SPEARMAN'S *RHO* AND KENDALL'S *TAU* FOR MULTIVARIATE DATA SETS

HAMANI EL MAACHE AND YVES LEPAGE
Université de Montréal

A class of U-statistics matrices is introduced to obtain the distribution of the matrices of the Spearman and Kendall correlation coefficients between the components of a random vector. These results are used to construct nonparametric tests of independence between two sets of variables based on three measures of multivariate relationship. The tests are illustrated by an example and a simulation study is performed to compare the tests based on Kendall's matrix with those based on Spearman's matrix.

1. Introduction

Let $F(x) = F(x^{[1]}, x^{[2]})$ be the continuous c.d.f. (cumulative distribution function) of a random vector $X = (X^{[1]}, X^{[2]})'$, where $x = (x^{(1)}, \ldots, x^{(m)})' \in \mathbb{R}^m$, $m \geq 2$, $x^{[1]} \in \mathbb{R}^p$, $x^{[2]} \in \mathbb{R}^q$ $(p + q = m)$ and $F^{[k]}(x^{[k]})$ $(k = 1, 2)$ denote the marginal c.d.f. of $X^{[k]}$. The objective of this paper is to detect deviation from the null hypothesis of independence that is, to test H_0: $F(x) = F^{[1]}(x^{[1]})F^{[2]}(x^{[2]})$ against appropriate classes of alternatives $\mathrm{H}_{1:n}$. A nonparametric approach to this problem was explored by Puri, Sen and Gokhale (1970) who defined a class of association parameters based on componentwise ranking. The statistic they proposed uses the elements of the matrix $D_n = \left(\begin{smallmatrix} D_{11} & D_{12} \\ D_{21} & D_{22} \end{smallmatrix}\right)$, where

$$(1.1) \qquad D_n^{(i,j)} = \frac{1}{n} \sum_{\alpha=1}^{n} J\left(\frac{R_\alpha^{(i)}}{n}\right) J\left(\frac{R_\alpha^{(j)}}{n}\right), \quad i, j = 1, \ldots, m.$$

Here, $R_\alpha^{(i)}$ is the rank of $X_\alpha^{(i)}$, that denote the ith coordinate of the vector X_{α},; the symbol α will run over the sample (from X) with $\alpha = 1, \ldots, n$ and J represents an arbitrary standardized score function. Puri, Sen and Gokhale (1970) established the joint asymptotic multivariate normality of the vector formed by the elements of D_n.

When the score function is $J(u) = J_0(u) = \sqrt{12}(u - \frac{1}{2})$, then

$$(1.2) \quad D_n^{(i,j)} = \frac{12}{n(n^2-1)} \sum_{\alpha=1}^{n} \left(R_\alpha^{(i)} - \frac{n+1}{2}\right)\left(R_\alpha^{(j)} - \frac{n+1}{2}\right), \qquad i, j = 1, \ldots, m,$$

AMS subject classifications: Primary 62H20; Secondary 62H15..

Keywords and phrases: U-statistics, Spearman *rho* correlation coefficient, Kendall *tau* correlation coefficient, measures of multivariate association, rank correlation, tests of independence..

which reduces to Spearman's rank correlation with asymptotic mean given by Spearman's coefficient (see Hoeffding, 1948, p. 318)

$$(1.3)\quad \varrho^{(i,j)} = 3\iint [2F^{(i)}(x^{(i)}) - 1][2F^{(j)}(x^{(j)}) - 1]\, dF^{(i,j)}(x^{(i)}, x^{(j)}), \qquad i, j = 1, \ldots, m,$$

where $F^{(i)}(x^{(i)})$ and $F^{(i,j)}(x^{(i)}, x^{(j)})$ denote the marginals c.d.f. of $X_\alpha^{(i)}$ and $(X_\alpha^{(i)}, X_\alpha^{(j)})'$ respectively.

They based their test of independence on the statistic $S^J = |D_n| \times (|D_{11}|\,|D_{22}|)^{-1}$, where $|A|$ denotes the determinant of A. They also showed that under H_0, $-n \log S^J \xrightarrow{\mathcal{L}} \mathcal{X}^2_{pq}$. With J_0, the statistic S^J is a generalization of Spearman's *rho* for multivariate data sets.

Using the results of Puri, Sen and Gokhale (1970) with $J_0(u)$, Cléroux, Lazraq and Lepage (1995) and Lazraq, Lepage and Cléroux (1995) proposed other tests of independence between two or more random vectors which are based on the measures of multivariate association proposed by Escoufier (1973) and Cramer and Nicewander (1979).

In the present paper, we present an approach based an original concept of U-statistics matrix inspired from Hoeffding (1948) to the problem of detecting dependence between two random vectors. This theoretical tool allows us to deduce the asymptotic distribution of a general association matrix. The first application is to construct the association matrix with Kendall's *tau* and study its relationship with Spearman's *rho*. We also propose nonparametric tests of independence between two random vectors based on three known measures of multivariate relationship with the Kendall and Spearman association matrices. We obtain the asymptotic distribution of the tests statistics under the null hypothesis and under a sequence of alternatives. In order to assess the behavior of the tests, a Monte Carlo study is performed to compare the empirical level and the empirical power of the tests based on Kendall's matrix with those based on Spearman's matrix.

Some multivariate generalizations of the Kendall's *tau* correlation coefficient have been studied in the literature by Hays (1960), Simon (1977) and Joe (1990). They have used the Kendall's *tau* correlation coefficient to test the total independence but not for the independence of two or more random vectors.

The paper is organized as follows. In Section 2, we give the asymptotic distribution of the matrices of U-statistics and deduce those for Spearman's matrix and Kendall's matrix. Section 3 is concerned with the three known measures of multivariate relationship: some properties and their asymptotic distributions under the null hypothesis and under a sequence of alternatives are given. In Section 4, we propose some tests of independence based on Spearman's and Kendall's matrices. We illustrate all the tests by an example.

Finally, Section 5 contains an empirical comparison of the new tests based on Kendall's matrix with the competitors based on Spearman's matrix. The results of this paper, can easily be extended to test the independence between several random vectors.

2. U-statistics matrix

Let $X_1, \ldots, X_n$ be n independent random vectors, $X_\alpha = (X_\alpha^{(1)}, \ldots, X_\alpha^{(m)})'$, $\alpha = 1, \ldots, n$, from an unknown continuous c.d.f. F. Let $\Phi^{(i,j)}(x_1, \ldots, x_{r^{(i,j)}})$, for $i = 1, \ldots, p$ and $j = 1, \ldots, q$, be symmetric function with $r^{(i,j)}$ ($r^{(i,j)} \in \mathbb{N}$) arguments. Let

$$U_n^{(i,j)} = \frac{1}{\binom{n}{r^{(i,j)}}} \sum_{\beta \in B} \Phi^{(i,j)}(X_{\beta_1}, \ldots, X_{\beta_{r^{(i,j)}}}),$$

where $B = \{\beta = (\beta_1, \ldots, \beta_{r^{(i,j)}}) \mid 1 \leq \beta_1 < \cdots < \beta_{r^{(i,j)}} \leq n\}$, be a U-statistic for the parameter $\gamma^{(i,j)}$ of degree $r^{(i,j)}$ based on the symmetric kernel $\Phi^{(i,j)}$. Let

(2.1)

$$\Phi_1^{(i,j)}(x) = \mathrm{E}[\Phi^{(i,j)}(x, X_2, \ldots, X_{\beta_{r^{(i,j)}}})], \quad \text{for } i = 1, \ldots, p \text{ and } j = 1, \ldots, q.$$

We note that $\mathrm{E}[U_n^{(i,j)}] = \mathrm{E}[\Phi_1^{(i,j)}(X)] = \gamma^{(i,j)}$ (see Hoeffding, 1948)). We now define the matrices of U-statistics U_n, of degrees R and of parameters Γ by respectively

$$U_n = \begin{pmatrix} U^{(1,1)} & \ldots & U^{(1,q)} \\ \vdots & \ddots & \vdots \\ U^{(p,1)} & \ldots & U^{(p,q)} \end{pmatrix}, \quad R = \begin{pmatrix} r^{(1,1)} & \ldots & r^{(1,q)} \\ \vdots & \ddots & \vdots \\ r^{(p,1)} & \ldots & r^{(p,q)} \end{pmatrix}$$

and

$$\Gamma = \begin{pmatrix} \gamma^{(1,1)} & \ldots & \gamma^{(1,q)} \\ \vdots & \ddots & \vdots \\ \gamma^{(p,1)} & \ldots & \gamma^{(p,q)} \end{pmatrix}.$$

Consider $\operatorname{vec} U_n$, as the vector formed by stacking the columns of U_n. The asymptotic multivariate normality of $\operatorname{vec} U_n$ follows from Theorem 7.1 of Hoeffding (1948).

