ZASTOSOWANIA MATEMATYKI
APPLICATIONES MATHEMATICAE
20,2 (1990), pp. 191–202

W. WYSOCKI (Warszawa)

A GEOMETRIC APPROACH TO MEASURING DEPENDENCE BETWEEN RANDOM VECTORS

0. Introduction. Measurement of stochastic dependence between random vectors is one of the central problems of mathematical statistics. Usually, we are rather dealing with dependence between real-valued random variables, say X and Y. A popular measure of the dependence of X on Y is the correlation coefficient

$$\varrho_0(X, Y) = \operatorname{cov}(X, Y)/\sigma(X)\sigma(Y).$$

Geometrically, $\varrho(X, Y)$ can be interpreted in $(L^2(\Omega, \mathscr{F}, P), \langle,\rangle)$ as the cosine of the angle between centered random variables X and Y.

According to me, the most important generalizations of ϱ_0 for random vectors X and Y were done by Höschel [1] and by Jupp and Mardia [2].

Höschel proposed a system of postulates, introduced a measure of dependence satisfying this system, and showed that it is unique. However, Höschel's measure is complicated from the numerical point of view and is not commonly used.

Jupp and Mardia proposed a measure $\varrho_0'(X, Y)$ which directly generalizes the definition of ϱ_0:

$$(\varrho_0'(X, Y))^2 = \operatorname{tr}(\Sigma_{11}^{-1}\Sigma_{12}\Sigma_{22}\Sigma_{12}^T),$$

where Σ_{11}, Σ_{22}, Σ_{12} are covariance matrices corresponding to $\sigma^2(X)$, $\sigma^2(Y)$, $\operatorname{cov}(X, Y)$, respectively, in the bivariate case. Thus, $\varrho_0'(X, Y)$ "normalizes" the covariance matrix Σ_{12}; it takes values from $[0, \min(p, q)]$, where p and q are dimensions of X and Y, respectively.

Other generalizations of ϱ_0 for random vectors X and Y were reviewed by Jupp and Mardia (though this review did not contain Höschel's measure). In particular, they mentioned the first canonical correlation, equal to the maximal correlation coefficient between linear combinations of the components of X and Y. This measure is most commonly used by statisticians, mainly because it is simple to be calculated.

As far as we know, geometric interpretations were not given for any measures discussed by Jupp and Mardia.

The present paper introduces the correlation coefficient $\varrho(X, Y)$ for random vectors X and Y, which measures the proximity of the linear spaces generated by the components of X and Y, respectively. The geometric reflexions given in Section 1 are the starting point for the definition of ϱ, leading to a volume measure of the angle between unitary subspaces. Theorem 2 in Section 1 is the most important theorem of the paper. The correlation coefficient ϱ is proposed in Section 2 as the cosine of the angle between the subspaces spanned by the centered components of X and Y, respectively. The properties of ϱ are discussed, and the relations between ϱ and the first canonical correlation are presented. In Section 3 an estimator $\hat{\varrho}_n$ of ϱ is introduced and some asymptotic properties of $\hat{\varrho}_n$ under the assumption of stochastic independence of all components of (X, Y) are proved.

Summing up, we want to stress the difference between statistical inspirations, exemplified by the generalization of ϱ_0 proposed by Jupp and Mardia, and the geometrical inspirations, exemplified by ϱ proposed in this paper.

1. The cosine of an angle between finite-dimensional unitary subspaces. Let $(\boldsymbol{E}, \langle,\rangle)$ be a real unitary space and let $\boldsymbol{A}$ and $\boldsymbol{B}$ be finite-dimensional subspaces of $\boldsymbol{E}$ with $\dim \boldsymbol{A} = p$ and $\dim \boldsymbol{B} = q$. Let

$$\boldsymbol{C} = \boldsymbol{A} \cap \boldsymbol{B}, \quad \boldsymbol{E}_0 = \boldsymbol{A} + \boldsymbol{B}$$

with $\dim \boldsymbol{C} = p_0$, $\dim \boldsymbol{E}_0 = n = p+q-p_0$.

Assume that $\boldsymbol{C} \neq \boldsymbol{A}$ and $\boldsymbol{C} \neq \boldsymbol{B}$. Then $p_0 < \min(p, q)$.

