mathbf{x}_n) = \boldsymbol{\tau}_n^{[(\ell_1,u_1),\ldots,(\ell_s,u_s)]}(\mathbf{x}_1, \ldots, \mathbf{x}_n) = \begin{pmatrix} \tau_n(\ell_1, u_1) \\ \vdots \\ \tau_n(\ell_s, u_s) \end{pmatrix}. \tag{3.4}$$

To use the results of the previous section define the mapping $\boldsymbol{\Psi} : \mathbb{R}^d \times \mathbb{R}^d \to \mathbb{R}^s$ by the formula

$$\begin{aligned} \boldsymbol{\Psi}(\mathbf{x}_1, \mathbf{x}_2) &= \boldsymbol{\Psi}(\mathbf{x}_1, \mathbf{x}_2)^{[(\ell_1,u_1),\ldots,(\ell_s,u_s)]} \\ &= \begin{pmatrix} \operatorname{sign}\Big(\mathbf{x}_2(\ell_1) - \mathbf{x}_1(\ell_1)\Big)\operatorname{sign}\Big(\mathbf{x}_2(u_1) - \mathbf{x}_1(u_1)\Big) \\ \vdots \\ \operatorname{sign}\Big(\mathbf{x}_2(\ell_s) - \mathbf{x}_1(\ell_s)\Big)\operatorname{sign}\Big(\mathbf{x}_2(u_s) - \mathbf{x}_1(u_s)\Big) \end{pmatrix}. \end{aligned} \tag{3.5}$$

Observe that

$$\boldsymbol{\tau} = E\Big(\boldsymbol{\Psi}(\mathbf{X}_1, \mathbf{X}_2)\Big) = \begin{pmatrix} \tau(\ell_1, u_1) \\ \vdots \\ \tau(\ell_s, u_s) \end{pmatrix} \tag{3.6}$$

and that

$$\boldsymbol{\tau}_n = \mathbf{U}_n = \mathbf{U}_n(\mathbf{x}_1, \ldots, \mathbf{x}_n) = \frac{2}{n(n-1)} \sum_{1 \le i < j \le n} \boldsymbol{\Psi}(\mathbf{x}_i, \mathbf{x}_j). \tag{3.7}$$

It follows from (3.7), (2.2) and (2.3) that the convergence in distribution

$$\sqrt{n}(\boldsymbol{\tau}_n - \boldsymbol{\tau}) \to N_s(\mathbf{0}, \mathbf{V}) \tag{3.8}$$

holds as $n \to \infty$. Here $\mathbf{V} = \mathbf{V}_{\boldsymbol{\Psi}}$ is the matrix defined by means of equality (2.3) and the mappping (3.5). This convergence was established in Theorem 2.3 of [12] where also a more detailed expression for the asymptotic covariance matrix $\mathbf{V}$ is presented.

Finally, if $0 < \alpha < 1$ and $\mathbf{A}$ is a p.s.d. symmetric $s \times s$ matrix then $cr_\alpha(\mathbf{A})$ will denote the critical value defined by the equality

$$P\Big(\max_{i=1,\ldots,s} |\mathbf{y}(i)| > cr_\alpha(\mathbf{A}) \Big| \mathbf{y} \sim N_s(\mathbf{0}, \mathbf{A})\Big) = \alpha. \tag{3.9}$$

The simplest way to obtain this constant appears to be through simulation.

Theorem 3.1. (I) Suppose $\mathbf{A}_n$ is a p.s.d. symmetric $s \times s$ matrix and (cf. (3.2))

$$I_n(\mathbf{A}_n) = \left\{ \mathbf{z} \in \mathbb{R}^s;\ \max_{j=1,\ldots,s} |\pi_j(\mathbf{z}) - \tau_n(\ell_j, u_j)| \le \frac{cr_\alpha(\mathbf{A}_n)}{\sqrt{n}} \right\} \tag{3.10}$$

is the corresponding s-dimensional rectangle. Let

$$F_j(t) = P(\pi_j(\mathbf{X}) \le t) \tag{3.11}$$

stand for the distribution function of the jth coordinate of the population. If $\mathbf{A}_n$ is a consistent estimate of the matrix $\mathbf{V}$ from (3.8) and there exist real numbers $t_1, \dots, t_d$ such that

$$0 < F_j(t_j) < 1, \quad j = 1, \dots, d, \tag{3.12}$$

then

$$\lim_{n\to\infty} P\Big(\boldsymbol{\tau} \in I_n(\mathbf{A}_n)\Big) = 1 - \alpha. \tag{3.13}$$

(II) Suppose that $\mathbf{X}$ posseses a density f with respect to the Lebesque measure μ_L on $\mathbb{R}^d$ and the function f is continuous on the set $\mathcal{N} = \{\mathbf{x} \in \mathbb{R}^d; f(\mathbf{x}) > 0\}$. If $\mathcal{N} = (a_1, b_1) \times \cdots \times (a_d, b_d)$, where $-\infty \le a_j < b_j \le +\infty$ for $j = 1, \dots, d$, then the matrix $\mathbf{V}$ is positive definite.

