ON THE GENERAL NOTION OF MAXIMAL CORRELATION

by

PÉTER CSÁKI and JÁNOS FISCHER

Introduction

In their recent papers [3], [4] the authors discussed some properties of the maximal correlation of two random variables as introduced by H. GEBELEIN [5]. The natural idea of the straightforward generalization of this measure of connection for random vectors, for stochastic processes and even for σ-algebras has been already realized in [8] and [9]. The present paper aims at adding some further properties to those described in the alluded works.

In § 1 a notion involving the maximal correlation is considered in the general Hilbert space. § 2 deals with topological and metrical problems. § 3 is to show how the maximal correlation may be defined in the non-commutative probability theory developed in [11]. § 4 is about some properties of the maximal correlation between σ-algebras. In § 5 the intuitive background of the postulates given in § 1 is revealed. Finally, § 6 contains some particular cases and examples.

§ 1. General notions

To commence with, the notions and facts from the Hilbert space theory needed and the symbols used in this paper are explained.

1.1. Let H_1 and H_2 be two real Hilbert spaces with scalar products $(x_1, x_2)_1$ $(x_1 \in H_1, x_2 \in H_1)$ and $(y_1, y_2)_2$ $(y_1 \in H_2, y_2 \in H_2)$, the respective norms denoted by $\| x \|_1$ $(x \in H_1)$ and $\| y \|_2$ $(y \in H_2)$.

Let $\boldsymbol{T}$ be a linear bounded operator whose domain is the space H_1 and which takes on its values in the space H_2. It is known that there exists a uniquely determined linear bounded operator $\boldsymbol{T}^*$ whose domain is the space H_2 and it takes on its values in the space H_1, furthermore

$$(1.1) \qquad (\boldsymbol{T} x, y)_2 = (x, \boldsymbol{T}^* y)_1$$

holds for all $x \in H_1$ and $y \in H_2$. The operator $\boldsymbol{T}^*$ will be called the adjoint of the operator $\boldsymbol{T}$. The equality $\| \boldsymbol{T} \| = \| \boldsymbol{T}^* \|$ is valid.

The real number λ is called an eigenvalue of the pair of operators $\boldsymbol{T}$, $\boldsymbol{T}^*$ and the pair of elements $x \in H_1$, $y \in H_2$ is said to be a pair of eigenelements (of $\boldsymbol{T}$, $\boldsymbol{T}^*$) belonging to λ if

$$(1.2) \qquad \begin{aligned} \boldsymbol{T} x &= \lambda y \\ \boldsymbol{T}^* y &= \lambda x \end{aligned}$$

excluding the case when both x and y are equal to the zero element.

As it is known, the operator $\boldsymbol{T}$ being completely continuous, the members of the pairs of eigenelements form complete orthogonal systems in the respective spaces and $\|\boldsymbol{T}\|$ is the greatest eigenvalue.

The square root of the quantity

$$|||\boldsymbol{T}|||^2 = |||\boldsymbol{T}^*|||^2 = \sum_{i,k} (\boldsymbol{T} x_i, y_k)_2 \tag{1.3}$$

is called the double norm of the operators $\boldsymbol{T}$ and $\boldsymbol{T}^*$, where $\{x_i\}$ and $\{y_k\}$ are complete orthonormal systems in the spaces H_1 and H_2, respectively. It is known that — $\{\lambda_i\}$ being the sequence of eigenvalues —

$$|||\boldsymbol{T}|||^2 = \sum_i \lambda_i^2 \tag{1.4}$$

is equivalent to the statement that the members of pairs of eigenelements furnish complete orthogonal systems in the respective spaces. If even $|||\boldsymbol{T}||| < \infty$ then the operator $\boldsymbol{T}$ is completely continuous.

1.2. Henceforth, let us treat the Hilbert spaces as subspaces of the same given H in which the scalar product will be denoted by (x, y) and the norm by $\|x\|$. Let $\mathscr{H}$ denote the set of all subspaces of H except the subspace containing the single element zero. Let $\boldsymbol{P}_{H_0}$ denote the operator of the orthogonal projection on the subspace $H_0 \in \mathscr{H}$ i.e. the linear bounded operator, which has the following properties:

1. Having $x \in H$,

$$(x, y) = (x_0, y) \tag{1.5}$$

for some $x_0 \in H_0$ and for all $y \in H_0$ if and only if $x_0 = \boldsymbol{P}_{H_0}x$.

2. Having $x \in H$,

$$\min_{y \in H_0} \|x - y\| = \|x - x_0\| \tag{1.6}$$

for some $x_0 \in H_0$ if and only if $x_0 = \boldsymbol{P}_{H_0}x$.

Further, it is kown that $\boldsymbol{P}_{H_0}^2 = \boldsymbol{P}_{H_0}^* = \boldsymbol{P}_{H_0}$.

As to the relative position of two subspaces $H_1 \in \mathscr{H}$ and $H_2 \in \mathscr{H}$ the quantity

$$\delta(H_1, H_2) = \inf_{\substack{x \in H_1 \\ \|x\| = 1}} \|x - \boldsymbol{P}_{H_2} x\| \tag{1.7}$$

may be esteemed characteristic, though this quantity does not represent a distance in the metrical sense as this will be shown in § 2. However, the basis of the postulates given in Section 1.3 will be the quantity (1.7) because certain intuitive meaning can be attributed to it for the probability theory in some particular cases (see § 5.).

Now, some simple properties of the quantity (1.7) are listed.

Theorem 1.1. *For any subspaces $H_1 \in \mathscr{H}$ and $H_2 \in \mathscr{H}$*

1° $0 \leqq \delta(H_1, H_2) \leqq 1$;

2° $\delta(H_1, H_2) = \delta(H_2, H_1)$;

3° $\delta(H_1, H_2) = 0$ *if* $H_1 \cap H_2 \in \mathscr{H}$;

4° $\delta(H_1, H_2) = 1$ *if and only if* H_1 *and* H_2 *are orthogonal;*

5° $\delta(H_1, H_2) \geqq \delta(H_1', H_2')$ *if* $H_1 \subset H_1'$ *and* $H_2 \subset H_2'$.