Let $\boldsymbol{A}^{\perp}$ and $\boldsymbol{B}^{\perp}$ denote orthogonal complements of $\boldsymbol{A}$ and $\boldsymbol{B}$ in $\boldsymbol{E}_0$, and let $\boldsymbol{A}_0$ and $\boldsymbol{B}_0$ denote orthogonal complements of $\boldsymbol{C}$ in $\boldsymbol{A}$ and $\boldsymbol{B}$. Further, let

$$\begin{aligned} &(a_1, a_2, \ldots, a_{p-p_0}), \quad (b_1, b_2, \ldots, b_{q-p_0}), \\ &(c_1, c_2, \ldots, c_{p-p_0}), \quad (d_1, d_2, \ldots, d_{q-q_0}) \end{aligned} \tag{1}$$

denote the bases of $\boldsymbol{A}_0$, $\boldsymbol{B}_0$, $\boldsymbol{B}^{\perp}$, $\boldsymbol{A}^{\perp}$, and let U_0, $\tilde{U}_0$, W_1, W_2, W_3, W_4 be the following matrices:

$$\begin{aligned} U_0 &= [\langle a_i, c_j \rangle], & i,j &= 1, 2, \ldots, p-p_0, \\ \tilde{U}_0 &= [\langle b_i, d_j \rangle], & i,j &= 1, 2, \ldots, q-p_0, \\ W_1 &= [\langle a_i, a_j \rangle], & i,j &= 1, 2, \ldots, p-p_0, \\ W_2 &= [\langle c_i, c_j \rangle], & i,j &= 1, 2, \ldots, p-p_0, \\ W_3 &= [\langle b_i, b_j \rangle], & i,j &= 1, 2, \ldots, q-p_0, \\ W_4 &= [\langle d_i, d_j \rangle], & i,j &= 1, 2, \ldots, q-p_0. \end{aligned}$$

THEOREM 1. *The expressions*

$$|\det U_0|/(\det W_1 \cdot \det W_2)^{1/2}, \quad |\det \tilde{U}_0|/(\det W_3 \cdot \det W_4)^{1/2}$$

have the following properties:

(i) *they are invariant on the bases of* A_0, $B^{\perp}$ *and* B_0, $A^{\perp}$, *respectively;*

(ii) $$0 < \frac{|\det U_0|}{(\det W_1 \cdot \det W_2)^{1/2}} = \frac{|\det \tilde{U}_0|}{(\det W_3 \cdot \det W_4)^{1/2}} \leqslant 1.$$

Proof. (i) Let $(a'_1, a'_2, \ldots, a'_{p-p_0})$ and $(c'_1, c'_2, \ldots, c'_{p-p_0})$ be other bases of A_0 and $B^{\perp}$, and let P_1 and P_2 be matrices of transformations from non-primed to primed bases. It is known that

$$W'_1 = P_1^T W_1 P_1, \qquad W'_2 = P_2^T W_2 P_2,$$

where

$$W'_1 = [\langle a'_i, a'_j \rangle], \qquad W'_2 = [\langle c'_i, c'_j \rangle].$$

Then

(2) $$\det W'_1 = (\det P_1)^2 \det W_1,$$

(3) $$\det W'_2 = (\det P_2)^2 \det W_2.$$

Similarly, putting $U'_0 = [\langle a'_i, c'_j \rangle]$, we have

$$U'_0 = P_1^T U_0 P_2.$$

From the last equality divided by $(\det W_1 \cdot \det W_2)^{1/2}$, in view of (2) and (3), we obtain

$$\frac{|\det U'_0|}{(\det W'_1 \cdot \det W'_2)^{1/2}} = \frac{|\det U_0|}{(\det W_1 \cdot \det W_2)^{1/2}}.$$

The proof in the case of B_0 and $A^{\perp}$ is analogous.

(ii) In view of (i), we can restrict ourselves to the case where the vectors (1) are orthonormal. Let

(4) $$(a_{p-p_0+1}, a_{p-p_0+2}, \ldots, a_p)$$

be an orthonormal basis of C, and let

$$(a_1, a_2, \ldots, a_{p-p_0}) \quad \text{and} \quad (a_{p+1}, a_{p+2}, \ldots, a_n)$$

be complements of (4) for orthonormal subspaces A and B, respectively. Let

(5) $$(c_1, c_2, \ldots, c_{p-p_0}, a_{p-p_0+1}, \ldots, a_n)$$

be an orthonormal basis of E_0. The system of vectors

(6) $$(a_1, a_2, \ldots, a_n)$$

is linearly independent, and hence is a basis of E_0. The matrix of coefficients of the vectors (6) on the basis (5) is of the form

$$U = \begin{bmatrix} U_0^T & 0 \\ U_1 & I_q \end{bmatrix},$$

and therefore $\det U = \det U_0$. Since (6) is a basis of E_0, the Gramian of (6) satisfies the inequality