P r o o f. (I) Suppose that $\{n_t\}_{t=1}^{\infty}$ is an increasing sequence of positive integers. Since $\mathbf{A}_n$ is a consistent estimate of $\mathbf{V}$, there exists a subsequence $\{n_{t_w}\}_{w=1}^{\infty}$ such that

$$\mathbf{A}_{n_{t_w}} \to \mathbf{V}$$

almost surely. Thus almost surely $N_s(\mathbf{0}, \mathbf{A}_{n_{t_w}}) \to N_s(\mathbf{0}, \mathbf{V})$ in the sense of weak convergence of probability measures. Taking into account (3.5), (2.1), (2.3) and (3.12) one finds that the diagonal elements of the matrix $\mathbf{V}$ are positive and therefore the boundary of the s-dimensional interval $\langle c_1, h_1\rangle \times \cdots \times \langle c_s, h_s\rangle$ has zero mass under the $N_s(\mathbf{0}, \mathbf{V})$ distribution for every real numbers $c_v < h_v$, $v = 1, \dots, s$. This together with the mentioned weak convergence yields

$$P\Big(\langle -c, c\rangle^s \Big| N_s(\mathbf{0}, \mathbf{A}_{n_{t_w}})\Big) \to P\Big(\langle -c, c\rangle^s \Big| N_s(\mathbf{0}, \mathbf{V})\Big)$$

almost surely. Therefore $cr_\alpha(\mathbf{A}_{n_{t_w}}) \to cr_\alpha(\mathbf{V})$ almost surely and by the Egoroff theorem almost uniformly. Hence by (3.8)

$$\begin{aligned} \lim_{n\to\infty} P\Big(\sqrt{n} \max_{j=1,\dots,s} |\tau_{n_{t_w}}(\ell_j, u_j) - \tau(\ell_j, u_j)| > cr_\alpha(\mathbf{A}_{n_{t_w}})\Big) \\ = P\Big(\max_{j=1,\dots,s} |\mathbf{y}(j)| > cr_\alpha(\mathbf{V}) \Big| N_s(\mathbf{0}, \mathbf{V})\Big) = \alpha \end{aligned}$$

and the convergence (3.13) is proved.

(II) Fix an integer $j \in \{1, \dots, d\}$. Let z be a real number and

$$K_z = \{(z_1, \dots, z_{j-1}, z_{j+1}, \dots, z_d)^\top; (z_1, \dots, z_{j-1}, z, z_{j+1}, \dots, z_d)^\top \in \mathcal{N}\}.$$

By Fubini's theorem the density of the distribution function (3.11) is

$$f_j(z) = \int_{K_z} f(z_1, \dots, z_{j-1}, z, z_{j+1}, \dots, z_d)\, \mathrm{d}z_1 \dots \mathrm{d}z_{j-1} \mathrm{d}z_{j+1} \dots \mathrm{d}z_d \tag{3.14}$$

if this integral is a real number and $f_j(z) = 0$ otherwise, and the second possibility occurs with one-dimensional Lebesgue measure zero. Fix numbers $c_j < g_j \in (a_j, b_j)$. Then for each number $x \in \langle c_j, g_j \rangle$

$$F_j(x) = F_j(c_j) + \int_{c_j}^{x} f_j(t)\,\mathrm{d}t$$

and according to Theorem 9.3 from [9] for almost all $x \in \langle c_j, g_j \rangle$

$$F_j'(x) = f_j(x). \tag{3.15}$$

Therefore this equality holds on (a_j, b_j) a.e. with respect to the one-dimensional Lebesgue measure and obviously f_j is positive on (a_j, b_j) a.e. with respect to the Lebesgue measure.

Since the distribution function of the random vector $(\pi_{\ell_v}(\mathbf{X}), \pi_{u_v}(\mathbf{X}))$ and its marginal distribution functions are continuous, for $\mathbf{x} \in \mathbb{R}^d$

$$\begin{aligned}\pi_v(\mathbf{\Psi}_1(\mathbf{x})) &= \int \operatorname{sign}\Big(\pi_{\ell_v}(\mathbf{X}_2) - \pi_{\ell_v}(\mathbf{x})\Big)\operatorname{sign}\Big(\pi_{u_v}(\mathbf{X}_2) - \pi_{u_v}(\mathbf{x})\Big)\,\mathrm{d}P^{\mathbf{X}_2} \\ &= 2\Big(P(\pi_{\ell_v}(\mathbf{X}_2) < \pi_{\ell_v}(\mathbf{x}), \pi_{u_v}(\mathbf{X}_2) < \pi_{u_v}(\mathbf{x})) \\ &\quad + P(\pi_{\ell_v}(\mathbf{X}_2) > \pi_{\ell_v}(\mathbf{x}), \pi_{u_v}(\mathbf{X}_2) > \pi_{u_v}(\mathbf{x}))\Big) - 1\end{aligned} \tag{3.16}$$

and the mapping $\mathbf{\Psi}_1$ is continuous on $\mathcal{N}$.

Let $\mathbf{c} = (c_1, \dots, c_s)^\top \in \mathbb{R}^s$ be a vector such that $\mathbf{c}^\top \boldsymbol{\zeta}_1 \mathbf{c} = 0$. Then with the probability $P^{\mathbf{X}_1} = 1$

$$\mathbf{c}^\top \mathbf{\Psi}_1(\mathbf{x}) = \sum_{j=1}^{s} c_j \pi_j(\mathbf{\Psi}_1(\mathbf{x})) = \mathbf{c}^\top \boldsymbol{\tau}. \tag{3.17}$$

The assumptions imply that each point from $\mathcal{N}$ is a limit of points from $\mathcal{N}$ for which (3.17) holds and therefore (3.17) holds for each point from $\mathcal{N}$.

Without loss of generality assume that $\ell_1 \le \ell_2 \le \ldots \le \ell_s$. Let $\ell_1 = \ldots = \ell_{k_1} < \ell_{k_1+1}$. Fix a number $x \in (a_{\ell_1}, b_{\ell_1})$ and assume that $\mathbf{x} \in \mathcal{N}$ is such that $\pi_{\ell_1}(\mathbf{x}) = x$. Since (3.16) holds, for each $v = 1, \dots, k_1$

$$\lim_{\pi_{u_v}(\mathbf{x}) \nearrow b_{u_v}} \pi_v(\mathbf{\Psi}_1(\mathbf{x})) = 2P(\pi_{\ell_1}(\mathbf{X}_2) < \pi_{\ell_1}(\mathbf{x})) - 1,$$

$$\lim_{\pi_{u_v}(\mathbf{x}) \searrow a_{u_v}} \pi_v(\mathbf{\Psi}_1(\mathbf{x})) = 2P(\pi_{\ell_1}(\mathbf{X}_2) > \pi_{\ell_1}(\mathbf{x})) - 1.$$