Theorem 2.1. *If the kernel function $\Phi^{(i,j)}$ for the parameter $\gamma^{(i,j)}$ of degree $r^{(i,j)}$ is such that*

$$\mathrm{E}[\Phi^{(i,j)}(X_1, \ldots, X_{r^{(i,j)}})] = \gamma^{(i,j)} \quad \text{and} \quad \mathrm{E}\big[\big(\Phi^{(i,j)}(X_1, \ldots, X_{r^{(i,j)}})\big)^2\big] < \infty,$$

for $i = 1, \ldots, p$ and $j = 1, \ldots, q$, then $\sqrt{n}(\operatorname{vec} U_n - \operatorname{vec} \Gamma) \xrightarrow{\mathcal{L}} \mathcal{N}_{pq}(0, \Omega)$ where the elements of Ω are given by

$$m^{(ij,kl)} = r^{(i,j)} r^{(k,l)} [\mathrm{E}[\Phi_1^{(i,j)}(X_1) \Phi_1^{(k,l)}(X_1)] - \gamma^{(i,j)} \gamma^{(k,l)}]$$

$\Phi_1^{(i,j)}(x)$ *and* $\Phi_1^{(k,l)}(x)$ *are given in* (2.1) *for* $i,k=1,\ldots,p$ *and* $j,l=1,\ldots,q$.

We can also deduce from Hoeffding (1948) that $\operatorname{vec} U_n \xrightarrow{P} \operatorname{vec}\Gamma$.

Spearman's matrix

To express the rank correlation in terms of indicators, we define the signum function as $s(x)=1$ if $x>0$, 0 if $x=0$ and -1 if $x<0$. Then we can define the U-statistic

$$\mathcal{S}_n^{(i,j)} = \frac{1}{\binom{n}{3}} \sum\sum\sum_{1\le\alpha<\beta<\nu\le n} \Psi^{(i,j)}(\,X_\alpha, X_\beta, X_\nu)$$

for Spearman's coefficient $\varrho^{(i,j)}$ of degree 3 based on the kernel function

$$\Psi^{(i,j)}(X_1,X_2,X_3) = \frac{1}{2} \sum\sum\sum_{1\le\alpha\ne\beta\ne\nu\le 3} s\big(X_\alpha^{(i)} - X_\beta^{(i)}\big) s\big(X_\alpha^{(j)} - X_\nu^{(j)}\big).$$

Here, we have (see Hoeffding, 1948, p. 320)

$$\begin{aligned}(2.2)\quad \Psi_1^{(i,j)}(X_\alpha) &= [1-2F^{(i)}(X_\alpha^{(i)})][1-2F^{(j)}(X_\alpha^{(j)})] \\ &\quad + 4\int [F^{(i,j)}(x^{(i)}, X_\alpha^{(j)}) - F^{(i)}(x^{(i)})F^{(j)}(X_\alpha^{(j)})]\, dF^{(i)}(x^{(i)}) \\ &\quad + 4\int [F^{(i,j)}(X_\alpha^{(i)}, x^{(j)}) - F^{(i)}(X_\alpha^{(i)})F^{(j)}(x^{(j)})]\, dF^{(j)}(x^{(j)})\end{aligned}$$

where $\Psi_1^{(i,j)}(x) = \mathrm{E}[\Psi^{(i,j)}(x,X_2,X_3)]$. For $i=j$, we have $\mathcal{S}_n^{(i,j)} = \varrho^{(i,j)} = 1$. Obviously $\mathcal{S}_n^{(i,j)}$ is an unbiased estimator of $\varrho^{(i,j)}$ while $D^{(i,j)}$ given by (1.2)) is not.

The matrix $\mathcal{S}_n = (\mathcal{S}_n^{(i,j)})_{i,j=1,\ldots,m}$ will be called Spearman's matrix for the parameter matrix $P = (\varrho^{(i,j)})_{i,j=1\ldots,m}$. For all $i\ne j$ the degree is 3 and zero for $i=j$. The application of Theorem 2.1 leads immediately to the following theorem.

Theorem 2.2. *The random vector* $\sqrt{n}(\operatorname{vec}\mathcal{S}_n - \operatorname{vec}P)$ *has a limiting* m^2-*multivariate normal distribution* $\mathcal{N}_{m^2}(O,\Sigma_{\mathcal{S}})$ *where the elements of* $\Sigma_{\mathcal{S}}$ *are given by* $\sigma_{\mathcal{S}}^{(ij,kl)} = 9\sum_{h=1}^3\sum_{h'=1}^3 \operatorname{Cov}(V_1^{(i,j),h}, V_1^{(k,l),h'})$ *with* $i,j,k,l=1,\ldots,m$,

$$V_1^{(i,j),1} = [1-2F^{(i)}(X_1^{(i)})][1-2F^{(j)}(X_1^{(j)})],$$

$$V_1^{(i,j),2} = 4\int [F^{(i,j)}(x^{(i)}, X_1^{(j)}) - F^{(i)}(x^{(i)})F^{(j)}(X_1^{(j)})]\, dF^{(i)}(x^{(i)})$$

and

$$V_1^{(i,j),3} = 4\int [F^{(i,j)}(X_1^{(i)}, x^{(j)}) - F^{(i)}(X_1^{(i)})F^{(j)}(x^{(j)})]\, dF^{(j)}(x^{(j)}).$$

Kendall's matrix

Kendall's *tau* is a measure defined by the product moment correlation of signs of concordance,

$$K_n^{(i,j)} = \frac{1}{\binom{n}{2}} \sum\sum_{1\le\alpha<\beta\le n} s(X_\beta^{(i)} - X_\alpha^{(i)})s(X_\beta^{(j)} - X_\alpha^{(j)}),$$

while Spearman's rank correlation coefficient is the product moment correlation between $F^{(i)}(X^{(i)})$ and $F^{(j)}(X^{(j)})$ $(i, j = 1, \ldots, m)$ (see Cléroux, Lazraq and Lepage, 1995, p. 719). Thus, Theorem 4.1 in Puri, Sen and Gokhale (1970) cannot be used to obtain the asymptotic multivariate normality of the elements of the Kendall's matrix. The element $K_n^{(i,j)}$ is a U-statistic of degree 2 based on the symmetric kernel

$$\Phi^{(i,j)}(X_1, X_2) = s(X_2^{(i)} - X_1^{(i)})s(X_2^{(j)} - X_1^{(j)})$$

for Kendall's coefficient defined as

$$\tau^{(i,j)} = 4 \iint F^{(i,j)}(x^{(i)}, x^{(j)})\, dF^{(i,j)}(x^{(i)}, x^{(j)}) - 1.$$

Here also (see Hoeffding, 1948, p. 316), we have

$$\begin{aligned}(2.3)\quad \Phi_1^{(i,j)}(X_\alpha) &= 1 - 2F^{(i)}(X_\alpha^{(i)}) - 2F^{(j)}(X_\alpha^{(j)}) + 4F^{(i,j)}(X_\alpha^{(i)}, X_\alpha^{(j)}) \\ &= [1 - 2F^{(i)}(X_\alpha^{(i)})][1 - 2F^{(j)}(X_\alpha^{(j)})] \\ &\quad + 4\big[F^{(i,j)}(X_\alpha^{(i)}, X_\alpha^{(j)}) - F^{(i)}(X_\alpha^{(i)})F^{(j)}(X_\alpha^{(j)})]\end{aligned}$$

where $\Phi_1^{(i,j)}(x) = \mathrm{E}[\Phi^{(i,j)}(x, X_2)]$. For $i = j$, we have $K_n^{(i,j)} = \tau^{(i,j)} = 1$.