(7) $$W(a_1, a_2, \ldots, a_n) = (\det U_0)^2 = (\det U)^2 > 0.$$

Putting $(b_1, \ldots, b_{q-p_0}) = (a_{p+1}, \ldots, a_n)$, we deduce similarly that

$$\tilde{U} = \begin{bmatrix} I_q & \tilde{U}_1 \\ 0 & \tilde{U}_0^T \end{bmatrix},$$

where $\tilde{U}$ corresponds to the presentation of (6) in the orthonormal basis $(a_1, \ldots, a_p, d_1, \ldots, d_{q-p_0})$, and $\tilde{U}_1$ is some matrix defined by the context. Consequently, $\det U_0 = \det \tilde{U}$, and the corresponding Gramian satisfies

$$W(a_1, \ldots, a_n) = (\det \tilde{U}_0)^2 = (\det \tilde{U})^2 > 0.$$

This fact and (7) imply that $|\det U_0| = |\det \tilde{U}_0| > 0$. Since $|\det U_0|$, equal to $|\det \tilde{U}_0|$, is the square root of a Gramian of a linearly independent system of n normed vectors, we obtain

$$|\det U_0| = |\det \tilde{U}_0| \leqslant 1.$$

Definition 1. For any unitarian subspaces A and B, the expression

$$\cos \sphericalangle (A, B) = \begin{cases} (1-(\det U_0)^2)^{1/2} & \text{if } p_0 < \min(p, q), \\ 1 & \text{if } p_0 = \min(p, q) \end{cases}$$

is called the *cosine* of the angle between A and B, and $\arccos(\cos \sphericalangle (A, B))$ is called the *volume* of the angle between A and B and denoted by $\operatorname{vol} \sphericalangle (A, B)$.

This definition covers the previously excluded case in which $p_0 = \min(p, q)$, i.e., $A \subset B$ or $B \subset A$. It is evident that

$$0 \leqslant \operatorname{vol} \sphericalangle (A, B) \leqslant \pi/2,$$

$$\operatorname{vol} \sphericalangle (A, B) = \operatorname{vol} \sphericalangle (B, A) = \operatorname{vol} \sphericalangle (A_0, B_0).$$

Remark 1. Unitarian subspaces A and B are parallel iff

$$\operatorname{vol} \sphericalangle (A, B) = 0.$$

The orthogonality of A and B implies that

$$\operatorname{vol} \sphericalangle (A, B) = \pi/2;$$

on the other hand, the equality $\operatorname{vol} \sphericalangle (A, B) = \pi/2$ implies only that A_0 and B_0 are orthogonal.

In the sequel we show another way of deriving $|\det U_0|$.

Let

$$(x_1, x_2, \ldots, x_p) \quad \text{and} \quad (y_1, y_2, \ldots, y_q)$$

be any bases of A and B, respectively. We define

$$\Sigma_{11} = [\langle x_i, x_j \rangle], \quad i, j = 1, 2, \ldots, p,$$
$$\Sigma_{22} = [\langle y_i, y_j \rangle], \quad i, j = 1, 2, \ldots, q,$$
$$\Sigma_{12} = [\langle x_i, y_j \rangle], \quad i = 1, 2, \ldots, p, j = 1, 2, \ldots, q.$$

Let $(\varrho_i^2, i = 1, \ldots, p)$ and $(\tilde{\varrho}_j, j = 1, \ldots, q)$ denote the eigenvalues of

$$\Sigma = \Sigma_{11}^{-1}\Sigma_{12}\Sigma_{22}^{-1}\Sigma_{12}^T \quad \text{and} \quad \tilde{\Sigma} = \Sigma_{22}^{-1}\Sigma_{12}^T\Sigma_{11}^{-1}\Sigma_{12},$$

respectively.

THEOREM 2. *The eigenvalues of Σ and $\tilde{\Sigma}$ which are different from* 1 *satisfy the equalities*

$$(\det U_0)^2 = \prod_{i=1}^{p-p_0} (1-\varrho_i^2), \quad (\det U_0)^2 = \prod_{j=1}^{q-p_0} (1-\tilde{\varrho}_j^2).$$

Proof. We show first how we come to consider Σ and $\tilde{\Sigma}$.

Let h be a functional on $(\boldsymbol{H}, \|\cdot\|_0)$,

$$\boldsymbol{H} = \boldsymbol{A} \times \boldsymbol{B}, \quad \|(x, y)\|_0 = \|x\| + \|y\|, \quad \|x\| = (\langle x, x \rangle)^{1/2},$$

such that

$$h((x, y)) = \langle x, y \rangle.$$

We look for local conditional maxima of h on the compact set

$$\mathscr{H} = \{(x, y): \langle x, x \rangle = \langle y, y \rangle = 1\}.$$

Thus we get two systems of equations:

$$(\Sigma - \varrho^2 I_p)\bar{x}^T = \boldsymbol{0}^T, \quad \bar{x} I_p \bar{x}^T = 1, \quad \varrho \neq 0,$$
$$(\tilde{\Sigma} - \varrho^2 I_q)\bar{y}^T = \boldsymbol{0}^T, \quad \bar{y} I_q \bar{y}^T = 1, \quad \varrho \neq 0,$$

where $\bar{x}$ and $\bar{y}$ are the counterparts of vectors x and y if $\boldsymbol{A}$ and $\boldsymbol{B}$ are replaced by R^p and R^q, respectively. Non-trivial solutions appear when the matrices of the systems are singular. The eigenvalues of Σ and $\tilde{\Sigma}$ satisfy

$$0 < \varrho_{p-q_0+1}^2 \leqslant \ldots \leqslant \varrho_p^2 \leqslant 1,$$
$$0 < \tilde{\varrho}_{q-q_0+1}^2 \leqslant \ldots \leqslant \tilde{\varrho}_{q-q_0+j}^2 \leqslant \ldots \leqslant \tilde{\varrho}_q^2 \leqslant 1,$$

where $q_0 = \operatorname{rank} \Sigma_{12}$.