Proof. The statement 1° follows from the inequalities

$$0 \leqq \|x - \boldsymbol{P}_{H_2} x\|^2 = 1 - \|\boldsymbol{P}_{H_2} x\|^2 \leqq 1 \qquad \|x\| = 1.$$

For the proof of 2°, let $x \in H_1$, $\|x\| = 1$ be arbitrary. If $\boldsymbol{P}_{H_2} x = 0$ then $\|x - \boldsymbol{P}_{H_2} x\| = 1 \geqq \delta(H_2, H_1)$; on the other hand, if $\boldsymbol{P}_{H_2} x \neq 0$ then

$$\|x - \boldsymbol{P}_{H_2} x\|^2 = 1 - \|\boldsymbol{P}_{H_2} x\|^2 = 1 - \frac{(\boldsymbol{P}_{H_2} x, \boldsymbol{P}_{H_2} x)^2}{\|\boldsymbol{P}_{H_2} x\|^2} = 1 - \frac{(\boldsymbol{P}_{H_1} \boldsymbol{P}_{H_2} x, x)^2}{\|\boldsymbol{P}_{H_2} x\|^2} \geqq$$

$$\geqq 1 - \frac{\|\boldsymbol{P}_{H_1} \boldsymbol{P}_{H_2} x\|^2}{\|\boldsymbol{P}_{H_2} x\|^2} = 1 - \|\boldsymbol{P}_{H_1} y\|^2 = \|y - \boldsymbol{P}_{H_1} y\|^2 \geqq \delta^2(H_2, H_1),$$

where $y = \dfrac{\boldsymbol{P}_{H_2} x}{\|\boldsymbol{P}_{H_2} x\|}$. Therefore $\delta(H_1, H_2) \geqq \delta(H_2, H_1)$. The converse inequality may be verified analogously, thus 2° is proved.

If $H_1 \cap H_2 \in \mathscr{H}$ there exists $x_0 \in H_1 \cap H_2$ such that $\|x_0\| = 1$ for which $\boldsymbol{P}_{H_2} x_0 = x_0$, hence $\|x_0 - \boldsymbol{P}_{H_2} x_0\| = 0$ and this proves 3°.

If H_1 and H_2 are orthogonal then $\|\boldsymbol{P}_{H_2} x\|^2 = (x, \boldsymbol{P}_{H_2} x) = 0$ holds, i.e. $\|x - \boldsymbol{P}_{H_2} x\| = 1$ for all $x \in H_1$, $\|x\| = 1$, thus $\delta(H_1, H_2) = 1$. Conversely, if $\delta(H_1, H_2) = 1$ then $\|\boldsymbol{P}_{H_2} x\|^2 = 1 - \|x - \boldsymbol{P}_{H_2} x\|^2 = 0$ for all $x \in H_1$, $\|x\| = 1$, i.e. $(x, y) = (\boldsymbol{P}_{H_2} x, y) = 0$ for all $x \in H_1$, $y \in H_2$ — thus 4° is proved.

Finally, if $H_2 \subset H_2'$ then $\boldsymbol{P}_{H_2} = \boldsymbol{P}_{H_2} \boldsymbol{P}_{H_2'}$, hence if even $H_1 \subset H_1'$

$$\delta^2(H_1, H_2) \geqq \inf_{\substack{x \in H_1' \\ \|x\| = 1}} \|x - \boldsymbol{P}_{H_2} x\|^2 = \inf_{\substack{x \in H_1' \\ \|x\| = 1}} \{1 - \|\boldsymbol{P}_{H_2} \boldsymbol{P}_{H_2'} x\|^2\} \geqq$$

$$\geqq \inf_{\substack{x \in H_1' \\ \|x\| = 1}} \{1 - \|\boldsymbol{P}_{H_2'} x\|^2\} = \inf_{\substack{x \in H_1' \\ \|x\| = 1}} \|x - \boldsymbol{P}_{H_2'} x\|^2 = \delta^2(H_1', H_2'),$$

thus all the statements of the theorem are proved.

1.3. In order to characterize the "similarity of position" of subspaces let a real number $\alpha(H_1, H_2)$ correspond to every pair of subspaces $H_1 \in \mathscr{H}$, $H_2 \in \mathscr{H}$ such that the following Postulates are fulfilled:

1. $\alpha(H_1, H_2) \geqq \alpha(H_1', H_2')$ if and only if $\delta(H_1, H_2) \leqq \delta(H_1', H_2')$.

2. $\alpha(H_1, H_2) = 0$ if the subspaces H_1 and H_2 are orthogonal.

3. $\alpha(H_1, H_2) = 1$ if $\delta(H_1, H_2) = 0$.

Evidently, these Postulates are satisfied by an α if and only if it is a strictly decreasing function of δ, taking on the value 1 at the point 0 and the value 0 at the point 1. However, the choice of such a function of δ is equivalent to the choice of the scale measuring the relative position of subspaces. Therefore α is uniquely determined but scaling. It is advantageous to choose the scale such that $\alpha(H_1, H_2) = \mathbf{S}(H_1, H_2)$, where

$$\mathbf{S}(H_1, H_2) = \sqrt{1 - \delta^2(H_1, H_2)}. \tag{1.8}$$

Theorem 1.1. may be written by the aid of $\mathbf{S}(H_1, H_2)$ in the following form:

Theorem 1.1'. *For any pair of subspaces* $H_1 \in \mathscr{H}$ *and* $H_2 \in \mathscr{H}$

1° $0 \leq \mathbf{S}(H_1 H_2) \leq 1$;

2° $\mathbf{S}(H_1, H_2) = \mathbf{S}(H_2, H_1)$;

3° $\mathbf{S}(H_1, H_2) = 1$ *if* $H_1 \cap H_2 \in \mathscr{H}$;

4° $\mathbf{S}(H_1, H_2) = 0$ *if and only if* H_1 *and* H_2 *are orthogonal*;

5° $\mathbf{S}(H_1, H_2) \leq \mathbf{S}(H_1', H_2')$ *if* $H_1 \subset H_1'$ *and* $H_2 \subset H_2'$.