As it has been shown, there exists a number $x \in (a_{\ell_1}, b_{\ell_1})$ such that both (3.15) and the inequality $f_j(x) > 0$ hold for $j = \ell_1$. Thus letting $\pi_{u_v}(\mathbf{x})$ go to b_{u_v}, $v = 1, \dots, k_1$ and differentiating the limit of (3.17) with respect to $\pi_{\ell_1}(\mathbf{x})$ at $\pi_{\ell_1}(\mathbf{x}) = x$ we obtain that

$$0 = \sum_{j=1}^{k_1} c_j 2 f_{\ell_1}(x), \qquad \sum_{j=1}^{k_1} c_j = 0.$$

Similarly, fixing $i \in \{1, \dots, k_1\}$, letting $\pi_{u_i}(\mathbf{x})$ go to a_{u_i}, $\pi_{u_v}(\mathbf{x})$ go to b_{u_v}, $v = 1, \dots, k_1$, $v \neq i$, and differentiating the limit of of (3.17) with respect to $\pi_{\ell_1}(\mathbf{x})$ at $\pi_{\ell_1}(\mathbf{x}) = x$ we obtain that

$$0 = \sum_{j=1, j\neq i}^{k_1} c_j 2 f_{\ell_1}(x) - c_i 2 f_{\ell_1}(x), \qquad c_i = \sum_{j=1, j\neq i}^{k_1} c_j.$$

Hence $c_i = -c_i = 0$ for $i = 1, \dots, k_1$, and therefore

$$\mathbf{c}^\top \boldsymbol{\tau} = \sum_{j=k_1+1}^{s} c_j \pi_j(\boldsymbol{\Psi}_1(\mathbf{x})).$$

Repeating this procedure yields $c_1 = \ldots = c_s = 0$. This means that the matrix $\boldsymbol{\zeta}_1$ is regular. □

Classical confidence sets for parameters are ellipsoids formed by means of quantiles of the chi-square distribution, but in this case such an approach would result in the situation with slower convergence of coverage probability to $1 - \alpha$. Nowadays, with computers at our disposal, the use of critical constants as in (3.9) is not as complicated a problem as it was in the past. We remark that the constants of the type (3.9) were used for statistical inference in [11].

The advantage of the multivariate confidence region (3.10) for $\boldsymbol{\tau}$ from (3.6) is that it yields the multiple comparisons rule for the null hypothesis that for each $v \in \{1, \dots, s\}$ the random variables $\pi_{\ell_v}(\mathbf{X})$ and $\pi_{u_v}(\mathbf{X})$ are independent. Indeed, although two random variables need not be independent if their Kendall's correlation coefficient is zero, in the case when this coefficient is different from zero they are not independent. Therefore if the interval $(\tau_n(\ell_v, u_v) - \frac{c r_\alpha(\mathbf{A}_n)}{\sqrt{n}}, \tau_n(\ell_v, u_v) + \frac{c r_\alpha(\mathbf{A}_n)}{\sqrt{n}})$ does not contain zero, then the random variables $\pi_{\ell_v}(\mathbf{X})$, $\pi_{u_v}(\mathbf{X})$ are declared to be dependent. Moreover, because of the dimensionality s of the parameter of interest, the multivariate products of intervals are easier to interpret than the s-dimensional ellipsoids, because in contradistinction to the confidence ellipsoids, the limits of the particular coordinate need not be computed by means of remaining coordinates.

The estimators $\tilde{\mathbf{V}}_n$, $\hat{\mathbf{V}}_n$ defined by means of (2.6), (2.8) and (3.5) are asymptotically equivalent, but (as simulations show), the estimator $\tilde{\mathbf{V}}_n$ generally yields shorter multivariate intervals and $\hat{\mathbf{V}}_n$ generally yields larger probability of coverage.

Now we are going to deal with confidence interval for a linear combination of Kendall's population coefficients.

Corollary 3.1. Let $\mathbf{c} \in \mathbb{R}^s$ be a non-zero vector. Then $\sqrt{n}(\mathbf{c}^\top \boldsymbol{\tau}_n - \mathbf{c}^\top \boldsymbol{\tau}) \to N(0, \sigma^2)$ in distribution, where $\sigma^2 = \mathbf{c}^\top \mathbf{V} \mathbf{c}$ and $\mathbf{V}$ is the matrix appearing in (3.8). Therefore for any consistent estimate $\mathbf{A}_n = \mathbf{A}_n(\mathbf{X}_1, \dots, \mathbf{X}_n)$ of the matrix $\mathbf{V}$ the statistic $\sigma_n^2 = \mathbf{c}^\top \mathbf{A}_n \mathbf{c}$ is a consistent estimate of σ^2 and if $\sigma^2 > 0$ (which is true under the assumptions of Theorem 3.1(II)), then

$$\lim_{n\to\infty} P\Big(\mathbf{c}^\top \boldsymbol{\tau} \in (\mathbf{c}^\top \boldsymbol{\tau}_n - g_{1-\alpha/2} \frac{\sigma_n}{\sqrt{n}}, \mathbf{c}^\top \boldsymbol{\tau}_n + g_{1-\alpha/2} \frac{\sigma_n}{\sqrt{n}})\Big) = 1 - \alpha. \tag{3.18}$$

Here $g_{1-\alpha/2}$ denotes the $1 - \alpha/2$ quantile of the standard $N(0, 1)$ distribution.