It is easy to show that Σ is equivalent to Σ', and $\tilde{\Sigma}$ to $\tilde{\Sigma}'$ (i.e., they have the same eigenvalues), where Σ' and $\tilde{\Sigma}'$ correspond to the new bases of $\boldsymbol{A}$ and $\boldsymbol{B}$. Therefore, in this proof we restrict ourselves to orthonormal bases. Then

$$\Sigma = \Sigma_{12}\Sigma_{12}^T \quad \text{and} \quad \tilde{\Sigma} = \Sigma_{12}^T\Sigma_{12}.$$

Let $q_0 = \operatorname{rank} \Sigma_{12}$. It is known that each of the matrices $\Sigma_{12}\Sigma_{12}^T$ and $\Sigma_{12}^T\Sigma_{12}$has q_0 positive eigenvalues and both sequences of positive eigenvalues are identical:

$$\varrho_{p-q_0+i}^2 = \tilde{\varrho}_{q-q_0+i}^2, \quad i = 1, 2, \ldots, q_0.$$

Let us choose the bases as indicated in the proof of (ii) in Theorem 1. Therefore

(8) $$(a_{p-p_0+1}, \ldots, a_p)$$

is an orthonormal basis of $\boldsymbol{C}$, and

(9) $$(a_1, \ldots, a_{p-p_0}), \quad (a_{p+1}, \ldots, a_n)$$

are the complements of (8) to the orthonormal bases of $\boldsymbol{A}$ and $\boldsymbol{B}$, respectively. Let A_0, B_0 and C be the matrices of coordinates of the vectors (9) and (8) in the orthonormal basis

(10) $$(e_1, \ldots, e_n)$$

of $\boldsymbol{E}_0$. The Gramm matrix W of the system

$$(a_1, \ldots, a_{p-p_0}, \ldots, a_n)$$

in the basis (10) can be expressed as follows:

$$W = \begin{bmatrix} I_{p-p_0} & 0 & A_0^T B_0 \\ 0 & I_p & 0 \\ B_0^T A_0 & 0 & I_{q-p_0} \end{bmatrix}.$$

Let

$$Z = \begin{bmatrix} I_{p-p_0} & 0 & -A_0^T B_0 \\ 0 & I_{p_0} & 0 \\ 0 & 0 & I_{p-q_0} \end{bmatrix}.$$

Then

$$ZW = \begin{bmatrix} I_{p-p_0} - A_0^T B_0 B_0^T A_0 & 0 & 0 \\ 0 & I_{p_0} & 0 \\ B_0^T A_0 & 0 & I_{q-p_0} \end{bmatrix}.$$

Hence

$$\det ZW = \det Z \det W = \det W = \det(I_{p-p_0} - A_0^T B_0 B_0^T A_0) = \prod_{i=1}^{p-p_0} (1-\varrho_i^2),$$

where $\varrho_i^2 (i = 1, 2, \ldots, p-p_0)$ are the eigenvalues of the non-negative definite matrix $A_0^T B_0 B_0^T A_0$. It follows that

(11) $$(\det U_0)^2 = \prod_{i=1}^{p-p_0} (1-\varrho_i^2).$$

It remains to show that the sequence of eigenvalues of Σ being non-decreasingly ordered and different from 1 is equal to such a sequence of $A_0^T B_0 B_0^T A_0$. Let

$$Z_1 = [A_0 \quad C] \quad \text{and} \quad Z_2 = [C \quad B_0].$$

Then

$$\Sigma = Z_1^T Z_2 = \begin{bmatrix} A_0^T B_0 B_0^T A_0 & \mathbf{0} \\ \mathbf{0} & I_{p_0} \end{bmatrix}.$$

Consequently, the family of eigenvalues of Σ consists of the family of eigenvalues of $A_0^T B_0 B_0^T A_0$ and of $p_0 = \dim \boldsymbol{C}$ eigenvalues equal to 1. Thus, in view of (11), the proof is complete.