Theorem 1.2.

$$\mathbf{S}(H_1, H_2) = \sup_{\substack{x \in H_1, \|x\|=1 \\ y \in H_2, \|y\|=1}} (x, y) . \tag{1.9}$$

Proof. If $x \in H_1$, $\| x \| = 1$ and $y \in H_2$, $\| y \| = 1$ then

$$(x, y) = (\boldsymbol{P}_{H_2} x, y) \leq \|\boldsymbol{P}_{H_2} x\| = \sqrt{1 - \|x - \boldsymbol{P}_{H_2} x\|^2} \leq \tag{1.10}$$
$$\leq \sqrt{1 - \delta^2(H_1, H_2)} = \mathbf{S}(H_1, H_2) .$$

In the case of $\mathbf{S}(H_1, H_2) = 0$ the statement follows from inequality (1.10). If $\mathbf{S}(H_1, H_2) > 0$, let the sequence of $x_n \in H_1$ be such that $\| x_n \| = 1$ and $\| x_n - \boldsymbol{P}_{H_2} x_n \| \to \delta(H_1, H_2)$ i.e. $\| \boldsymbol{P}_{H_2} x_n \| \to \mathbf{S}(H_1, H_2)$; hence the sequence $\{x_n\}$ may be chosen so that $\boldsymbol{P}_{H_2} x_n \neq 0$. Choosing $y_n = \dfrac{\boldsymbol{P}_{H_2} x_n}{\|\boldsymbol{P}_{H_2} x_n\|}$,

$$(x_n, y_n) = \frac{(x_n, \boldsymbol{P}_{H_2} x_n)}{\|\boldsymbol{P}_{H_2} x_n\|} = \|\boldsymbol{P}_{H_2} x_n\| \to \mathbf{S}(H_1, H_2)$$

from which the statement follows.

1.4. Let $\boldsymbol{P}_{H_2}^{H_1}$ denote the restriction of the operator $\boldsymbol{P}_{H_2}$ to the subspace H_1, i.e. the domain of $\boldsymbol{P}_{H_2}^{H_1}$ is equal to the subspace H_1 and $\boldsymbol{P}_{H_2}^{H_1} x = \boldsymbol{P}_{H_2} x$ if $x \in H_1$.

Theorem 1.3.

$$(\boldsymbol{P}_{H_2}^{H_1})^* = \boldsymbol{P}_{H_1}^{H_2} \tag{1.11}$$

Proof. If $x \in H_1$ and $y \in H_2$, then

$$(\boldsymbol{P}_{H_2}^{H_1} x, y) = (\boldsymbol{P}_{H_2} x, y) = (x, y) = (x, \boldsymbol{P}_{H_1} y) = (x, \boldsymbol{P}_{H_1}^{H_2} y) .$$

Theorem 1.4.

$$\mathbf{S}(H_1, H_2) = \|\boldsymbol{P}_{H_2}^{H_1}\| = \|\boldsymbol{P}_{H_1}^{H_2}\| . \tag{1.12}$$

Proof.

$$\| \boldsymbol{P}_{H_2}^{H_1} \| = \sup_{\substack{x \in H_1 \\ \|x\|=1}} \| \boldsymbol{P}_{H_2}^{H_1} x \| = \sup_{\substack{x \in H_1 \\ \|x\|=1}} \sqrt{1 - \| x - \boldsymbol{P}_{H_2} x \|^2} = S(H_1, H_2).$$

The following theorem specifies the cases for which the supremum is attainable in (1.9).

Theorem 1.5.

(1.13) $$\mathbf{S}(H_1, H_2) = (x_0, y_0) \qquad x_0 \in H_1, \|x_0\| = 1; y_0 \in H_2, \|y_0\| = 1$$

if and only if x_0, y_0 *form a pair of eigenelements — belonging to the eigenvalue* $S = \mathbf{S}(H_1, H_2)$ *— of the pair of operators* $\boldsymbol{P}_{H_2}^{H_1}, \boldsymbol{P}_{H_1}^{H_2}$.

Proof. If (1.13) holds then for any $y \in H_2$

$$\|x_0 - y\|^2 = 1 - 2(x_0, y) + \|y\|^2 \geqq 1 - 2S\|y\| + \|y\|^2 =$$
$$= 1 - S^2 + (S - \|y\|)^2 \geqq 1 - S^2 = \|x_0 - Sy_0\|^2,$$

wherefrom owing to (1.5)

$$\boldsymbol{P}_{H_2}^{H_1} x_0 = S y_0$$

and similarly

$$\boldsymbol{P}_{H_1}^{H_2} y_0 = S x_0.$$

Conversely, if x_0, y_0 form a pair of normed eigenelements,

$$(x_0, y_0) = (\boldsymbol{P}_{H_2}^{H_1} x_0, y_0) = (Sy_0, y_0) = \mathbf{S}(H_1, H_2),$$

with which the statement is proved.

Theorem 1.6.

$$\sup_{\|x\|=1} \{\|\boldsymbol{P}_{H_1} x\| \, \|\boldsymbol{P}_{H_2} x\| - \|x - \boldsymbol{P}_{H_1} x\| \, \|x - \boldsymbol{P}_{H_2} x\|\} = \mathbf{S}(H_1, H_2).$$

Proof.