P r o o f. This corollary is an immediate consequence of the convergence (3.8) and of the regularity of the matrix $\mathbf{V}$ following from Theorem 3.1. □

The dependence of the components of the d-dimensional random vector $\mathbf{X}$ can be characterized by the coefficient of agreement

$$T_d = T_d(\mathbf{X}) = \frac{1}{d(d-1)} \sum_{i \neq j} \tau(i,j) = \frac{2}{d(d-1)} \sum_{1 \leq i < j \leq d} \tau(i,j) \tag{3.19}$$

which was introduced in [8]. It is estimated by its sample counterpart

$$T_{d,n} = \frac{2}{d(d-1)} \sum_{1 \leq i < j \leq d} \tau_n(i,j) \tag{3.20}$$

which was studied in [2] and [3]. To describe the variation and the limiting distribution of the statistic (3.20) assume that the coordinates (3.1) are those appearing above the diagonal of the matrix $((i,j))_{i,j=1}^d$. In accordance with this assumption and (3.5) let

$$\mathbf{\Psi}(\mathbf{x}_1, \mathbf{x}_2) = \mathbf{\Psi}(\mathbf{x}_1, \mathbf{x}_2)^{[(1,2),(1,3)\ldots,(1,d),(2,3),\ldots,(2,d),\ldots,(d-1,d)]}. \tag{3.21}$$

In what follows for any matrix $\mathbf{M}$ by $\mathbf{M}(i,j)$ we mean the element of $\mathbf{M}$ on the position (i,j).

Corollary 3.2. Let $\boldsymbol{\zeta}_1$, $\boldsymbol{\zeta}_2$ be defined by means of (3.21) and (2.1) (where the mean $\boldsymbol{\mu} = E(\mathbf{\Psi}(\mathbf{X}_1, \mathbf{X}_2)) = E(\mathbf{\Psi}(\mathbf{X}_1, \mathbf{X}_2)^{[(1,2),(1,3)\ldots,(1,d),(2,3),\ldots,(2,d),\ldots,(d-1,d)]})$ is the vector of all population Kendall's coefficients).

(I) The equality

$$\begin{gathered} \operatorname{var}(\sqrt{n}T_{d,n}) = \frac{1}{s^2} \sum_{i=1}^{s} \sum_{j=1}^{s} \mathbf{V}_n(i,j), \quad \mathbf{V}_n = 4\frac{n-2}{n-1}\boldsymbol{\zeta}_1 + \frac{2}{n-1}\boldsymbol{\zeta}_2, \\ s = \frac{(d-1)d}{2}, \end{gathered} \tag{3.22}$$

holds and

$$\sqrt{n}(T_{d,n} - T_d) \to N(0, \sigma^2) \text{ in distribution}, \quad \sigma^2 = \frac{1}{s^2} \sum_{i=1}^{s} \sum_{j=1}^{s} \mathbf{V}(i,j), \tag{3.23}$$

where $\mathbf{V} = 4\boldsymbol{\zeta}_1$.

(II) Suppose that the number σ^2 in (3.23) is positive (which is true under the assumptions of Theorem 3.1(II)). If $\mathbf{A}_n = \mathbf{A}_n(\mathbf{X}_1, \ldots . \mathbf{X}_n)$ is a consistent estimate of the matrix $\mathbf{V}$, then $\sigma_n^2 = (1/s^2) \sum_{i=1}^{s} \sum_{j=1}^{s} \mathbf{A}_n(i,j)$ is a consistent estimate of σ^2 and

$$\begin{gathered} \lim_{n \to \infty} P\Big(T_d \in I^{T_{d,n}}(\mathbf{A}_n)\Big) = 1 - \alpha, \\ I^{T_{d,n}}(\mathbf{A}_n) = (T_{d,n} - g_{1-\alpha/2}\frac{\sigma_n}{\sqrt{n}}, T_{d,n} + g_{1-\alpha/2}\frac{\sigma_n}{\sqrt{n}}). \end{gathered} \tag{3.24}$$

In this framework, if $\mathbf{A}_n = \tilde{\mathbf{V}}_n$ denotes the matrix defined by means of (3.21), (3.5) and (2.6) or $\mathbf{A}_n = \hat{\mathbf{V}}_n$ denotes the matrix defined by means of (3.21), (3.5) and (2.8), (2.9), then (3.24) holds.

Proof. Obviously

$$T_{d,n} = \frac{2}{d(d-1)} \mathbf{1}^\top \boldsymbol{\tau}_n \tag{3.25}$$

where $\boldsymbol{\tau}_n = \boldsymbol{\tau}_n^{[(1,2),(1,3)\dots,(1,d),(2,3),\dots,(2,d),\dots,(d-1,d)]}$ are the sample Kendall's coefficients (3.4) and $\mathbf{1} \in \mathbb{R}^{d(d-1)/2}$ is the column vector whose coordinates equal 1. Hence (3.22) follows from (3.7) and (2.7). Similarly (3.23), (3.24) follow from Corollary 3.1. Since according to the previous Section the matrices $\tilde{\mathbf{V}}_n$ and $\hat{\mathbf{V}}_n$ are consistent estimators of the matrix $\mathbf{V}$, the Corollary is proved. □

In contrast to the approach of this paper, formulas for variance, asymptotic variance and the asymptotic normality of the sample coefficient of agreement (3.20) are derived in [3] in terms of the underlying copula models, but neither confidence intervals are mentioned nor the positivity of the asymptotic variance is mentioned there. We remark that the conditions imposed in Theorem 3.1(II) on the density of the multivariate distributions are fulfilled perhaps by every continuous multivariate distribution used in statistical inference.

The data used in the next example are taken from DASL (The Data and Story Library), Carnegie Mellon University, Story Names: *Air Pollution and Mortality.*

Example. The following data were recorded in 59 metropolitan areas of the USA. The monitored variables are: v1 – mean of January temperature (Fahrenheit), v2 – mean of July temperature (Fahrenheit), v3 – relative humidity, v4 – annual rainfall (inches), v5 – age adjusted mortality, v6 – median education, v7 – population density, v8 – percentage of non whites, v9 – percentage of white collar workers, v10 – population, v11 – population per household, v12 – median income, v13 – HC pollution potential, v14 – Nitrous Oxide pollution potential, v15 – Sulfur Dioxide pollution potential. In difference from the 60 areas in the original data, the observations from Forth Worth are here omitted because for this area 2 items are missing. Therefore we consider 59 areas. The data for the 59 areas can be found in the public domain http://ugrad.stat.ubc.ca/~stat447j/datasets/mortality.txt.