EXAMPLE. Let $(\boldsymbol{E}, \langle\ \rangle)$ be an arbitrary real unitary space with the orthogonal system of vectors $(e_1, e_2, e_3, e_4, e_5)$. Basing on Theorem 2, we calculate vol $\sphericalangle (\boldsymbol{A}, \boldsymbol{B})$ between unitary subspaces

$$\boldsymbol{A} = L\{e_1+e_3,\ e_3+e_4+e_5\} \quad \text{and} \quad \boldsymbol{B} = L\{e_1,\ e_1+e_3+e_5,\ e_2+e_4\}.$$

We have

$$\Sigma_{11} = \begin{bmatrix} 2 & 1 \\ 1 & 3 \end{bmatrix}, \quad \Sigma_{11}^{-1} = \begin{bmatrix} 0.6 & -0.2 \\ -0.2 & 0.4 \end{bmatrix},$$

$$\Sigma_{22} = \begin{bmatrix} 1 & 1 & 0 \\ 1 & 3 & 0 \\ 0 & 0 & 2 \end{bmatrix}, \quad \Sigma_{22}^{-1} = \begin{bmatrix} 1.5 & -0.5 & 0 \\ -0.5 & 0.5 & 0 \\ 0 & 0 & 0.5 \end{bmatrix},$$

$$\Sigma_{12} = \begin{bmatrix} 1 & 2 & 0 \\ 0 & 2 & 1 \end{bmatrix}.$$

Then

$$\Sigma = \begin{bmatrix} 0.7 & 0.1 \\ 0.1 & 0.8 \end{bmatrix}.$$

The eigenvalues of Σ satisfy the equation

$$(\varrho^2 - 0.7)(\varrho^2 - 0.8) - 0.1^2 = 0.$$

Consequently,

$$\varrho_1^2 = 0.75 - 0.05\sqrt{5}, \qquad \varrho_2^2 = 0.75 + 0.05\sqrt{5},$$

$$\cos^2(\text{vol} \sphericalangle (\boldsymbol{A}, \boldsymbol{B})) = 1 - (1-\varrho_1^2)(1-\varrho_2^2) = 0.95,$$

$$\text{vol} \sphericalangle (\boldsymbol{A}, \boldsymbol{B}) = \arccos\sqrt{0.95}.$$

2. A correlation coefficient between random vectors. In this section we define a measure of dependence $\varrho(X, Y)$ between random vectors X and Y, which corresponds to Definition 1. The introduced measure can be geometrically interpreted in $\big(L^2(\Omega, \mathscr{F}, P), \langle\ \rangle\big)$ as the cosine of the angle between the linear subspaces spanned over the centered coordinates of X and Y, respectively. It can be also interpreted as a measure of "mutual closeness" of these subspaces.

We assume that the covariance matrices of X and Y exist. Therefore, as a natural real unitary space E we take $(L^2(\Omega, \mathcal{F}, P), \langle\ \rangle)$ with the inner product defined by

$$\langle \dot{Z}_1, \dot{Z}_2 \rangle = \mathrm{E}[Z_1 Z_2],$$

where $\dot{Z}_i$ $(i = 1, 2)$ is the class of abstraction represented by the random variable Z_i $(i = 1, 2)$.

We start with geometrical interpretations in the case of the correlation coefficient for random variables and of the multiple correlation. The first of these two measures of dependence is defined for any $Z_1, Z_2 \in L^2(\Omega, \mathcal{F}, P)$ by

$$\varrho_0(Z_1, Z_2) = \frac{\mathrm{cov}(Z_1, Z_2)}{\sigma(Z_1)\sigma(Z_2)} = \frac{\mathrm{E}(Z_1 - \mathrm{E}Z_1)(Z_2 - \mathrm{E}Z_2)}{\sigma(Z_1)\sigma(Z_2)},$$

where the operations cov and σ define the covariance of a pair of random variables and the dispersion of a random variable. The multiple correlation coefficient between a random variable Z_0 and a q-dimensional random vector $Y = (Y_1, \ldots, Y_q)$ for Z_0, $Y_j \in L^2(\Omega, \mathcal{F}, P)$ is defined by

$$\varrho_1(Z_0, Y) = \sup\{\varrho_0(Z_0, Z)\colon Z \in L(Y_1, \ldots, Y_q) \subset L^2(\Omega, \mathcal{F}, P)\}.$$

Then $\varrho_0(Z_1, Z_2)$ is interpreted in $(L^2(\Omega, \mathcal{F}, P), \langle\ \rangle)$ as the cosine of the angle between the random variables $Z_i - \mathrm{E}Z_i$ $(i = 1, 2)$ while $\varrho_1(Z_0, Y)$ is interpreted as the cosine of the angle between the random variable $Z_0 - \mathrm{E}Z_0$ and the linear space spanned on $Y_j - \mathrm{E}Y_j$ $(j = 1, \ldots, q)$.