The matrix $K_n = (K_n^{(i,j)})_{i,j=1,\ldots,m}$ will be called Kendall's matrix for the parameter matrix $\Lambda = (\tau^{(i,j)})_{i,j=1,\ldots,m}$. For all $i \neq j$, the degree is 2 while it is zero for $i = j$. The application of Theorem 2.1 leads immediately to the following theorem.

Theorem 2.3. *The random vector* $\sqrt{n}(\mathrm{vec}\, K_n - \mathrm{vec}\, \Lambda)$ *has a limiting* m^2*-multivariate normal distribution* $\mathcal{N}_{m^2}(O, \Sigma_K)$ *where the elements of* Σ_K *are given by* $\sigma_K^{(ij,kl)} = 4\sum_{h=1}^2 \sum_{h'=1}^2 \mathrm{Cov}(U_1^{(i,j),h}, U_1^{(k,l),h'})$ *with* i, j, k *and* $l = 1, \ldots, m$,

$$U_1^{(i,j),1} = [1 - 2F^{(i)}(X_1^{(i)})][1 - 2F^{(j)}(X_1^{(j)})]$$

and

$$U_1^{(i,j),2} = 4[F^{(i,j)}(X_1^{(i)}, X_1^{(j)}) - F^{(i)}(X_1^{(i)})F^{(j)}(X_1^{(j)})].$$

If we insert the rank $R_\alpha^{(i)}$ of $X_\alpha^{(i)}$ defined by

$$R_\alpha^{(i)} = \frac{n+1}{2} + \frac{1}{2}\sum_{\beta=1}^{n} s(X_\alpha^{(i)} - X_\beta^{(i)})$$

in (1.2), we have

$$D_n = \frac{n-2}{n+1}\mathcal{S}_n + \frac{3}{n+1}K_n$$

(see Hoeffding, 1948, p. 318). Then, $\sqrt{n}(\operatorname{vec} D_n - \operatorname{vec} P)$ and $\sqrt{n}(\operatorname{vec} \mathcal{S}_n - \operatorname{vec} P)$ have the same limiting distribution given by Theorem 2.2; we find here the result given by Puri, Sen and Gokhale (1970).

Let us now partition P and Λ and their analogue sample matrices $\mathcal{S}_n$ and K_n in following way:

$$P = \begin{pmatrix} P_{11} & P_{12} \\ P_{21} & P_{22} \end{pmatrix}, \quad \Lambda = \begin{pmatrix} \Lambda_{11} & \Lambda_{12} \\ \Lambda_{21} & \Lambda_{22} \end{pmatrix}, \quad K_n = \begin{pmatrix} K_{11} & K_{12} \\ K_{21} & K_{22} \end{pmatrix}$$

and

$$\mathcal{S}_n = \begin{pmatrix} \mathcal{S}_{11} & \mathcal{S}_{12} \\ \mathcal{S}_{21} & \mathcal{S}_{22} \end{pmatrix}$$

where M_{21} ($M = P, \Lambda, \mathcal{S}$ or K) is of order $q \times p$. Now, we have $P_{21} = \Lambda_{21} = O$ under H_0. Under H_0, $X_\alpha^{(i)}$ and $X_\alpha^{(j)}$ are independent for $i = p+1, \dots, m$ and $j = 1, \dots, p$.

Theorem 2.4. *Under* H_0 *and when* $n \to \infty$, *we have*

$$\sqrt{n}\operatorname{vec} K_{21} \xrightarrow{\mathcal{L}} Z^{(\tau)} \quad \textit{where } Z^{(\tau)} \textit{ follows a } \mathcal{N}_{pq}(O, \tfrac{4}{9}P_{11} \otimes P_{22})$$

and

$$\sqrt{n}\operatorname{vec} \mathcal{S}_{21} \xrightarrow{\mathcal{L}} Z^{(\varrho)} \quad \textit{where } Z^{(\varrho)} \textit{ follows a } \mathcal{N}_{pq}(O, P_{11} \otimes P_{22}).$$

Proof. From Theorem 2.3, we note that under H_0 the random vector $\sqrt{n}\operatorname{vec} K_{21}$ has a limiting multivariate normal distribution $\mathcal{N}_{pq}(O, A)$ where the elements of A are given for $i, j, k, l = 1, \dots, m$ by

$$\begin{aligned} \sigma_K^{(ij,kl)} &= 4\,\mathrm{E}(U_1^{(i,j),1} U_1^{(k,l),1}) \\ &= 4\,\mathrm{E}[1 - 2F^{(i)}(X_1^{(i)})][1 - 2F^{(k)}(X_1^{(k)})] \\ &\qquad\qquad \times \mathrm{E}[1 - 2F^{(j)}(X_1^{(j)})][1 - 2F^{(l)}(X_1^{(l)})] \\ &= \tfrac{4}{9}\varrho^{(i,k)}\varrho^{(j,l)}. \end{aligned}$$

Thus, $A = \frac{4}{9}P_{11} \otimes P_{22}$. In a similar way, we can obtain the limiting multivariate distribution of $\sqrt{n}\operatorname{vec} \mathcal{S}_{21}$ under H_0. □

We shall now study the asymptotic distribution of M_{21} ($M = \mathcal{S}$ or K) under a sequence of alternatives $\{\mathrm{H}_{1:n}, n = 1, 2, \dots\}$ (see Puri, Sen and Gokhale, 1970) which specifies that

$$\mathrm{H}_{1:n}: \ F(x) = F^{[1]}(x^{[1]})F^{[2]}(x^{[2]})\left(1 + \frac{\Omega^{([1],[2])}\big(F^{[1]}(x^{[1]}), F^{[2]}(x^{[2]})\big)}{\sqrt{n}}\right)$$

where $\Omega^{([1],[2])}$ is some function of $\big(F^{[1]}(x^{[1]}), F^{[2]}(x^{[2]})\big)$ and $\Omega^{([1],[2])} \neq 0$. $\mathrm{H}_{1:n}$ implies that for $i = p+1, \dots, m$ and $j = 1, \dots, p$,

$$\text{(2.4)} \quad F^{(i,j)}(x^{(i)}, x^{(j)}) = F^{(i)}(x^{(i)})F^{(j)}(x^{(j)}) \times \left(1 + \frac{\Omega^{(i,j)}\big(F^{(i)}(x^{(i)}), F^{(j)}(x^{(j)})\big)}{\sqrt{n}}\right)$$

where $\Omega^{(i,j)}$ is a function of $(F^{(i)}, F^{(j)})$ and $\Omega^{(i,j)} \neq 0$; it also implies that

$$\text{(2.5)} \quad F^{(ij,kl)}(x^{(i)}, x^{(j)}, x^{(k)}, x^{(l)}) = F^{(i,k)}(x^{(i)}, x^{(k)})F^{(j,l)}(x^{(j)}, x^{(l)}) \times \left(1 + \frac{\Omega^{(ij,kl)}\big(F^{(i,k)}(x^{(i)}, x^{(k)}), F^{(j,l)}(x^{(j)}, x^{(l)})\big)}{\sqrt{n}}\right)$$

where $F^{(ij,kl)}$ is the c.d.f. of the $(X^{(i)}, X^{(j)}, X^{(k)}, X^{(l)})$ for $j, l = 1, \dots, p$; $i, k = p+1, \dots, m$, and $\Omega^{(ij,kl)} \neq 0$ is a function of $(F^{(i,k)}, F^{(j,l)})$.