Our task is to compute from these observations the 95% confidence region for Kendall's correlation coefficients of age adjusted mortality and the rest of the variables, i. e., the 95% confidence region for the vector

$$\boldsymbol{\tau} = (\tau(5,j);\, j = 1,\dots,15,\, j \neq 5)^\top \tag{3.26}$$

by means of the product of intervals $I_n(\tilde{\mathbf{V}}_n)$ from (3.10) based on (2.6), and to compute for the coefficient of agreement the 95% confidence interval $I^{T_{d,n}}(\tilde{\mathbf{V}}_n)$ from (3.24) based on (2.6).

The coordinates (3.1) are in this case the pairs $(5,j)$, $j = 1,\dots,15$, $j \neq 5$. For these data the dimension of the vector of Kendall's coefficients $s = 14$, the estimate (3.4) of the vector (3.26) and the matrix $\tilde{\mathbf{V}}_n$ (rounded for typographical reasons to 3 decimal places) are

$$\boldsymbol{\tau}_n = (0.057,\, 0.262,\, -0.093,\, 0.265,\, -0.369,\, 0.177,\, 0.429,\, -0.233,\, 0.064,\, 0.254,\, -0.157,\, 0.226,\, 0.297,\, 0.340)^\top,$$

$$\tilde{\mathbf{V}}_n = \begin{pmatrix}
0.651 & 0.183 & -0.034 & 0.116 & 0.045 & -0.094 & 0.236 & 0.134 & 0.184 & 0.040 & 0.056 & 0.204 & 0.193 & -0.077 \\
0.183 & 0.514 & -0.213 & 0.201 & -0.024 & -0.037 & 0.219 & 0.046 & -0.039 & 0.047 & -0.011 & -0.254 & -0.225 & -0.149 \\
-0.034 & -0.213 & 0.608 & -0.112 & 0.045 & -0.069 & -0.068 & 0.006 & 0.063 & -0.039 & 0.010 & 0.144 & 0.107 & -0.142 \\
0.116 & 0.201 & -0.112 & 0.583 & -0.221 & -0.068 & 0.128 & -0.168 & -0.160 & 0.099 & -0.215 & -0.226 & -0.210 & -0.111 \\
0.045 & -0.024 & 0.045 & -0.221 & 0.387 & 0.003 & -0.026 & 0.235 & 0.159 & -0.151 & 0.220 & 0.095 & 0.067 & -0.012 \\
-0.094 & -0.037 & -0.069 & -0.068 & 0.003 & 0.467 & -0.023 & 0.004 & 0.097 & -0.203 & 0.081 & 0.016 & 0.052 & 0.221 \\
0.236 & 0.219 & -0.068 & 0.128 & -0.026 & -0.023 & 0.299 & 0.017 & 0.065 & 0.039 & -0.009 & -0.030 & -0.023 & -0.050 \\
0.134 & 0.046 & 0.006 & -0.168 & 0.235 & 0.004 & 0.017 & 0.343 & 0.192 & -0.117 & 0.235 & 0.104 & 0.098 & -0.036 \\
0.184 & -0.039 & 0.063 & -0.160 & 0.159 & 0.097 & 0.065 & 0.192 & 0.476 & -0.139 & 0.230 & 0.239 & 0.236 & 0.106 \\
0.040 & 0.047 & -0.039 & 0.099 & -0.151 & -0.203 & 0.039 & -0.117 & -0.139 & 0.419 & -0.132 & -0.049 & -0.042 & -0.105 \\
0.056 & -0.011 & 0.010 & -0.215 & 0.220 & 0.081 & -0.009 & 0.235 & 0.230 & -0.132 & 0.416 & 0.143 & 0.123 & 0.024 \\
0.204 & -0.254 & 0.144 & -0.226 & 0.095 & 0.016 & -0.030 & 0.104 & 0.239 & -0.049 & 0.143 & 0.538 & 0.486 & 0.211 \\
0.193 & -0.225 & 0.107 & -0.210 & 0.067 & 0.052 & -0.023 & 0.098 & 0.236 & -0.042 & 0.123 & 0.486 & 0.493 & 0.227 \\
-0.077 & -0.149 & -0.142 & -0.111 & -0.012 & 0.221 & -0.050 & -0.036 & 0.106 & -0.105 & 0.024 & 0.211 & 0.227 & 0.396
\end{pmatrix}.$$

By means of a random sample of size 5000 from the normal $N_{14}(\mathbf{0}, \tilde{\mathbf{V}}_n)$ distribution we obtain the simulation estimate of the constant from (3.9)

$$cr_{0.05}(\tilde{\mathbf{V}}_n) = 2.0172.$$

From these quantities the confidence region $I_n(\tilde{\mathbf{V}}_n)$ from (3.10) based on (2.6) equals

$$\begin{array}{lll}
\tau(5,1) \in (-0.205, 0.320), & \tau(5,2) \in (-0.001, 0.524), & \tau(5,3) \in (-0.356, 0.170), \\
\tau(5,4) \in (0.003, 0.528), & \tau(5,6) \in (-0.631, -0.106), & \tau(5,7) \in (-0.086, 0.440), \\
\tau(5,8) \in (0.166, 0.692), & \tau(5,9) \in (-0.495, 0.030), & \tau(5,10) \in (-0.199, 0.326), \\
\tau(5,11) \in (-0.009, 0.516), & \tau(5,12) \in (-0.420, 0.105), & \tau(5,13) \in (-0.037, 0.488), \\
\tau(5,14) \in (0.0343, 0.560), & \tau(5,15) \in (0.077, 0.602). &
\end{array}$$

Thus the multiple comparisons rule detects $(5,4)$, $(5,6)$, $(5,8)$, $(5,14)$, $(5,15)$ as the pairs of indexes of dependent coordinate random variables, because the corresponding subintervals of the multivariate confidence region do not contain zero.