Now, let $X = (X_1, \ldots, X_p)$ and $Y = (Y_1, \ldots, Y_q)$ be random vectors (p- and q-dimensional, respectively) defined on $(\Omega, \mathcal{F}, P)$, with

$$\dim L(X_1, \ldots, X_p) = p, \qquad \dim L(Y_1, \ldots, Y_q) = q,$$

$X_i, Y_j \in L^2(\Omega, \mathcal{F}, P)$, $i = 1, \ldots, p$, $j = 1, \ldots, q$. Let

$$\begin{aligned}
\Sigma_{11} &= [\mathrm{cov}(X_i, X_j)], \quad && i, j = 1, 2, \ldots, p,\\
\Sigma_{22} &= [\mathrm{cov}(Y_i, Y_j)], \quad && i, j = 1, 2, \ldots, q,\\
\Sigma_{12} &= [\mathrm{cov}(X_i, Y_j)], \quad && i = 1, \ldots, p,\ j = 1, \ldots, q,
\end{aligned}$$

$$\Sigma = \Sigma_{11}^{-1}\Sigma_{12}\Sigma_{22}^{-1}\Sigma_{12}^T.$$

We deal with the non-decreasingly ordered eigenvalues of Σ which are different from 1:

$$1 > \varrho_1^2 \geqslant \varrho_2^2 \geqslant \ldots \geqslant \varrho_{p-p_0}^2 \geqslant 0,$$

where $p_0 = \dim \mathcal{X} \cap \mathcal{Y}$, and

$$\mathcal{X} = L\{X_1 - \mathrm{E}X_1, \ldots, X_p - \mathrm{E}X_p\}, \quad \mathcal{Y} = L\{Y_1 - \mathrm{E}Y_1, \ldots, Y_q - \mathrm{E}Y_q\}.$$

Definition 2. The expression

$$\varrho(X, Y) = \begin{cases} 1 & \text{if all eigenvalues of } \Sigma \text{ are equal to } 1,\\ \left(1 - \prod_{i=1}^{p-p_0} (1 - \varrho_i^2)\right)^{1/2} & \text{otherwise} \end{cases}$$

is called the *correlation coefficient* between the random vectors X and Y.

Remark 2. In view of Definition 1 and Theorem 2, $\varrho(X, Y)$ can be interpreted in $(L^2(\Omega, \mathscr{F}, P), \langle\ \rangle)$ as the cosine of the angle between subspaces $\mathscr{X}$ and $\mathscr{Y}$.

Obviously,

$$0 \leqslant \varrho(X, Y) \leqslant 1.$$

For $p = 1$, $p = q$, $\varrho = \varrho_1$ so that Definition 2 generalizes the notion of the multiple correlation coefficient.

THEOREM 3. *Let X and Y be random vectors with singular covariance matrices. Then*

(i) *for any linear isomorphisms $u \in I(R^p)$, $v \in I(R^q)$,*

$$\varrho(u \circ X, v \circ Y) = \varrho(X, Y);$$

(ii) $\varrho(X, Y) = \varrho(Y, X)$;

(iii) *if X and Y are uncorrelated (i.e., $\Sigma_{12} = 0$), then $\varrho(X, Y) = 0$;*

(iv) *if $\varrho(X, Y) = 0$ and $p_0 = \dim(\mathscr{X} \cap \mathscr{Y}) = 0$, then X and Y are uncorrelated;*

(v) $\varrho(X, Y) = 1$ *iff there exists a linear transformation $u \in L(R^q, R^p)$ such that*

$$X - \mathrm{E}X = u\circ(Y - \mathrm{E}Y) \text{ a.e.}$$

Proof. (i) Put

$$X' = (X'_1, \ldots, X'_p) = u \circ X = AX^T, \qquad Y' = (Y'_1, \ldots, Y'_q) = v \circ Y = BY^T,$$

where A and B are the respective non-singular matrices, and let

$$\mathscr{X}' = L\{X'_1 - \mathrm{E}X'_1, \ldots, X'_p - \mathrm{E}X'_p\}, \qquad \mathscr{Y}' = L\{Y'_1 - \mathrm{E}Y'_1, \ldots, Y'_q - \mathrm{E}Y'_q\}.$$

Due to the geometrical interpretation of ϱ it suffices to prove that $\mathscr{X} = \mathscr{X}'$ and $\mathscr{Y} = \mathscr{Y}'$. Since A is non-singular, we have

$$\sum_{i=1}^{p} t_i X'_i = \sum_{i=1}^{p} t_i \Big(\sum_{j=1}^{p} a_{ij} X_j\Big) = \sum_{j=1}^{p} \Big(\sum_{i=1}^{p} a_{ij} t_i\Big) X_j = \sum_{j=1}^{p} t'_j X_j, \qquad t_i, t'_j \in R^1,$$

which implies $L\{X_1, \ldots, X_p\} = L\{X'_1, \ldots, X'_p\}$. The proof for Y is analogous.

The proofs of (ii) and (iii) are obvious, while (iv) is a consequence of Remark 1. To prove (v), note that $\varrho(X, Y) = 1$ iff $\mathscr{X}$ is parallel to $\mathscr{Y}$ or, equivalently, there exists a $(p \times q)$-matrix A such that

$$(X - \mathrm{E}X)^T = A(Y - \mathrm{E}Y)^T.$$

The next theorem states the relations between ϱ and the canonical correlations (as defined for instance in [3]).