Let for $i = p+1, \dots, m$ and $j = 1, \dots, p$,

$$\text{(2.6)} \quad \begin{aligned} dF^{(i,j)} &= f^{(i,j)}(x^{(i)}, x^{(j)})dx^{(i)}dx^{(j)} \\ &= \frac{\partial^2 F^{(i,j)}(x^{(i)}, x^{(j)})}{\partial x^{(i)} \partial x^{(j)}}\, dx^{(i)}\, dx^{(j)} \\ &= dF^{(i)}\, dF^{(j)}\left(1 + \frac{1}{\sqrt{n}}\omega_{ij}(F^{(i)}(x^{(i)}), F^{(j)}(x^{(j)}))\right), \end{aligned}$$

where the function ω_{ij} is obtained by differentiating (2.4)). In a similar way, let for $i = p+1, \dots, m$ and $j = 1, \dots, p$,

$$\text{(2.7)} \quad \begin{aligned} f^{(ij,kl)}&(x^{(i)}, x^{(j)}, x^{(k)}, x^{(l)}) \\ &= \frac{\partial^4 F^{(ij,kl)}(x^{(i)}, x^{(j)}, x^{(k)}, x^{(l)})}{\partial x^{(i)} \partial x^{(j)} \partial x^{(k)} \partial x^{(l)}} \\ &= \frac{\partial^2 F^{(i,k)}(x^{(i)}, x^{(k)})}{\partial x^{(i)} \partial x^{(k)}} \frac{\partial^2 F^{(j,l)}(x^{(j)}, x^{(l)})}{\partial x^{(j)} \partial x^{(l)}} + \frac{1}{\sqrt{n}}\omega_{ij,kl}. \end{aligned}$$

To simplify the notations, we set

$$\begin{aligned} dF^{(ij,kl)} &= f^{(ij,kl)}(x^{(i)}, x^{(j)}, x^{(k)}, x^{(l)})\, dx^{(i)}\, dx^{(j)}\, dx^{(k)}\, dx^{(l)} \\ &= \frac{\partial^4 F^{(ij,kl)}(x^{(i)}, x^{(j)}, x^{(k)}, x^{(l)})}{\partial x^{(i)} \partial x^{(j)} \partial x^{(k)} \partial x^{(l)}}\, dx^{(i)}\, dx^{(j)}\, dx^{(k)}\, dx^{(l)} \end{aligned}$$

and

$$dF^{(i,k)} = f^{(i,k)}(x^{(i)}, x^{(k)})dx^{(i)}dx^{(k)}.$$

Let $B = (\beta^{(i,j)})$ be the $q \times p$ matrix where

$$\beta^{(i,j)} = \iint F^{(i)}(x^{(i)})F^{(j)}(x^{(j)})\Omega^{(i,j)}\big(F^{(i)}(x^{(i)}), F^{(j)}(x^{(j)})\big)\, dF^{(i,j)}.$$

Using (1.3) and $\Psi_1^{(i,j)}$ defined by (2.2) where

$$\mathrm{E}[\Psi_1^{(i,j)}(X_1)] = \varrho^{(i,j)} = 3\iint [2F^{(i)}(x^{(i)}) - 1][2F^{(j)}(x^{(j)}) - 1]\, dF^{(i,j)},$$

we obtain under $\mathrm{H}_{1:n}$,

$$\begin{aligned}\varrho^{(i,j)} &= \frac{1}{3}\varrho^{(i,j)} + 8\iint \big[F^{(i,j)}(x^{(i)}, x^{(j)}) - F^{(i)}(x^{(i)})F^{(j)}(x^{(j)})\big]dF^{(i,j)} \\ &= \frac{1}{3}\varrho^{(i,j)} + \frac{8}{\sqrt{n}}\beta^{(i,j)} = \frac{12\beta^{(i,j)}}{\sqrt{n}}.\end{aligned}$$

In a similar way, using $\Phi_1^{(i,j)}$ defined by (2.3), we obtain under $\mathrm{H}_{1:n}$,

$$\tau^{(i,j)} = \frac{1}{3}\varrho^{(i,j)} + \frac{4}{\sqrt{n}}\beta^{(i,j)} = \frac{8}{\sqrt{n}}\beta^{(i,j)}.$$

We thus have shown the following lemma.

Lemma 2.1. *Under* $\mathrm{H}_{1:n}$*, we have*

$$\Lambda_{21} = \frac{8}{\sqrt{n}}B \quad \textit{and} \quad P_{21} = \frac{12}{\sqrt{n}}B.$$

The next theorem gives the limiting distribution of K_{21} and $\mathcal{S}_{21}$ under the sequence $\mathrm{H}_{1:n}$.

Theorem 2.5. *Under* $\mathrm{H}_{1:n}$ *and when* $n \to \infty$*, we have*

$$\begin{aligned}&\sqrt{n}\,\mathrm{vec}\,K_{21} \xrightarrow{\mathcal{L}} Z^{(\tau)} \quad \textit{where } Z^{(\tau)} \textit{ follows a } \mathcal{N}_{pq}(8\,\mathrm{vec}\,B, \tfrac{4}{9}P_{11}\otimes P_{22}),\\ &\sqrt{n}\,\mathrm{vec}\,\mathcal{S}_{21} \xrightarrow{\mathcal{L}} Z^{(\varrho)} \quad \textit{where } Z^{(\varrho)} \textit{ follows a } \mathcal{N}_{pq}(12\,\mathrm{vec}\,B, P_{11}\otimes P_{22}).\end{aligned}$$

Proof. From Theorem 2.3, the random vector $\sqrt{n}\,\mathrm{vec}\,K_{21}$ has a limiting multivariate distribution with mean vector $\mathrm{E}[\sqrt{n}\,\mathrm{vec}\,K_{21}] = 8\,\mathrm{vec}\,B$ and covariance matrix $\frac{4}{9}P_{11}\otimes P_{22}$ whose its elements are

$$\sigma_K^{(ij,kl)} = 4\,\mathrm{Cov}(U_1^{(i,j),1}, U_1^{(k,l),1}) = \tfrac{4}{9}\varrho^{(i,k)}\varrho^{(j,l)}.$$

Using the expression for $dF^{(ij,kl)}$ given by equation (2.7), we have

$$\begin{aligned}
&\mathrm{Cov}(U_1^{(i,j),1}, U_1^{(k,l),1}) \\
&\quad = \mathrm{E}[U_1^{(i,j),1} U_1^{(k,l),1}] - \mathrm{E}[U_1^{(i,j),1}]\,\mathrm{E}[U_1^{(k,l),1}] \\
&\quad = \left(\iint [1 - 2F^{(i)}(x^{(i)})][1 - 2F^{(k)}(x^{(k)})]\, dF^{(i,k)}\right) \\
&\qquad\qquad \times \left(\iint [1 - 2F^{(j)}(x^{(j)})][1 - 2F^{(l)}(x^{(l)})]\, dF^{(j,l)}\right) \\
&\qquad + \frac{1}{\sqrt{n}} \iiiint [1 - 2F^{(i)}(x^{(i)})][1 - 2F^{(j)}(x^{(j)})][1 - 2F^{(k)}(x^{(k)})] \\
&\qquad\qquad \times [1 - 2F^{(l)}(x^{(l)})]\omega_{ij,kl}\, dx^{(i)}\, dx^{(j)}\, dx^{(k)}\, dx^{(l)} \\
&\qquad - \frac{8}{n}\beta^{(i,j)}\beta^{(k,l)} \\
&\quad = \frac{1}{9}\varrho^{(i,k)}\varrho^{(j,l)} \\
&\qquad + \frac{1}{\sqrt{n}} \iiiint [1 - 2F^{(i)}(x^{(i)})][1 - 2F^{(j)}(x^{(j)})][1 - 2F^{(k)}(x^{(k)})] \\
&\qquad\qquad \times [1 - 2F^{(l)}(x^{(l)})]\omega_{ij,kl}\, dx^{(i)}\, dx^{(j)}\, dx^{(k)}\, dx^{(l)} \\
&\qquad - \frac{8}{n}\beta^{(i,j)}\beta^{(k,l)} \\
&\quad = \frac{1}{9}\varrho^{(i,k)}\varrho^{(j,l)} + O(n^{-1/2}).
\end{aligned}$$

The result follows from Serfling (1980) (Lemma A, p. 20). In a similar way, we have the limiting distribution of $\sqrt{n}\,\mathrm{vec}\,\mathcal{S}_{21}$ from Theorem 2.2. □

3. Measures of association

We now apply the measures of multivariate relationship proposed by Escoufier (1973), Stewart and Love (1968) and Cramer and Nicewander (1979) to the Kendall and Spearman matrices.