Since the computation of the interval $I^{T_{d,n}}(\tilde{\mathbf{V}}_n)$ requires the computation of all $15 \times 14/2 = 105$ Kendall's correlation coefficients $\tau(i,j)$, $1 \le i < j \le 15$ and the computation of the 105×105 matrix $\tilde{\mathbf{V}}_n$, we present only the final numerical results. For these data

$$T_{d,n} = 0.0764, \quad \sigma_n = 0.1295$$

and therefore an asymptotic 95% confidence interval for the coefficient agreement is

$$T_d \in (0.043, 0.110).$$

4. SOME SIMULATION RESULTS ON CONFIDENCE REGIONS

All simulations estimates presented in this section are carried out by means of MATLAB. They are based on $N = 5000$ trials and the nominal probability of coverage is $1-\alpha = 0.95$ in each case.

Let $\boldsymbol{\Psi}$ be the mapping (3.5). In the following tables simulation estimates of the probability of the coverage $P(\boldsymbol{\tau} \in I_n(\mathbf{A}_n))$ of the population Kendall's correlation coefficient (3.6) by the multivariate rectangle (3.10) are presented. The tables include also simulation estimates of the mean of the length of this multivariate interval and simulation estimates concerning the coefficient of agreement. By the length of (3.10) we mean the number $le(I_n(\mathbf{A}_n)) = 2\frac{cr_\alpha(\mathbf{A})}{\sqrt{n}}$ and by the length of $I^{T_{d,n}}(\mathbf{A}_n)$ from (3.24) we mean the number $le(I^{T_{d,n}}(\mathbf{A}_n)) = 2g_{1-\alpha/2}\sigma_n/\sqrt{n}$. The symbol $\sigma(\xi)$ denotes the simulation estimate of the standard deviation of the random variable ξ.

In accordance with the previous text, $\tilde{\mathbf{V}}_n$ denotes the matrix defined by means of (3.5) and (2.6), $\hat{\mathbf{V}}_n$ denotes the matrix defined by means of (3.5) and (2.8), (2.9).

An explicit formula for the value of Kendall's tau is available for a special type of distributions, known as meta-elliptical distributions (studied in [7], [1] and in the papers quoted ibidem). Suppose that $\mathbf{W}$ is a symmetric positive definite $d \times d$ matrix and $\mathbf{X}$ has the normal $N_d(\mathbf{0}, \mathbf{W})$ distribution or the multivariate $t_d(\mathbf{0}, \mathbf{W}, \nu)$ distribution with ν degrees of freedom (having the density $f(\mathbf{x}) = \det(\mathbf{W})^{-1/2}K(d,\nu)(1 + \mathbf{x}^\top\mathbf{W}^{-1}\mathbf{x}/\nu)^{-(\nu+d)/2}$, which includes the multivariate Cauchy distributions determined with this density when ν=1). Since the Kendall's coefficient is invariant under monotone increasing transformations of the coordinate random variables, it follows from the properties of the meta-elliptical distributions that the population Kendall's coefficient of $\mathbf{X}$

$$\tau(i,j) = \frac{2}{\pi}\arcsin\Big(\frac{\mathbf{W}(i,j)}{\sqrt{\mathbf{W}(i,i)\mathbf{W}(j,j)}}\Big), \qquad i,j = 1,\dots,d, \tag{4.1}$$

where $\mathbf{W}(i,j)$ denotes the element of $\mathbf{W}$ on the position (i,j).

For $r \in (0,1)$ let $\mathbf{\Sigma}_d(r)$ denote the $d \times d$ matrix having on the position (i,j) the number 1 if $i = j$ and r otherwise, i. e., $\mathbf{\Sigma}_d(r)$ corresponds to the equicorrelated case. Note that according to Section 1f, Exercise 1.1 of [13] this matrix is positive definite provided that $-1/(d-1) < r < 1$.

Now we assume that $d = 7$ and

$$\begin{pmatrix} \ell_1 & \ell_2 & \dots & \ell_s \\ u_1 & u_2 & \dots & u_s \end{pmatrix} = \begin{pmatrix} 1 & 1 & 1 & 1 & 2 & 2 & 2 & 2 & 3 & 3 & 3 & 3 \\ 4 & 5 & 6 & 7 & 4 & 5 & 6 & 7 & 4 & 5 & 6 & 7 \end{pmatrix},$$

i. e., the dimension of the vector of Kendall's correlation coefficients (3.4) is $s = 12$.

Further, let

$$\mathbf{W} = \mathbf{K}\mathbf{K}^\top, \quad \mathbf{K} = \begin{pmatrix} 1 & 0 & 0 & 0 & 0 & 0 & 0 \\ r & 1 & 0 & 0 & 0 & 0 & 0 \\ r^2 & r & 1 & 0 & 0 & 0 & 0 \\ r^3 & r^2 & r & 1 & 0 & 0 & 0 \\ r^4 & r^3 & r^2 & r & 1 & 0 & 0 \\ r^5 & r^4 & r^3 & r^2 & r & 1 & 0 \\ r^6 & r^5 & r^4 & r^3 & r^2 & r & 1 \end{pmatrix}, \quad r = 0.7,$$

i. e., $\mathbf{W}_7$ is an autoregressive matrix of order 1. For this matrix the Kendall coefficients (4.1) in $\boldsymbol{\tau} = (\tau(1,4), \tau(1,5), \tau(1,6), \tau(1,7), \tau(2,4), \tau(2,5), \tau(2,6), \tau(2,7), \tau(3,4), \tau(3,5), \tau(3,6), \tau(3,7))'$ are

$$\begin{array}{llll} \tau(1,4) = 0.1624 & \tau(1,5) = 0.1113 & \tau(1,6) = 0.0771 & \tau(1,7) = 0.0537 \\ \tau(2,4) = 0.2901 & \tau(2,5) = 0.1962 & \tau(2,6) = 0.1352 & \tau(2,7) = 0.0939 \\ \tau(3,4) = 0.4737 & \tau(3,5) = 0.3093 & \tau(3,6) = 0.2104 & \tau(3,7) = 0.1453 \end{array}$$

and the coefficient of agreement $T_d = 0.2696$.