THEOREM 4. (i) *Let $q_0 = \operatorname{rank}\Sigma_{12}$. For any $i = 1, \ldots, q_0 - p_0$, if the* *number* ϱ_i *introduced in Definition 2 is different from* 0, *then* ϱ_i *is the canonical correlation of order* p_0+i *between* X *and* Y.

(ii) *For any* $i = 1, \ldots, q_0-p_0$, *we have* $\varrho(X, Y) > \varrho_i$.

Proof. (i) follows from the derivation and interpretation of Σ and its spectrum presented in [3]; (ii) follows from the obvious inequality

$$\prod_{i=1}^{q_0-p_0} (1-\varrho_i^2) < 1-\varrho_i^2.$$

3. Estimation of ϱ. Let $X = (X_1, \ldots, X_p)$ and $Y = (Y_1, \ldots, Y_q)$ be defined as in Section 2, and let

$$(E_n, \varepsilon_n, Q_n) = \big(R^p \times R^q, \mathscr{B}(R^p)\otimes\mathscr{B}(R^q), P_{X,Y}\big)^n$$

be the sample space corresponding to n observations of the pair (X, Y). Let X_{ij} $(i = 1, \ldots, p, j = 1, \ldots, m)$ and Y_{ij} $(i = 1, \ldots, q, j = 1, \ldots, m)$ be statistics defined on the sample space, such that X_{ij} (Y_{ij}) is the i-th coordinate of the j-th observation of X (of Y). Let $T_1 = [X_{ij}]$, $T_2 = [Y_{ij}]$, and let

$$\bar{T}_1 = (\bar{X}_1, \ldots, \bar{X}_p)^T, \quad \bar{T}_2 = (\bar{Y}_1, \ldots, \bar{Y}_q)^T$$

be the respective vectors of sample means; X_{ij}, $\bar{X}_i$, Y_{ij}, $\bar{Y}_i$ belong to $L^2(E_n, \varepsilon_n, Q_n)$. We define the sample analogues of Σ_{11}, Σ_{22}, Σ_{12}:

$$S_{11}^{(n)} = \frac{1}{n} T_1 T_1^T - \bar{T}_1 \bar{T}_1^T,$$

$$S_{22}^{(n)} = \frac{1}{n} T_2 T_2^T - \bar{T}_2 \bar{T}_2^T,$$

$$S_{12}^{(n)} = \frac{1}{n} T_1 T_2^T - \bar{T}_1 \bar{T}_2^T,$$

and the sample analogue of Σ:

$$S^{(n)} = (S_{11}^{(n)})^{-1} S_{12}^{(n)} (S_{22}^{(n)})^{-1} (S_{12}^{(n)})^T.$$

Let $\sigma(S^{(n)})$ denote the spectrum of $S^{(n)}$.

Lemma 1. *If* $S_{11}^{(n)}$ *and* $S_{22}^{(n)}$ *are positive definite a.e.* Q_n, *then* $\sigma(S^{(n)}) \subset [0, 1]$ *a.e.* Q_n. *Then we consider the sample analogue of* $\varrho(X, Y)$:

$$\hat{\varrho}_n = \begin{cases} 1 & \text{if any eigenvalue of } S^{(n)} \text{ is equal to } 1, \\ \Big[1 - \prod\limits_{1 \neq r^2 \in \sigma(S^{(n)})} (1-r^2)\Big]^{1/2} & \text{otherwise.} \end{cases}$$

Obviously, $\hat{\varrho}_n \in L^2(E_n, \varepsilon_n, Q_n)$.

We investigate the asymptotic behaviour of $\hat{\varrho}_n$ when n tends to infinity. First, we extend the sample space $(E_n, \varepsilon_n, Q_n)$ to the case of infinitely many observations of (X, Y), i.e., to

$$(E, \varepsilon, Q) = \big(R^p \times R^q, \mathscr{B}(R^p)\otimes\mathscr{B}(R^q), P_{X,Y}\big)^\infty.$$

THEOREM 5. *If $\{X_i, Y_j, i = 1, \ldots, p, j = 1, \ldots, q\}$ is a family of stochastically independent random variables, then $n(\hat{\varrho}_n)^2$ is weakly convergent to $V_0^2 \in L^1(E, \varepsilon, Q)$, where V_0^2 has the χ^2-distribution with $\nu_0 = pq$ degrees of freedom.*

Proof. Without no loss of generality we can assume that X_{ij} and Y_{ij} are standardized. In view of the assumption of mutual independence of the components of X and Y, there exists $N_0 \in \mathcal{N}$ such that for any $n > N_0$ all eigenvalues of $S^{(n)}$ are different from 1 a.e. Q. Then

$$\hat{\varrho}_n^2 = 1 - \det(I_p - S^{(n)}).$$

We use the fact that the statistic

$$Z_n = n[1 - \det(I_p - S_{12}^{(n)}(S_{12}^{(n)})^T)]$$

has the same weak limit as $n\hat{\varrho}_n^2$. In the sequel we calculate the weak limit of Z_n.