For the Escoufier's measure (1973), we have

$$\mathrm{RV}^{(\tau)} = \frac{\mathrm{tr}(K_{12}K_{12}')}{\sqrt{\mathrm{tr}(K_{11}^2)\,\mathrm{tr}(K_{22}^2)}} \quad \text{and} \quad \mathrm{RV}^{(\varrho)} = \frac{\mathrm{tr}(\mathcal{S}_{12}\mathcal{S}_{12}')}{\sqrt{\mathrm{tr}(\mathcal{S}_{11}^2)\,\mathrm{tr}(\mathcal{S}_{22}^2)}}.$$

The Stewart and Love's measure (1968) gives

$$\mathrm{SL}^{(\tau)} = \frac{\mathrm{tr}(K_{12}K_{22}^{-1}K_{12}')}{p} \quad \text{and} \quad \mathrm{SL}^{(\varrho)} = \frac{\mathrm{tr}(\mathcal{S}_{12}\mathcal{S}_{22}^{-1}\mathcal{S}_{12}')}{p}.$$

Finally with the Cramer and Nicewander's measure (1979), we have

$$\mathrm{CN}^{(\tau)} = \frac{\mathrm{tr}(K_{11}^{-1}K_{12}K_{22}^{-1}K_{12}')}{p} \quad \text{and} \quad \mathrm{CN}^{(\varrho)} = \frac{\mathrm{tr}(\mathcal{S}_{11}^{-1}\mathcal{S}_{12}\mathcal{S}_{22}^{-1}\mathcal{S}_{12}')}{p}.$$

The corresponding measures at the level of the population are defined by:

$$\rho\text{RV}^{(\tau)} = \frac{\text{tr}(\Lambda_{12}\Lambda'_{12})}{\sqrt{\text{tr}(\Lambda^2_{11})\,\text{tr}(\Lambda^2_{22})}} \quad \text{and} \quad \rho\text{RV}^{(\varrho)} = \frac{\text{tr}(P_{12}P'_{12})}{\sqrt{\text{tr}(P^2_{11})\,\text{tr}(P^2_{22})}}$$

for the Escoufier's measure,

$$\rho\text{SL}^{(\tau)} = \frac{\text{tr}(\Lambda_{12}\Lambda^{-1}_{22}\Lambda'_{12})}{p} \quad \text{and} \quad \rho\text{SL}^{(\varrho)} = \frac{\text{tr}(P_{12}P^{-1}_{22}P'_{12})}{p}$$

for the Stewart and Love's measure,

$$\rho\text{CN}^{((\tau)} = \frac{\text{tr}(\Lambda^{-1}_{11}\Lambda_{12}\Lambda^{-1}_{22}\Lambda'_{12})}{p} \quad \text{and} \quad \rho\text{CN}^{((\varrho)} = \frac{\text{tr}(P^{-1}_{11}P_{12}P^{-1}_{22}P'_{12})}{p}$$

for the Cramer and Nicewander's measure.

The main advantage of considering these transformed measures are that: (a) the individual data may be ordinal variables, (b) the scale of measurement for each variable may be different, (c) the classical hypotheses of multivariate normality or ellipticity of the parent population may be omitted, (d) they lead to a robust procedure against outliers. Moreover, the three measures applied to Kendall's matrix or Spearman's matrix have the following properties:

(i) $\rho M^{(\tau)} = \rho M^{(\varrho)} = 0$ if and only if $P_{21} = \Lambda_{12} = 0$, for $M =$ RV, SL and CN.

(ii) when $p = q = 1$, the three measures reduce to the square of Kendall's coefficient or to the square of Spearman's coefficient between the variables $X^{(1)}$ and $X^{(2)}$.

(iii) $0 \leq \rho M^{(s)} \leq 1$, for $s = \tau$, ϱ and $M =$ RV, SL and CN. The sample analogue of the measures, $M^{(s)}$, for $s = \tau$, ϱ and $M =$ RV, SL and CN, have the same properties.

For the proof of these properties and other results on measures of multivariate relationship, the reader is referred to Lazraq and Cléroux (1988). The testing problem is now restated as H_0: $\rho M^{(s)} = 0$ versus $\rho M^{(s)} > 0$, for $s = \tau$, ϱ and $M =$ RV, SL and CN.

In the following theorems we give the asymptotic distribution of our statistics under the null hypothesis and under a sequence of alternatives. We will show that they are represented as linear combinations of independent central $\mathcal{X}^2$ and noncentral $\mathcal{X}^2$ random variables respectively.

Theorem 3.1. *Let K_n and $\mathcal{S}_n$ be Kendall's and Spearman's matrices respectively obtained from a sample of size n drawn from a m-dimensional random vector with an arbitrary continuous c.d.f. $F(x)$. Then, under* H_0 *and when $n \to \infty$, we have*

(i) $$n\mathrm{RV}^{(\tau)} \xrightarrow{\mathcal{L}} \frac{4}{9}\frac{1}{\sqrt{\mathrm{tr}(\Lambda_{11}^2)\,\mathrm{tr}(\Lambda_{22}^2)}}\sum_{i=1}^{p}\sum_{j=1}^{q}\lambda_i\mu_j U_{ij}^2,$$

(ii) $$n\mathrm{RV}^{(\varrho)} \xrightarrow{\mathcal{L}} \frac{1}{\sqrt{\mathrm{tr}(P_{11}^2)\,\mathrm{tr}(P_{22}^2)}}\sum_{i=1}^{p}\sum_{j=1}^{q}\lambda_i\mu_j U_{ij}^2,$$

where the U_{ij}'s are iid $\mathcal{N}(0,1)$, $i = 1,\ldots,p$; $j = 1,\ldots,q$, random variables and λ_i and μ_j are the eigenvalues of P_{11} and P_{22} respectively.

(iii) $$n\mathrm{SL}^{(\tau)} \xrightarrow{\mathcal{L}} \frac{4}{9p}\sum_{i=1}^{p}\sum_{j=1}^{q}\lambda_i t_j^{(2)} U_{ij}^2,$$

(iv) $$n\mathrm{SL}^{(\varrho)} \xrightarrow{\mathcal{L}} \frac{1}{p}\sum_{i=1}^{p}\lambda_i Z_{q,i}^2,$$

where the $Z_{q,i}^2$'s are iid $\mathcal{X}_q^2$, $i = 1,\ldots,p$, random variables with q degrees of freedom, λ_i, $i = 1,\ldots,p$ are the the eigenvalues of P_{11} and $t_j^{(2)}$, $j = 1,\ldots,q$, are the eigenvalues of $\Lambda_{22}^{-1}P_{22}$.