(1.14) $$\|\boldsymbol{P}_{H_1} x\| \, \|\boldsymbol{P}_{H_2} x\| - \|x - \boldsymbol{P}_{H_1} x\| \, \|x - \boldsymbol{P}_{H_2} x\| \leqq \mathbf{S}(H_1, H_2),$$

for all $x \in H$, $\|x\| = 1$

holds as proved by H. P. KRAMER [9]. Let $x_n \in H_2$, $\| x_n \| = 1$ $(n = 1, 2, \ldots)$ be chosen for x in (1.14) such that $\| \boldsymbol{P}_{H_1} x_n \| \to \mathbf{S}(H_1, H_2)$; then the left side of (1.14) converges to $\mathbf{S}(H_1, H_2)$, wherefrom the statement follows.

Now, let H_1 and H_2 be two separable subspaces of H and $\{x_i\}$, $\{y_k\}$ be respective complete orthonormal systems. Introduce the quantity

(1.15) $$\mathbf{C}(H_1, H_2) = [\sum_{i,k} (x_i, y_k)^2]^{1/2}.$$

Theorem 1.7.

(1.16) $$\mathbf{C}(H_1, H_2) = |||\boldsymbol{P}_{H_2}^{H_1}||| = |||\boldsymbol{P}_{H_1}^{H_2}|||\,.$$

Proof. From the definitions (1.3) and (1.15)

$$|||\boldsymbol{P}_{H_2}^{H_1}|||^2 = \sum_{i,k} (\boldsymbol{P}_{H_2}^{H_1} x_i, y_k)^2 = \sum_{i,k} (x_i, y_k)^2\,,$$

$\{x_i\}$ and $\{y_k\}$ being complete orthonormal systems in H_1 and H_2, respectively.

§ 2. Topological problems

As mentioned in § 1, $\delta(H_1, H_2)$ defined by (1.6) represents no distance in the set $\mathscr{H}$. The existence of a non-vanishing function of δ defining a distance on $\mathscr{H}$ was first inquired in [9]. Theorem 2.1 which will be formulated in general form, excludes even the existence of a non-vanishing function of δ the neighbourhoods generated by which would form a basis of a topology. (The definition of topological space and related axioms of bases see e. g. [7] pp. 34—35.)

Theorem 2.1. *Let X be an arbitrary set. Let us suppose that for the function* $\varphi(x, y)$ *of two variables defined on X and for any* $x \in X$ *and* $y \in X$ *there exists a* $z \in X$ *such that*

$$\varphi(x, z) = \varphi(z, y) = \varphi(z, z) . \tag{2.1}$$

If f is a non-negative function defined on the range of φ *such that the subsets*

$$V(x, \varepsilon) = \{y : \gamma(x, y) < \varepsilon\} \qquad \varepsilon > 0$$

(the neighbourhood of x with radius ε *generated by* γ*) satisfy the axioms of bases where*

$$\gamma(x, y) = f(\varphi(x, y)) , \qquad x \in X, y \in X$$

then

$$\gamma(x, y) \equiv 0 .$$

Proof. Let $x \in X$, $y \in X$. In consequence of the condition on φ, there exists a $z \in X$ satisfying (2.1), wherefrom $\gamma(x, z) = \gamma(z, y) = \gamma(z, z)$ follows. According to the first axiom of bases $z \in V(z, \varepsilon)$ for all $\varepsilon > 0$, that is $\gamma(z, z) < \varepsilon$ for all $\varepsilon > 0$, i.e. $\gamma(x, z) = \gamma(z, y) = \gamma(z, z) = 0$, therefore

$$z \in V(x, \varepsilon') \qquad \text{for all } \varepsilon' > 0, \tag{2.2}$$

$$y \in V(z, \varepsilon'') \qquad \text{for all } \varepsilon'' > 0. \tag{2.3}$$

According to the third axiom of bases for any $\varepsilon > 0$ there exists $\varepsilon' > 0$ such that for any point of $V(x, \varepsilon')$ and thus by (2.2) for z, there exists $\varepsilon'' > 0$ such that $V(z, \varepsilon'') \subset V(x, \varepsilon)$. From this and (2.3), $y \in V(x, \varepsilon)$, that is $\gamma(x, y) < \varepsilon$ for all $\varepsilon > 0$ is valid, hence $\gamma(x, y) = 0$ for all $x \in X$ and $y \in X$.

From this Theorem it is evident, no non-vanishing function of φ with property (2.1) may generate a quasi-distance.

Corollary. *There is no non-vanishing function of the quantity defined by* (1.7) *to generate a topology in* $\mathcal{H}$.

Proof. For any $H_1 \in \mathcal{H}$ and $H_2 \in \mathcal{H}$

$$\delta(H_1, H) = \delta(H, H_2) = \delta(H, H) = 0 .$$

Hence δ satisfies the condition of Theorem 2.1, thus the statement holds.

§ 3. Remarks on the case of non-commutative probability theory

In this section a special Hilbert space is briefly discussed implying the functional ones arising from the ordinary or the non-commutative probability theory as particular cases (see [11]). In this abstract formulation we deal with the special projectors which provide the conditional expectation with norms equaling the maximal correlations.

Let L^0 be a real linear space in which the following further operations are defined:

1. Multiplication. An element $xy \in L^0$ corresponds to every $x \in L^0$ and $y \in L^0$ so that

a) $x(yz) = (xy)z$ for all $x \in L^0$, $y \in L^0$, $z \in L^0$,

b) there exists a unit-element $e \in L^0$ such that $xe = ex = x$ for all $x \in L^0$,

c) $x(y + z) = xy + xz$ and $(x + y)z = xz + yz$ for all $x \in L^0$, $y \in L^0$, $z \in L^0$,

d) $x(\alpha\, y) = (\alpha\, x)y = \alpha(xy)$ for all $x \in L^0$, $y \in L^0$ and any real number α.