r	-0.16	0	0.40	0.80	0.99
$P(\boldsymbol{\tau} \in I_n(\tilde{\mathbf{V}}_n))$	0.957	0.960	0.959	0.969	0.989
$E(\,le(I_n(\tilde{\mathbf{V}}_n))\,)$	0.631	0.639	0.595	0.418	0.144
$\sigma(\,le(I_n(\tilde{\mathbf{V}}_n))\,)$	0.019	0.018	0.029	0.050	0.024
$P(\,\boldsymbol{\tau} \in I_n(\hat{\mathbf{V}}_n)\,)$	0.968	0.973	0.972	0.980	0.996
$E(\,le(I_n(\hat{\mathbf{V}}_n))\,)$	0.654	0.662	0.618	0.443	0.164
$\sigma(\,le(I_n(\hat{\mathbf{V}}_n))\,)$	0.018	0.017	0.028	0.048	0.024
$P(T_d \in I^{T_{d,n}}(\tilde{\mathbf{V}}_n)\,)$	0.982	0.927	0.937	0.934	0.954
$E(\,le(I^{T_{d,n}}(\tilde{\mathbf{V}}_n))\,)$	0.019	0.092	0.200	0.186	0.066
$\sigma(\,le(I^{T_{d,n}}(\tilde{\mathbf{V}}_n))\,)$	0.003	0.020	0.024	0.032	0.016
$P(T_d \in I^{T_{d,n}}(\hat{\mathbf{V}}_n)\,)$	0.993	0.938	0.941	0.947	0.982
$E(\,le(I^{T_{d,n}}(\hat{\mathbf{V}}_n))\,)$	0.023	0.096	0.205	0.195	0.076
$\sigma(\,le(I^{T_{d,n}}(\hat{\mathbf{V}}_n))\,)$	0.003	0.020	0.023	0.030	0.015

Tab. 1. Sampling from multivariate normal $N_7(\mathbf{0}, \mathbf{\Sigma}_7(r))$ distribution, $n = 40$.

r	-0.16	0	0.40	0.80	0.99
$P(\boldsymbol{\tau} \in I_n(\tilde{\mathbf{V}}_n))$	0.949	0.946	0.953	0.962	0.956
$E(\,le(I_n(\tilde{\mathbf{V}}_n))\,)$	0.788	0.795	0.756	0.583	0.212
$\sigma(\,le(I_n(\tilde{\mathbf{V}}_n))\,)$	0.029	0.028	0.037	0.070	0.084
$P(\,\boldsymbol{\tau} \in I_n(\hat{\mathbf{V}}_n)\,)$	0.957	0.955	0.960	0.970	0.979
$E(\,le(I_n(\hat{\mathbf{V}}_n))\,)$	0.803	0.810	0.771	0.598	0.225
$\sigma(\,le(I_n(\hat{\mathbf{V}}_n))\,)$	0.027	0.027	0.036	0.069	0.082
$P(T_d \in I^{T_{d,n}}(\tilde{\mathbf{V}}_n)\,)$	0.905	0.895	0.927	0.932	0.886
$E(\,le(I^{T_{d,n}}(\tilde{\mathbf{V}}_n))\,)$	0.033	0.114	0.241	0.250	0.097
$\sigma(\,le(I^{T_{d,n}}(\tilde{\mathbf{V}}_n))\,)$	0.012	0.035	0.032	0.030	0.039
$P(T_d \in I^{T_{d,n}}(\hat{\mathbf{V}}_n)\,)$	0.926	0.900	0.931	0.937	0.912
$E(\,le(I^{T_{d,n}}(\hat{\mathbf{V}}_n))\,)$	0.035	0.116	0.244	0.255	0.103
$\sigma(\,le(I^{T_{d,n}}(\hat{\mathbf{V}}_n))\,)$	0.011	0.034	0.031	0.029	0.037

Tab. 2. Sampling from multivariate Cauchy $t_7(\mathbf{0}, \mathbf{\Sigma}_7(r), 1)$ distribution, $n = 40$.

The simulation estimates suggest that for both methods (based either on $\tilde{\mathbf{V}}_n$ or on $\hat{\mathbf{V}}_n$) the probability of the coverage is approximately the same but the use of $\tilde{\mathbf{V}}_n$ yields better length.

n	30	40	50	60	80	100
$P(\boldsymbol{\tau} \in I_n(\tilde{\mathbf{V}}_n))$	0.960	0.954	0.956	0.951	0.947	0.948
$E(\,le(I_n(\tilde{\mathbf{V}}_n))\,)$	0.708	0.601	0.531	0.480	0.412	0.366
$\sigma(\,le(I_n(\tilde{\mathbf{V}}_n))\,)$	0.039	0.029	0.023	0.019	0.015	0.012
$P(\,\boldsymbol{\tau} \in I_n(\hat{\mathbf{V}}_n)\,)$	0.974	0.966	0.967	0.961	0.954	0.955
$E(\,le(I_n(\hat{\mathbf{V}}_n))\,)$	0.743	0.625	0.548	0.494	0.420	0.372
$\sigma(\,le(I_n(\hat{\mathbf{V}}_n))\,)$	0.037	0.028	0.023	0.019	0.014	0.012
$P(T_d \in I^{T_{d,n}}(\tilde{\mathbf{V}}_n)\,)$	0.932	0.937	0.937	0.938	0.940	0.940
$E(\,le(I^{T_{d,n}}(\tilde{\mathbf{V}}_n))\,)$	0.234	0.202	0.180	0.164	0.142	0.127
$\sigma(\,le(I^{T_{d,n}}(\tilde{\mathbf{V}}_n))\,)$	0.032	0.024	0.019	0.015	0.011	0.009
$P(T_d \in I^{T_{d,n}}(\hat{\mathbf{V}}_n)\,)$	0.942	0.944	0.942	0.942	0.943	0.943
$E(\,le(I^{T_{d,n}}(\hat{\mathbf{V}}_n))\,)$	0.243	0.207	0.184	0.167	0.144	0.128
$\sigma(\,le(I^{T_{d,n}}(\hat{\mathbf{V}}_n))\,)$	0.030	0.023	0.018	0.015	0.011	0.009