The general term of $S_{12}^{(n)}(S_{12}^{(n)})^T = [S_{ij}]_{i,j=1,\ldots,p}$ is of the form

$$S_{ij} = \sum_{l=1}^{q} \left(\frac{1}{n} X_i Y_l - \bar{X}_i \bar{Y}_l\right)\left(\frac{1}{n} X_j Y_l - \bar{X}_j \bar{Y}_l\right),$$

where

$$\frac{1}{n} X_i Y_l = \frac{1}{n} \sum_{j=1}^{n} X_{ij} Y_{lj}.$$

Since, for any $A = [A_{ij}]_{i,j=1,\ldots,p}$,

$$\det A = \sum_{i_1,\ldots,i_p \in \{1,\ldots,p\}} \delta_{i_1 \ldots i_p} a_{1 i_1} \ldots a_{p i_p}$$

(where the function $\delta_{i_1 \ldots i_p}$ is equal to -1 when $(i_1, \ldots, i_p)$ is an odd permutation of $\{1, \ldots, p\}$, and is equal to 1 when $(i_1, \ldots, i_p)$ is an even permutation of $\{1, \ldots, p\}$, and is equal to 0 otherwise), we have

$$(12) \qquad nZ_n^2 = n - n \prod_{i=1}^{p} (1 - S_{ii}) - n(\ldots)$$

$$= n - n\Big(1 - \sum_{i_1} S_{i_1 i_1} + \sum_{i_1 i_2} S_{i_1 i_1} S_{i_2 i_2} + \ldots + (-1)^p \prod_{i=1}^{p} S_{ii}\Big) - n(\ldots)$$

$$= n \sum_{i_1} S_{i_1 i_1} - n \sum_{i_1 i_2} S_{i_1 i_1} S_{i_2 i_2} + \ldots + (-1)^{p+1} \prod_{i=1}^{p} S_{ii} - n(\ldots).$$

Consider the first component of (12):

$$(13) \qquad n \sum_{i=1}^{n} S_{ii} = n \sum_{i=1}^{p} \sum_{l=1}^{q} \left(\frac{1}{n} X_i Y_l - \bar{X}_i \bar{Y}_l\right)^2$$

$$= \sum_{i=1}^{p} \sum_{l=1}^{q} \left(\left(\frac{X_i Y_l}{\sqrt{n}}\right)^2 - 2 X_i Y_l \bar{X}_i \bar{Y}_l + n(\bar{X}_i \bar{Y}_l)^2\right).$$

By the Central Limit Theorem we have

$$\left(\frac{X_i Y_j}{\sqrt{n}}\right)^2 \Rightarrow V_{ij}^2, \qquad V_{ij}^2 \in L^1(E, \varepsilon, Q).$$

Further, applying the Central Limit Theorem and the weak laws of large numbers to the remaining components of (13) we obtain

$$2X_i Y_j \bar{X}_i \bar{Y}_j \to 0, \qquad n(\bar{X}_i \bar{Y}_j)^2 \to 0.$$

Therefore

$$n \sum_{i=1}^{p} \sum_{l=1}^{q} \left(\frac{1}{n} X_i Y_l - \bar{X}_i \bar{Y}_l\right)^2 \Rightarrow \sum_{i=1}^{p} \left(\sum_{l=1}^{q} V_{il}^2\right) = \sum_{i=1}^{p} V_i^2 = V_0^2,$$

where

$$V_i^2 = \sum_{j=1}^{q} V_{ij}^2.$$

It is easy to show that the statistics $V_{ij}^2 \in L^1(E, \varepsilon, Q)$, $i = 1, \ldots, p$, $j = 1, \ldots, q$, are stochastically independent and have the χ^2-distribution with 1 degree of freedom. Then $V_0^2 = \sum_{i=1}^{p} V_i^2$ has the χ^2-distribution with ν_0 degrees of freedom, where $\nu_0 = pq$.

The remaining components of (12), namely

$$n \sum_{i_1 i_2} S_{i_1 i_1} S_{i_2 i_2}, \ldots, n(-1)^{p+1} \prod_{i=1}^{p} S_{ii}, \; n(\ldots),$$

have weak limits equal to 0, which completes the proof.

References

[1] H. P. Höschel, *Ein axiomatischer Zugang zu Abhängigkeitsmassen für Zufallsvektoren und die Spurkorrelation*, Math. Operationsforsch. Statist. 7 (1976), pp. 789–802.

[2] P. E. Jupp and K. V. Mardia, *A general correlation coefficient for directional data and related regression problems*, Biometrika 67 (1980), pp. 163–173.

[3] K. V. Mardia, *Measures of multivariate skewness and kurtosis with applications*, ibidem 57 (1970), pp. 510–530.

INSTITUTE OF COMPUTER SCIENCE
POLISH ACADEMY OF SCIENCES
00-901 WARSAW

Received on 1986.11.11