2. Involution. An element $x' \in L^0$ corresponds to every $x \in L^0$ so that

a) $(x')' = x$ for all $x \in L^0$,

b) $(x + y)' = x' + y'$ for all $x \in L^0$ and $y \in L^0$,

c) $(\alpha\, x)' = \alpha\, x'$ for all $x \in L^0$ and any real number α,

d) $(x\, y)' = y'\, x'$ for all $x \in L^0$ and $y \in L^0$.

It is easy to see that the unit-element is unique and $0\, x = 0$ for all $x \in L^0$. The elements $x \in L^0$ for which $x' = x$ are called self-adjoint. They form a linear subspace which contains the unit-element and the zero element. The elements of the set

$$\mathscr{P} = \{xx' : x \in L^0\} \tag{3.1}$$

are called positive elements. The unit-element and zero element are positive ones and the positive elements are self-adjoint.

Let μ be a real-valued function defined on the linear subspace $L^1 \subset L^0$ with the following properties:

1. μ is linear on L^1.

2. If $xx' \in L^1$ then $\mu(xx') \geq 0$ and if $\mu(xx') = 0$ then $x = 0$.

3. If $xx' \in L^1$ and $yy' \in L^1$ then $xy' + yx' \in L^1$.

4. $e \in L^1$ and $\mu\,(e) = 1$.

Put

$$L^2 = \{x : xx' \in L^1\}\,. \tag{3.2}$$

From the property 3 it follows that L^2 is a linear subspace of L^0.

For elements $x \in L^2$ and $y \in L^2$ let the scalar product and the norm be

$$(x, y) = \mu(xy' + yx') \qquad \text{and} \qquad \|x\| = \sqrt{(x, x)}, \tag{3.3}$$

respectively, with which the space L^2 forms a — not necessarily complete — Hilbert space.

Further, let $\boldsymbol{P}$ be a linear transformation on L^1 to itself with the following properties:

1. If $x \in L^2 \cap L^1$ then $x(\boldsymbol{P}\, x)' + (\boldsymbol{P}\, x)\, x' \in L^1$.

2. If $x(\boldsymbol{P}\, y)' + (\boldsymbol{P}\, y)\, x' \in L^1$ then $\boldsymbol{P}[x(P\, y)' + (\boldsymbol{P}\, y)\, x'] =$
$= \boldsymbol{P}\, x(\boldsymbol{P}\, y)' + \boldsymbol{P}\, y(\boldsymbol{P}\, x)'$.

3. $\mu(\boldsymbol{P}\, x) = \mu(x)$.

4. $\boldsymbol{P}\, e = e$.

Theorem 3.1. *If* $x \in L^2 \cap L^1$ *then* $\boldsymbol{P}\, x \in L^2$.

Proof.

$$2\, \boldsymbol{P}\, x(\boldsymbol{P}\, x)' = \boldsymbol{P}[x(\boldsymbol{P}\, x)' + (\boldsymbol{P}\, x)\, x'] \in L^1 .$$

Theorem 3.2. *If* $x \in L^2 \cap L^1$ *then* $\boldsymbol{P}x$ *is the orthogonal projection of* x *on the subspace*

$$H = \{\boldsymbol{P}\, x : x \in L^2 \cap L^1\} . \tag{3.4}$$

Proof.

$$(\boldsymbol{P}\, x, \boldsymbol{P}\, y) = \mu(\boldsymbol{P}\, x(\boldsymbol{P}\, y)' + \boldsymbol{P}\, y(\boldsymbol{P}\, x)') = \mu(\boldsymbol{P}[x(\boldsymbol{P}\, y)' + (\boldsymbol{P} y)\, x']) =$$
$$= \mu(x(\boldsymbol{P}\, y)' + (\boldsymbol{P}\, y)\, x') = (x, \boldsymbol{P}\, y)$$

for all $\boldsymbol{P}y \in L^2$, hence the statement follows from (1.5).

Therefore the restriction of $\boldsymbol{P}$ to $L^2 \cap L^1$ is the projection $\boldsymbol{P}_H$. However, not every projection has the above properties. These particular projectors represent the conditional expectation either in the ordinary or the non-commutative probability theory. The quantity

$$\mathbf{S}(H_{1,0}, H_{2,0}) = \|\boldsymbol{P}^{H_{1,0}}_{H_{2,0}}\| = \sup_{\substack{x \in H_{1,0}\ \|x\| = 1 \\ y \in H_{2,0}\ \|y\| = 1}} \mu(xy' + yx') ,$$

is called the maximal correlation of H_1 and H_2, where $H_{1,0}$ and $H_{2,0}$ are the orthogonal complements of the subspaces of the above types H_1 and H_2, respectively, to the unit-element.

§ 4. Maximal correlation of σ-algebras

4.1. In the present section let us consider the case of (commutative) probability theory. Let $(\Omega, F, \mathbf{P})$ be a probability space and $L^2 = L^2(\Omega, F, \mathbf{P})$, further $L^2_{F_0} = L^2(\Omega, F_0, \mathbf{P})$, where F_0 is an arbitrary sub-σ-algebra of the σ-algebra F. Let us denote the subspace of the elements with zero expected values of the space $L^2_{F_0}$ by $L^2_{F_0,0}$. It should be remarked that not every subspace of L^2 is of the form $L^2_{F_0}$. (R. R. Bahadur [1] has given necessary and sufficient conditions for a subspace to be of this form.)

Let F_1 and F_2 be two sub-σ-algebras of F. The quantity

$$\mathbf{S}(F_1, F_2) = \mathbf{S}(L^2_{F_1,0}, L^2_{F_2,0}) \tag{4.1}$$

is called the maximal correlation of these sub-σ-algebras.

It should be remarked that no topology may be introduced among the sub-σ-algebras of F by the aid of non-vanishing functions of the maximal correlation as it follows from Theorem 2.1 in consequence of $\mathbf{S}\ (F_1, F) = \mathbf{S}(F_2, F) = \mathbf{S}(F, F) = 1$.