Tab. 3. Sampling from the multivariate normal $N_7(\mathbf{0}, \mathbf{W})$ distribution.

n	30	40	50	60	80	100
$P(\boldsymbol{\tau} \in I_n(\tilde{\mathbf{V}}_n))$	0.948	0.948	0.949	0.947	0.945	0.953
$E(\,le(I_n(\tilde{\mathbf{V}}_n))\,)$	0.879	0.759	0.680	0.620	0.537	0.480
$\sigma(\,le(I_n(\tilde{\mathbf{V}}_n))\,)$	0.045	0.035	0.028	0.024	0.018	0.015
$P(\,\boldsymbol{\tau} \in I_n(\hat{\mathbf{V}}_n)\,)$	0.958	0.957	0.955	0.954	0.950	0.957
$E(\,le(I_n(\hat{\mathbf{V}}_n))\,)$	0.902	0.775	0.691	0.629	0.543	0.484
$\sigma(\,le(I_n(\hat{\mathbf{V}}_n))\,)$	0.044	0.034	0.027	0.023	0.018	0.015
$P(T_d \in I^{T_{d,n}}(\tilde{\mathbf{V}}_n)\,)$	0.919	0.929	0.929	0.939	0.941	0.938
$E(\,le(I^{T_{d,n}}(\tilde{\mathbf{V}}_n))\,)$	0.280	0.244	0.219	0.201	0.174	0.156
$\sigma(\,le(I^{T_{d,n}}(\tilde{\mathbf{V}}_n))\,)$	0.043	0.031	0.025	0.020	0.015	0.013
$P(T_d \in I^{T_{d,n}}(\tilde{\mathbf{V}}_n)\,)$	0.924	0.934	0.934	0.940	0.943	0.940
$E(\,le(I^{T_{d,n}}(\tilde{\mathbf{V}}_n))\,)$	0.286	0.248	0.221	0.203	0.175	0.157
$\sigma(\,le(I^{T_{d,n}}(\tilde{\mathbf{V}}_n))\,)$	0.042	0.031	0.025	0.020	0.015	0.012

Tab. 4. Sampling from multivariate Cauchy $t_7(\mathbf{0}, \mathbf{W}, 1)$ distribution.

ACKNOWLEDGEMENT

The author is grateful to two anonymous referees for their valuable comments.

This research was supported by the project 2/0038/12 and by the project 2/0047/15 from the Scientific Grant Agency VEGA of the Slovak republic.

(Received January 9, 2014)

REFERENCES

[1] B. Abdous, C. Genest, and B. Rémillard: Dependence properties of meta-elliptical distributions. In: Statistical Modelling and Analysis for Complex Data Problems (P. Duchesne and B. Rémillard,eds.), Springer, New York, 2005, pp. 1–15. DOI:10.1007/0-387-24555-3_1

[2] A. S. C. Ehrenberg: On sampling from a population of rankers. Biometrika *39* (1952), 82–87. DOI:10.1093/biomet/39.1-2.82

[3] C. Genest, J. Nešlehová, and N. Ben Ghorbal: Estimators based on Kendall's tau in multivariate copula models. Austral. and New Zealand J. Statist. *53* (2011), 157–177. DOI:10.1111/j.1467-842x.2011.00622.x

[4] L. A. Goodman: A simple simultaneous test procedure for quasi-independence in contingency tables. J. Royal Statist. Soc., Ser. C *20* 2(1971), 165–177. DOI:10.2307/2346464

[5] J. Hájek, Z. Šidák, and P. K. Sen: Theory of Rank Tests. Academic Press, San Diego 1999.

[6] W. Hoeffding: A class of statistics with asymptotically normal distribution. The Annals Math. Statist. *19* (1948), 293–325. DOI:10.1214/aoms/1177730196

[7] H. Hult and F. Lindskog: Multivariate extremes, aggregation and dependence in elliptical distributions. Adv. Appl. Probab. *34* (2002), 587–608. DOI:10.1239/aap/1033662167

[8] M. G. Kendall and B. Babington Smith: On the method of paired comparisons. Biometrika *31* (1940), 324–345. DOI:10.1093/biomet/31.3-4.324

[9] J. F. C. Kingman and S. J. Taylor: Introduction to Measure and Probability. Cambridge University Press, Cambridge 1966. DOI:10.1017/cbo9780511897214

[10] A. J. Lee: U-Statistics: Theory and Practice. Marcel Dekker, Inc., New York 1990.

[11] A. Liu, Q. Li, C. Liu, K. Yu, and K. F. Yu: A rank-based test for comparison of multidimensional outcomes JASA *105* (2010), 578–587. DOI:10.1198/jasa.2010.ap09114

[12] H. El Maache and Y. Lepage: Spearman's rho and Kendalls's tau for multivariate data sets. In: Lecture Notes – Monograph Series *42*, Mathematical Statistics and Applications, Festschrift for Constance van Eeden, Beachwood 2003, pp. 113–130.

[13] C. R. Rao: Linear Statistical Inference and its Applications. John Wiley and Sons, New York 1973. DOI:10.1002/9780470316436

[14] P. K. Sen: On some convergence of U-statistics. Cal. Statist. Assoc. Bull. *10* (1960), 1–18.

František Rublík, Institute of Measurement Science, Slovak Academy of Sciences, Dúbravská cesta 9, 84104 Bratislava. Slovak Republic.
e-mail: umerrubl@savba.sk