(v) $$n\mathrm{CN}^{(\tau)} \xrightarrow{\mathcal{L}} \frac{4}{9p}\sum_{i=1}^{p}\sum_{j=1}^{q} t_i^{(1)} t_j^{(2)} U_{ij}^2,$$

(vi) $$n\mathrm{CN}^{(\varrho)} \xrightarrow{\mathcal{L}} \frac{\mathcal{X}_{pq}^2}{p},$$

where $t_i^{(1)}$, $i = 1,\ldots,p$, are the eigenvalues of $\Lambda_{11}^{-1}P_{11}$.

Proof. (i) Since K_n converges in probability to Λ as $n \to \infty$, the submatrices K_{11} and K_{22} converges in probability to Λ_{11} and Λ_{22} respectively as $n \to \infty$. Furthermore, under H_0, $\sqrt{n}\,\mathrm{vec}\,K_{21}$ converges to $Z^{(\tau)}$ with distribution $\mathcal{N}_{pq}(O, \frac{4}{9}P_{11} \otimes P_{22})$ Theorem 2.3. Since

$$n\,\mathrm{tr}(K_{12}K_{21}) = (\sqrt{n}\,\mathrm{vec}\,K_{21})'(\sqrt{n}\,\mathrm{vec}\,K_{21}) \xrightarrow{\mathcal{L}} Z^{(\tau)\prime}Z^{(\tau)},$$

we deduce using classical results on quadratic form (see Baldessari, 1967 or Johnson and Kotz, 1970) that,

$$n\mathrm{RV}^{(\tau)} \xrightarrow{\mathcal{L}} \frac{4}{9}\frac{1}{\sqrt{\mathrm{tr}(\Lambda_{11}^2)\,\mathrm{tr}(\Lambda_{22}^2)}}\sum_{i=1}^{p}\sum_{j=1}^{q}\lambda_i\mu_j U_{ij}^2$$

where the U_{ij}'s are iid $\mathcal{N}(0,1)$, $i = 1, \ldots, p$; $j = 1, \ldots, q$, random variables and λ_i, μ_j are the eigenvalues of P_{11}, P_{22} respectively.

Noting that $n\,\mathrm{tr}(K_{12}K_{22}^{-1}K_{21}) = (\sqrt{n}\,\mathrm{vec}\,K_{21})'(I_p \otimes K_{22})^{-1}(\sqrt{n}\,\mathrm{vec}\,K_{21})$, we have

$$n\mathrm{SL}^{(\tau)} \xrightarrow{\mathcal{L}} \frac{4}{9p}\sum_{i=1}^{p}\sum_{j=1}^{q}\lambda_i t_j^{(2)} U_{ij}^2$$

where $t_j^{(2)}$, $j = 1, \ldots, q$, are the eigenvalues of $\Lambda_{22}^{-1}P_{22}$. For the case $n\mathrm{CN}^{(\tau)}$, we use

$$n\,\mathrm{tr}(K_{11}^{-1}K_{12}K_{22}^{-1}K_{21}) = (\sqrt{n}\,\mathrm{vec}\,K_{21})'(K_{11} \otimes K_{22})^{-1}(\sqrt{n}\,\mathrm{vec}\,K_{21}).$$

The proofs are analogous when Spearman's matrix is used. □

Theorem 3.2. *If the conditions of Theorem* 3.1 *are satisfied then under* $\mathrm{H}_{1:n}$ *and when* $n \to \infty$, *we have*

(i) $$n\mathrm{RV}^{(\tau)} \xrightarrow{\mathcal{L}} \frac{4}{9}\frac{1}{\sqrt{\mathrm{tr}(\Lambda_{11}^2)\,\mathrm{tr}(\Lambda_{22}^2)}}\sum_{i=1}^{p}\sum_{j=1}^{q}\lambda_i\mu_j U_{ij,1}^2,$$

where the $U_{ij,1}$*'s are independent* $\mathcal{N}(\delta_{ij}, 1)$, $i = 1, \ldots, p$, $j = 1, \ldots, q$, *random variables;*

(ii) $$n\mathrm{RV}^{(\varrho)} \xrightarrow{\mathcal{L}} \frac{1}{\sqrt{\mathrm{tr}(P_{11}^2)\,\mathrm{tr}(P_{22}^2)}}\sum_{i=1}^{p}\sum_{j=1}^{q}\lambda_i\mu_j U_{ij,2}^2,$$

where the $U_{ij,2}$*'s are independent* $\mathcal{N}(\sqrt{\frac{9}{4}\delta_{ij}^2}, 1)$, $i = 1, \ldots, p$, $j = 1, \ldots, q$, *random variables and* λ_i, μ_j *are the eigenvalues of* P_{11}, P_{22} *resp. corresponding to the normalized eigenvectors* a_i, b_j, $\delta_{ij}^2 = 64\,\mathrm{tr}(B'b_jb_j'P_{22}^{-1}BP_{11}^{-1}a_ia_i')$ *and* B *is the matrix defined in Lemma* 2.1*;*

(iii) $$n\mathrm{SL}^{(\tau)} \xrightarrow{\mathcal{L}} \frac{4}{9p}\sum_{i=1}^{p}\sum_{j=1}^{q}\lambda_i t_j^{(2)}\mathcal{X}_{1,ij}^2(\delta_{ij,1}^2),$$

where the $\mathcal{X}_{1,ij}^2(\delta_{ij,1}^2)$*'s are independent chi-squared random variables with one degree of freedom, with* $\delta_{ij,1}^2 = \mathrm{tr}(B'p_jp_j'P_{22}^{-1}BP_{11}^{-1}a_ia_i')$ *as noncentrality parameter and* p_j, $j = 1, \ldots, q$, *is the normalized eigenvector corresponding to the eigenvalue* $t_j^{(2)}$ *of* $\Lambda_{22}^{-1}P_{22}$*;*

(iv) $$n\mathrm{SL}^{(\varrho)} \xrightarrow{\mathcal{L}} \frac{1}{p}\sum_{i=1}^{p}\lambda_i\mathcal{X}_{i(q)}^2(\delta_i^2),$$

where the $\mathcal{X}^2_{i(q)}(\delta_i^2)$*'s,* $i = 1, \ldots, p$*, random variables are independent chi-squared random variables with* q *degrees of freedom and noncentrality parameter defined by* $\delta_i^2 = 64\,\mathrm{tr}(B'P_{22}^{-1}AP_{11}^{-1}a_ia_i')$*;*

(v)
$$n\mathrm{CN}^{(\tau)} \xrightarrow{\mathcal{L}} \frac{4}{9p}\sum_{i=1}^{p}\sum_{j=1}^{q} t_i^{(1)}t_j^{(2)}U_{ij,3}^2,$$

where the U_{ij}*'s are independent* $\mathcal{N}(\Delta_{ij}, 1)$*,* $i = 1, \ldots, p$*,* $j = 1, \ldots, q$*, random variables with* $\Delta_{ij}^2 = 64\,\mathrm{tr}(B'p_jp_j'P_{22}^{-1}BP_{11}^{-1}d_id_i')$ *as noncentrality parameter and* d_i*,* $i = 1, \ldots, p$*, is the normalized eigenvector corresponding to the eigenvalue* $t_i^{(1)}$ *of* $\Lambda_{11}^{-1}P_{11}$*;*

(vi)
$$n\mathrm{CN}^{(\varrho)} \xrightarrow{\mathcal{L}} \frac{\mathcal{X}^2_{pq}(\delta^2)}{p},$$

where the $\mathcal{X}^2_{pq}$ *random variable has* qp *degrees of freedom and noncentrality parameter defined as* $\delta^2 = \mathrm{tr}(B'P_{22}^{-1}BP_{11}^{-1})$*.*

The proof of this Theorem is analogous to Theorem 3.1, but a noncentrality parameter is introduced in the asymptotic distribution of $n\mathrm{RV}^{(s)}$, $n\mathrm{SL}^{(s)}$ and $n\mathrm{CN}^{(s)}$.