4.2. In the subsequent part of the paper the following symbols will be used for particular cases:

F_V for the smallest σ-algebra with respect to which every element of V is measurable, V being an arbitrary set of random variables; $F_{\xi_t} = F_{\{\xi_t\}}$ and

$F_\xi = F_{\{\xi\}}$, further the spaces

$$(4.2)\qquad \begin{aligned} L_V^2 &= L_{F_V}^2,\ L_{V,0}^2 = L_{F_V,0}^2\,,\\ L_{\xi_t}^2 &= L_{\{\xi_t\}}^2,\ L_{\xi_t,0}^2 = L_{\{\xi_t\},0}^2\,,\\ L_\xi^2 &= L_{\{\xi\}}^2,\ L_{\xi,0}^2 = L_{\{\xi\},0}^2\,; \end{aligned}$$

moreover, the projectors

$$(4.3)\qquad \begin{aligned} \boldsymbol{P}_V &= \boldsymbol{P}_{L_V^2}\,,\\ \boldsymbol{P}_{\xi_t} &= \boldsymbol{P}_{\{\xi_t\}}\,,\\ \boldsymbol{P}_\xi &= \boldsymbol{P}_{\{\xi\}}\,; \end{aligned}$$

finally,

$$(4.4)\qquad \begin{aligned} \mathbf{S}(\xi_t, \eta_t) &= \mathbf{S}(F_{\xi_t}, F_{\eta_t})\,,\\ \mathbf{S}(\xi, \eta) &= \mathbf{S}(F_\xi, F_\eta) \end{aligned}$$

and

$$(4.5)\qquad \begin{aligned} \boldsymbol{P}_{F_2}^{F_1} &= \boldsymbol{P}_{L_{F_2,0}^2}^{L_{F_1,0}^2}\\ \boldsymbol{P}_{\eta_t}^{\xi_t} &= \boldsymbol{P}_{L_{\eta_t,0}^2}^{L_{\xi_t,0}^2}\\ \boldsymbol{P}_\eta^\xi &= \boldsymbol{P}_{L_{\eta,0}^2}^{L_{\xi,0}^2} \end{aligned}$$

where ξ_t, η_t denote stochastic processes and ξ, η vector random variables (including single random variables as one-dimensional vectors), $\{\xi_t\}$, $\{\eta_t\}$ and $\{\xi\}$, $\{\eta\}$ denote the sets consisting of the components of ξ_t, η_t and ξ, η, respectively.

It is worthy of note that $\mathbf{S}(H_1, H_2)$ is equal to the correlation coefficient of the random variables $\zeta_1 \in L^2$, $\zeta_2 \in L^2$, if H_1 and H_2 are the sets of the linear functions with zero expected values of ζ_1 and ζ_2, respectively. If H is the set of the linear functions with zero expected values of $\zeta \in L^2$ and ξ is a random variable then $\mathbf{S}(H, L_{\xi,0}^2)$ is the correlation ratio of ζ on ξ.

4.3. Now, we consider the extreme cases 0 and 1 of maximal correlation.

The sub-σ-algebras F_1 and F_2 are called independent if every element of F_1 is independent of each one of F_2. It is easy to see that the independence of sub-σ-algebras (including that of stochastic processes and that of random variables) is equivalent to zero maximal correlation (see e. g. [2]).

The common elements of probability one and zero of two sub-σ-algebras of F are called their trivial common elements. Obviously, the empty set and the whole set Ω are the trivial common elements of any two sub-σ-algebras. If the sub-σ-algebras F_1 and F_2 have non-trivial common elements then $\mathbf{S}(F_1, F_2) = 1$, because of having such ones is equivalent to $L_{F_1,0}^2$ and $L_{F_2,0}^2$ having non-zero common elements. Especially, $\mathbf{S}(\xi, \eta) = 1$ if for the vector random variables ξ and η non-constant Borel-measurable vector-valued functions f and g exist such that $f(\xi) = g(\eta)$.

Since $F_1 \subset F_2$ if and only if $L_{F_1}^2 \subset L_{F_2}^2$, therefore $\mathbf{S}(F_1', F_2') \leq \mathbf{S}(F_1, F_2)$, whenever $F_1' \subset F_1$ and $F_2' \subset F_2$ as it follows from 5° of Theorem 1.1'.

Especially for vector random variables ξ and η $\mathbf{S}(f(\xi), g(\eta)) \leqq \mathbf{S}(\xi, \eta)$ where f and g are any Borel-measurable vector-valued functions, because of $F_{f(\xi)} \subset F_\xi$ and $F_{g(\eta)} \subset F_\eta$.

4.4. In this section it is dealt with the pair of operators $\boldsymbol{P}_{F_2}^{F_1}$ and $\boldsymbol{P}_{F_1}^{F_2}$.

Theorem 4.1. *For two normed random variables* $f \in L^2_{F_1,0}$ *and* $g \in L^2_{F_2,0}$ *the following statements are equivalent:*

1° *f and g form a pair of eigenfunctions of the pair of operators* $\boldsymbol{P}_{F_2}^{F_1}$, $\boldsymbol{P}_{F_1}^{F_2}$.

2° $||\boldsymbol{P}_{F_2}^{F_1} f|| = ||\boldsymbol{P}_{F_1}^{F_2} g|| = (f, g)$.

3° *f and g are linearly correlated*[1] *and* $||P_g^f f|| = ||P_{F_2}^{F_1} f||$; $||\boldsymbol{P}_f^g g|| = ||\boldsymbol{P}_{F_1}^{F_2} g||$.

Proof. Analogous to that of Theorem 1 in [4].

In the sequel, let us examine under what conditions the operators $\boldsymbol{P}_{F_2}^{F_1}$ and $\boldsymbol{P}_{F_1}^{F_2}$ form a pair of integral operators. The Theorem on this subject will be formulated in general.

Let $(\Omega_1, F_1, \mathbf{P}_1)$ and $(\Omega_2, F_2, \mathbf{P}_2)$ be two probability spaces and let $\boldsymbol{T}$ be a linear bounded operator defined on $L^2_{F_1}$ and taking on its values in $L^2_{F_2}$, for which

1. $\boldsymbol{T}\chi_{\Omega_1} = \chi_{\Omega_1}$,
2. if $A \in F_1$ then $\boldsymbol{T}\chi_A \geqq 0$,
3. $\int_{\Omega_2} \boldsymbol{T} f d\,\mathbf{P}_2 = \int_{\Omega_1} f d\,\mathbf{P}_1$ for all $f \in L^2_{F_1}$,

where χ_A denotes the indicator function of any $A \in F_1$.

Theorem 4.2. *There hold*

1° $\boldsymbol{T}^* \chi_{\Omega_1} = \chi_{\Omega_1}$,

2° *if* $B \in F_2$ *then* $\boldsymbol{T}^* \chi_B \geqq 0$,

3° $\int_{\Omega_1} \boldsymbol{T}^* g d\,\mathbf{P}_1 = \int_{\Omega_2} g d\,\mathbf{P}_2$ *for all* $g \in L^2_{F_2}$.

(For the adjoint $\boldsymbol{T}^*$ see (1.1).)

Proof. According to Condition 3, $(\boldsymbol{T}^* \chi_{\Omega_2}, f)_1 = (\chi_{\Omega_2}, \boldsymbol{T} f)_2 = (\chi_{\Omega_1}, f)_1$ for all $f \in L^2_{F_1}$ therefore 1° follows.

Since

$$\int_A \boldsymbol{T}^* \chi_B d\mathbf{P}_1 = (\chi_A, \boldsymbol{T}^* \chi_B)_1 = (\boldsymbol{T}\chi_A, \chi_B)_2 = \int_B \boldsymbol{T}\chi_A d\mathbf{P}_2 \geqq 0 \text{ for any } A \in F_1,$$

according to Condition 2, thus $\boldsymbol{T}^* \chi_B \geqq 0$.

Finally,

$$\int_{\Omega_1} \boldsymbol{T}^* g d\,\mathbf{P}_1 = (\chi_{\Omega_1}, \boldsymbol{T}^* g)_1 = (\boldsymbol{T}\chi_{\Omega_1}, g)_2 = (\chi_{\Omega_2}, g)_2 = \int_{\Omega_2} g d\mathbf{P}_2 .$$

[1] The correlation of f and g is called linear, if both the regressions of them are linear.

The conditions imply that the set function

$$\mathbf{P}(A \times B) = (\boldsymbol{T}\,\chi_A, \chi_B)_2 \tag{4.6}$$

is non-negative on intervals of $\Omega = \Omega_1 \times \Omega_2$, hence it can be uniquely extended as a measure on $F = F_1 \times F_2$.

Thus, being $\mathbf{P}(\Omega) = 1$, $(\Omega, F, \mathbf{P})$ is a probability space, whereas

$$\mathbf{P}_1(A) = \mathbf{P}(A \times \Omega_2) \qquad \text{for all} \quad A \in F_1,$$

$$\mathbf{P}_2(B) = \mathbf{P}(\Omega_1 \times B) \qquad \text{for all} \quad B \in F_2$$

are the marginals of $\mathbf{P}$. Hence, if $f = f(x) \in L^2_{F_1}$ and $f_1(x, y) = f(x)$ for all $y \in \Omega_2$, then

$$\int_{\Omega_2}\!\!\int f_1^2\, d\mathbf{P} = \int_{\Omega_1} f^2\, d\mathbf{P}_1,$$

which means that $L^2_{F_1}$ can be considered as a subspace of $L^2\,(\Omega, F, \mathbf{P})$ consisting of its elements not depending on the elements of Ω_2 while F_1 as the sub-σ-algebra of the elements $A \times \Omega_2$. $L^2_{F_2}$ can be characterized in a similar way.

Theorem 4.3.

$$\boldsymbol{T} = \boldsymbol{P}^{L^2_{F_1}}_{L^2_{F_1}}. \tag{4.7}$$

Proof. (4.6) implies

$$(\chi_A, \chi_B) = (\boldsymbol{T}\,\chi_A, \chi_B)$$

for each $A \in F_1$ and $B \in F_2$; thus

$$(f, g) = (\boldsymbol{T} f, g) \tag{4.8}$$

whenever $f \in L^2_{F_1}$ and $g \in L^2_{F_2}$ are step functions; by the continuity of the operator $\boldsymbol{T}$, (4.8) holds for any $f \in L^2_{F_1}$ and $g \in L^2_{F_2}$, whence the statement follows from (1.5).

The operator $\boldsymbol{T}$ is called an integral operator with kernel $K(x, y)$ if

$$\boldsymbol{T} f = \int_{\Omega_1} K(x, y) f(x)\, d\mathbf{P}_1(x) \in L^2_{F_2} \quad \text{for all} \quad f \in L^2_{F_1}.$$

Theorem 4.4. *$\boldsymbol{T}$ is an integral operator with kernel $K(x, y)$ if and only if*

$$\mathbf{P}(E) = \int_E\!\!\int K(x, y)\, d\mathbf{P}_1(x)\, d\mathbf{P}_2(y) \qquad \text{for all } E \in F,$$

i.e. $\mathbf{P} \ll \mathbf{P}_1 \times \mathbf{P}_2$.

Proof. Analogous to that of Theorem 1 in [3].

§ 5. The intuitive background of the postulates

In section 1.3 three postulates were given. Let us examine what is their intuitive background in spaces L^2 from the view-point of stochastic connection between two random variables.

As a matter of fact, the meaning of intensity of connection is not at all unambiguously determined. Thus, no system of postulates can be adapt to characterize it in all possible senses. This section is to explain the sense corresponding to our system of postulates.