4. Tests of independence of two vectors

The results of the preceding section can be used to construct asymptotic tests of independence between two vectors. We will test for M = RV, SL or CN and $s = \tau$ or ϱ, H_0: $\rho M^{(s)} = 0$, against $\rho M^{(s)} > 0$ at level α by rejecting H_0 if $nM^{(s)} > c_\alpha^{(s,M)}$ where $c_\alpha^{(s,M)}$ is the $100(1-\alpha)$th percentile of the corresponding distribution given in Theorem 3.1. Under $\mathrm{H}_{1:n}$, $nM^{(s)}$ converges in probability to $\rho M^{(s)}$ for M = RV, SL, CN and $s = \tau$, ϱ and thus the asymptotic power of each of these six tests converges to 1 when $n \to \infty$. Thus, each test is consistent.

The limiting distributions given in Theorem 3.1 are not easy to deal with and consequently, the percentiles will be computed by using Imhof's algorithm (Imhof, 1961). Moreover, in these distributions, P_{11}, P_{22}, Λ_{11} and Λ_{22} are usually unknown, we thus use instead the estimators $\mathcal{S}_{11}$, $\mathcal{S}_{22}$, K_{11} and K_{22}. Since the estimators are consistent, the asymptotic distributions remain unchanged.

Let us notice that the tests $nM^{(\tau)}$ (M = RV, SL and CN) based on the matrix of Kendall depend on the tests $M^{(\varrho)}$ based on the matrix of Spearman. For example, the asymptotic distribution of $n\mathrm{RV}^{(\tau)}$ and $n\mathrm{RV}^{(\varrho)}$ use the same eigenvalues resulting from the submatrices P_{11} and P_{22} of Spearman's matrix. They are asymptotically equivalent, up to a multiplicative coefficient which depends on Kendall matrix. In the case of total independence, this constant is $\frac{4}{9}$ and this is already mentioned by several authors (see, for example, Hájek and Šidák, 1967).

Description of the procedure

Given a sample of size n, $(X_1^{[1]}, X_1^{[2]})', \ldots, (X_n^{[1]}, X_n^{[2]})'$ where $X_i^{[1]}$: $p \times 1$ and $X_i^{[2]}$: $q \times 1$ for $i = 1, \ldots, n$.

Step 1: Compute K_{11}, K_{22}, K_{12}, K_{21} and $\mathcal{S}_{11}$, $\mathcal{S}_{22}$, $\mathcal{S}_{12}$, $\mathcal{S}_{21}$.

Step 2: Compute the required eigenvalues from the consistent estimators.

Step 3: Compute $nM^{(s)}$ for $M =$ RV, SL, CN and $s = \tau$, ϱ.

Step 4: For each distribution given by Theorem 3.1, obtain the $100(1-\alpha)$th percentile, $c_\alpha^{(s,M)}$, for $M =$ RV, SL, CN and $s = \tau$, ϱ, by using the Imhof (1961) algorithm.

Step 5: Reject H_0 at level α if $\rho M^{(s)} > c_\alpha^{(s,M)}$, for $M =$ RV, SL, CN and $s = \tau$, ϱ.

Example. The six tests are illustrated with sport data. The data consist of the 1984 Olympic track records of 55 nations for women as well as men (see Naik and Khattree, 1996). The data matrix for women is a 55×7 matrix with seven events represented: the 100 meters, 200 meters, 400 meters, 800 meters, 1500 meters, 3000 meters and marathon (which is 42195 meters). For the men the corresponding matrix is of order 55×8 differing from the women's events in that the 3000 meters was excluded but 5000 meters and 10000 meters were included.

As noted by Naik and Khattree (1996), to test athletic performances of women and men, the appropriate variable that may be more relevant in this context is the speed, defined as the "distance covered per unit of time." This variable succeeds in retaining the possibility of having different degrees of variability. We will therefore use the speed in the track events as the variable for the tests of independence between women and men performances. These two data sets are presented in Tables 1 and 2 of Naik and Khattree (1996).

First, we test the hypothesis H_0 of independence between $X^{[1]}$ and $X^{[2]}$ where $X^{[1]}$ is the vector formed by women performances and $X^{[2]}$ is the vector formed by men performances. We have $n = 55$, $p = 7$ and $q = 8$. Table 1 gives the value of the statistic, the 5% critical value and the observed critical value. Therefore, H_0 is strongly rejected.

5. Simulation study

In order to assess the behavior of the tests based on Kendall's matrix, a Monte-Carlo study is performed to compare its empirical level and its empirical power with those of the three competitors based on Spearman's matrix (see Cléroux, Lazraq and Lepage, 1995).

Table 1. Tests of independence between women and men performances, the value of the statistic, the 5% critical value and the observed critical value.

Matrix	Statistic	Value	Critical point $C_{0,05}$	Critical level
Spearman	$n\mathrm{RV}^{(\varrho)}$	44.88	4.33	1.19×10^{-7}
	$n\mathrm{SL}^{(\varrho)}$	42.14	14.32	1.19×10^{-7}
	$n\mathrm{CN}^{(\varrho)}$	17.18	10.63	1.25×10^{-6}
Kendall	$n\mathrm{RV}^{(\tau)}$	37.16	2.77	0
	$n\mathrm{SL}^{(\tau)}$	28.67	3.95	1.19×10^{-7}
	$n\mathrm{CN}^{(\tau)}$	8.21	1.57	1.19×10^{-7}

All the simulation programs were written in FORTRAN programming language. For ease of comparison, the study is restricted to the case $p = 2$, $q = 3$ and the nominal level 1%. The number of repetitions at each setting is 10 000. Two types of underlying distributions are imposed. In the family of elliptic distributions, we consider a multivariate distribution $\mathcal{N}_5(O, \Sigma)$ and an elliptic multivariate t_5. In the family of nonelliptic distributions, we consider a multivariate logistic U (see Johnson, 1987) and a general multivariate distribution constructed as follows: each component of the vector X is independently generated from the other, the first is $\mathcal{N}_1(O, 1)$, the second is uniform on $[0, 1]$ minus 0.5 and multiplied by $\sqrt{12}$, the third is an exponential distribution (with parameter 1) minus 1, the fourth is a beta (with parameters 2 and 2) minus 0.5 and multiplied by $\sqrt{20}$ and finally the fifth is a gamma distribution (with parameters 1 and 4) minus 4 and divided by 2.

Under H_0, we generate two independent random vectors $X^{[1]}$ and $X^{[2]}$. For the alternative hypothesis, we consider the linear transformation $Y = CX$ where C is such that $\Sigma = CC'$. The matrices considered here are

$$\Sigma_{11} = I_2, \quad \Sigma_{22} = I_3, \quad \Sigma_{12} = \Sigma_{21}' = C_{00}, C_{10}, C_{15} \text{ and } C_{20}$$

where the matrices C_{xy} represent 2×3 matrices with all elements being the real number $0.xy$; for example, all elements of C_{15} are equal to 0.15. This type of matrices was used and justified by Cléroux, Lazraq and Lepage (1995).

Table 2 summarizes the simulation results for the five distributions. In order to judge the empirical level of the asymptotic tests and their empirical power, an empirical level will be good if the nominal level 1% belongs to the 95% confidence interval. So that, for 10 000 repetitions, C_{00} column must vary between 79 and 121.

The first observation is that for Kendall's tests and Spearman's tests, the empirical power of each test increases with departure from the null hypothesis that is when the value xy of the matrices C_{xy} increases. The empirical levels of $nM^{(\varrho)}$ are in general slightly conservative while that of

Table 2: Empirical power (×10000) of the tests based on Spearman's matrix and Kendall's matrix at nominal level 1% for the multivariate distributions with $p = 2$ and $q = 3$.

			Multivariate normal				t_5				Multivariate				Logistic			
n	matrices	Tests	C_{00}	C_{10}	C_{15}	C_{20}	C_{00}	C_{10}	C_{15}	C_{20}	C_{00}	C_{10}	C_{15}	C_{20}	C_{00}	C_{10}	C_{15}	C_{20}
50	Spearman	$n\mathrm{RV}^{(\varrho)}$	91	530	1395	2596	91	627	1691	3050	96	440	1066	1995	91	2432	5738	8254
		$n\mathrm{SL}^{(\varrho)}$	82	543	1587	3603	98	621	1913	4056	90	455	1326	3007	105	1334	3729	6672
		$n\mathrm{CN}^{(\varrho)}$	89	504	1424	3137	96	568	1650	3521	88	430	1187	2584	115	846	2520	5119
	Kendall	$n\mathrm{RV}^{(\tau)}$	109	589	1489	2743	114	697	1842	3231	116	494	1162	2139	104	2561	5903	8367
		$n\mathrm{SL}^{(\tau)}$	107	590	1647	3415	119	713	2031	3947	115	527	1358	2801	120	2111	5146	7800
		$n\mathrm{CN}^{(\tau)}$	109	554	1552	3187	121	693	1913	3704	117	506	1288	2588	121	1908	4803	7476
100	Spearman	$n\mathrm{RV}^{(\varrho)}$	92	1453	4120	6994	88	1799	4802	7428	100	1168	3325	5710	108	5564	9151	9931
		$n\mathrm{SL}^{(\varrho)}$	100	1534	4745	8232	100	1877	5465	8539	103	1299	4049	7405	99	3673	8090	9751
		$n\mathrm{CN}^{(\varrho)}$	94	1467	4459	7931	97	1766	5151	8210	97	1241	3768	7062	100	2578	6857	9384
	Kendall	$n\mathrm{RV}^{(\tau)}$	98	1513	4195	7065	97	1899	4934	7541	107	1235	3387	5802	106	5632	9185	9939
		$n\mathrm{SL}^{(\tau)}$	104	1573	4629	7950	111	1964	5426	8315	105	1303	3899	6991	112	4895	8826	9879
		$n\mathrm{CN}^{(\tau)}$	105	1534	4490	7746	108	1888	5216	8127	106	1281	3739	6752	108	4488	8561	9830
200	Spearman	$n\mathrm{RV}^{(\varrho)}$	94	4086	8530	9845	103	4806	8973	9892	102	3492	7746	9616	100	9042	9986	10000
		$n\mathrm{SL}^{(\varrho)}$	98	4333	8964	9956	107	5076	9343	9969	100	3758	8476	9905	105	7880	9940	9999
		$n\mathrm{CN}^{(\varrho)}$	100	4212	8889	9949	102	4931	9229	9965	106	3657	8317	9892	101	6719	9816	9999
	Kendall	$n\mathrm{RV}^{(\tau)}$	103	4118	8549	9848	113	4905	8974	9903	103	3539	7795	9635	104	9056	9986	10000
		$n\mathrm{SL}^{(\tau)}$	104	4301	8871	9939	107	5089	9270	9964	104	3735	8300	9860	100	8651	9975	99999
		$n\mathrm{CN}^{(\tau)}$	102	4232	8809	9935	112	5004	9213	9955	103	3681	8219	9849	111	8353	9963	9999
300	Spearman	$n\mathrm{RV}^{(\varrho)}$	103	6679	9785	9996	107	7436	9910	10000	98	5961	9565	9990	110	9859	9999	10000
		$n\mathrm{SL}^{(\varrho)}$	115	6939	9874	9997	103	7698	9945	9998	99	6350	9793	9994	101	9489	9999	10000
		$n\mathrm{CN}^{(\varrho)}$	115	6837	9862	10000	97	7604	9946	10000	98	6250	9762	10000	91	8954	9996	10000
	Kendall	$n\mathrm{RV}^{(\tau)}$	101	6704	9785	9996	108	7488	9909	10000	102	6015	9568	9991	110	9864	9999	10000
		$n\mathrm{SL}^{(\tau)}$	110	6897	9844	9999	106	7666	9945	10000	101	6280	9746	10000	108	9736	9999	10000
		$n\mathrm{CN}^{(\tau)}$	115	6841	9840	9999	105	7603	9942	10000	101	6210	9719	10000	103	9635	9999	10000

$nM^{(\tau)}$ (M = RV, SL or CN) are liberal. The tests $nM^{(\varrho)}$ have an empirical power slightly inferior to $nM^{(\tau)}$ (M = RV, SL or CN). In each class of tests (Kendall or Spearman), we notice that the empirical power of the tests $n\mathrm{SL}^{(s)}$ ($s = \tau$ or ϱ) is greater than the empirical power of the other tests, but when the underlying distribution is logistic, the empirical power of $n\mathrm{RV}^{(s)}$ ($s = \tau$ or ϱ) is greater than the two others. In conclusion, the empirical power of each test, in a given class, depends on the underlying distribution. Nevertheless, one notices that the tests of Kendall's class are empirically more powerful than the tests of Spearman's class especially for small sample sizes and in the vicinity of the null hypothesis C_{00}.

REFERENCES

Baldessari, B. (1967). The distribution of quadratic forms of normal random variables. *Ann. Math. Statist* 38, 1700–1704.

Cléroux, R., Lazraq, A. and Lepage, Y. (1995). Vector correlation based on rank and a nonparametric test of no association between vectors. *Comm. Statist. Theory Methods* 24, 713–733.

Cramer, E.M. and Nicewander, G.R. (1979). Some symmetric invariant measure of multivariate association. *Psychometrika* 49, 403–423.

Escoufier, Y. (1973). Le traitement des variables vectorielles. *Biometrics* 29, 751–760.

Hájek, J. and Šidák, Z. (1967). *Theory of Rank Tests.* Academic Press, New York.

Hays, W.L. (1960). Note on Average *tau* as a Measures of Concordance. *J. Amer. Statist. Assoc.* 55, 331–341.

Hoeffding, W. (1948). A class of statistics with asymptotically normal distribution. *Ann. Math. Statistics* 19, 293–325.

Imhof, P. (1961). Computing the distribution of quadratic forms in normal variates. *Biometrika* **48**, 419–426.

Joe, H. (1990). Multivariate concordance. *J. Multivariate Anal.* 35, 12–30.

Johnson, M.E. (1987). *Multivariate Statistical Simulation.* John Wiley, New York.

Johnson, N.L. and Kotz, S. (1970). *Distributions in Statistics. Continuous Univariate Distributions.* 2. Houghton Mifflin Co., Boston.

Lazraq, A. and Cléroux, R. (1988). Étude comparative de différentes mesures de liaison entre deux vecteurs aléatoires et tests d'indépendance. *Statist. Anal. Données* 13, 15–18.

Lazraq, A., Lepage, Y. and Cléroux, R. (1995). Tests non paramétriques pour l'indépendance entre plusieurs vecteurs aléatoires. *Publ. Inst. Statist. Univ. Paris* 39, 57–77.

Naik, D.N. and Khattree, R. (1996). Revisiting Olympic track records: somme practical considerations in the principal component analysis. *Amer. Statist.* 50, 140–144.

Puri, M.L., Sen, P.K. and Gokhale, D.V. (1970). On a class of rank order tests for independence in multivariate distributions. *Sankhyā Ser. A* 32, 271–298.

Serfling, R.J. (1980). *Approximation Theorems of Mathematical Statistics.* John Wiley, New York.

Simon, G. (1977). A nonparametric test of total independence based on Kendall's tau. *Biometrika*, 64, 277–282.

Stewart, D. and Love, W. (1968). A general canonical correlation index. *Psychological Bulletin* 70, 160–163.

Hamani El Maache
Dép. de mathématiques et de statistique
Université de Montréal
C.P. 6128, succ. Centre-Ville
Montréal QC H3C 3J7
Canada
elmaach@dms.umontreal.ca

Yves Lepage
Dép. de mathématiques et de statistique
Université de Montréal
C.P. 6128, succ. Centre-Ville
Montréal QC H3C 3J7
Canada
lepage@dms.umontreal.ca