Let ξ and η be arbitrary standard random variables. As it is known, L^2_ξ and L^2_η are the spaces of functions (having finite standard deviations) of ξ and η, respectively; further $\boldsymbol{P}_\xi\, \eta$ represents the conditional expectation (regression) of η on ξ. If only the ξ-values are observed, the values of η may be inferred by the aid of $\boldsymbol{P}_\xi\, \eta$. Of course, the closer the connection between ξ and η, the better is the inference. On the other hand, the inference is better, when the root-mean-square error of it, i.e.

$$\|\eta - \boldsymbol{P}_\xi\, \eta\|, \tag{5.1}$$

is less. However, this standard deviation depends not only on the intensity of connection between ξ und η, but also on the scale, on which the values of η are measured, i.e. on the univalent transformations of η. Hence, the degree of dependence is regarded higher, when (5.1) may be lessened by normed univalent transformations of η, that is the quantity

$$\inf \|g - \boldsymbol{P}_\xi\, g\| \tag{5.2}$$

is less, where the infimum is taken for univalent $g \in L^2_{\eta,0}$, $\|g\| = 1$. Though (5.2) does not seem to be $\delta(L^2_{\xi,0}, L^2_{\eta,0})$, the following Theorem is true.

Theorem 5.1.

$$\sup\,(u, v) = \mathbf{S}(\xi, \eta)\,, \tag{5.3}$$

where the supremum is taken for $u \in L^2_{\xi,0}$ and $v \in L^2_{\eta,0}$ univalent functions of ξ and η, respectively, with $\|u\| = \|v\| = 1$.

Proof. Let $f_n \in L^2_{\xi,0}$ and $g_n \in L^2_{\eta,0}$ ($\|f_n\| = \|g_n\| = 1$) be such that $\lim\limits_{n=\infty} (f_n, g_n) = \mathbf{S}(\xi, \eta)$, further let $\varphi_{nm} \in L^2_{\xi,0}$, $\psi_{nm} \in L^2_{\eta,0}$ be step functions which take on only a finite number of values, such that $\lim\limits_{m=\infty} \|\varphi_{nm} - f_n\| = \lim\limits_{m=\infty} \|\psi_{nm} - g_n\| = 0$ for all n. Then $\lim\limits_{m=\infty} \|\varphi^*_{nm} - f_n\| = \lim\limits_{m=\infty} \|\psi^*_{nm} - g_n\| = 0$, too, where $\varphi^*_{nm} = \dfrac{\varphi_{nm}}{\|\varphi_{nm}\|}$, $\psi^*_{nm} = \dfrac{\psi_{nm}}{\|\psi_{nm}\|}$, for $\lim\limits_{m=\infty} \|\varphi^*_{nm} - \varphi_{nm}\|^2 = \lim\limits_{m=\infty} (1 - \|\varphi_{nm}\|)^2 = 0$ and

$$\lim_{m=\infty} \|\varphi^*_{nm} - f_n\| \leq \lim_{m=\infty} \|\varphi^*_{nm} - \varphi_{nm}\| + \lim_{m=\infty} \|\varphi_{nm} - f_n\| = 0\,,$$

which can be analogously verified for $\|\psi^*_{nm} - g\|$, too.

Let

$$u_{nmk} = \varepsilon_{nmk}\, t(\xi) + \varphi^*_{nm} \quad \text{and} \quad v_{nmk} = \varepsilon_{nmk}\, t(\eta) + \psi^*_{nm}\,,$$

where $t(x)$ is a univalent bounded measurable function: $|t(x)| \leq C$. The positive numbers ε_{nmk} are determined so that $\lim\limits_{k=\infty} \varepsilon_{nmk} = 0$ and $\varepsilon_{nmk} < \dfrac{\alpha_{nm}}{2C}$, where $\alpha_{nm} > 0$ are less than any of the absolute values of the non-zero difference of the values of the step functions φ^*_{nm}, ψ^*_{nm}. With this choice of ε_{nmk} the functions u_{nmk} and v_{nmk} are univalent, since if $u_{nmk}(x_1) = u_{nmk}(x_2)$, either

$$\alpha_{nm} < |\varphi^*_{nm}(x_1) - \varphi^*_{nm}(x_2)| = \varepsilon_{nmk}\, |t(x_1) - t(x_2)| < \frac{\alpha_{nm}}{2C} \cdot 2C = \alpha_{nm}\,,$$

which is impossible, or $\varphi^*_{nm}(x_1)—\varphi^*_{nm}(x_2)=0$, in which case $t(x_1)=t(x_2)$ and, owing to the univalence of $t(x)$, $x_1 = x_2$. Similarly, the functions v_{nmk} are also univalent. Because of

$$||u_{nmk} - \varphi^*_{nm}|| \leqq C\,\varepsilon_{nmk} \quad \text{and} \quad ||v_{nmk} - \psi^*_{nm}|| \leqq C\,\varepsilon_{nmk}\,,$$

and[2]

$$\lim_{k=\infty} ||u^*_{nmk} - u_{nmk}||^2 = \lim_{k=\infty} \left[(1 - ||u_{nmk} - \mathbf{M}(u_{nmk})||)^2 + \mathbf{M}^2(u_{nmk})\right] =$$

$$= \lim_{k=\infty} ||v^*_{nmk} - v_{nmk}||^2 = 0\,,$$

$\lim\limits_{k=\infty} ||u^*_{nmk} - \varphi^*_{nm}|| = \lim\limits_{k=\infty} ||v^*_{nmk} - \psi^*_{nm}|| = 0$ for all n and m, where u^*_{nmk} and v^*_{nmk} are the standardized of u_{nmk} and v_{nmk}.

Therefore, (u^*_{nmk}, v^*_{nmk}) may approach $\mathbf{S}(\xi, \eta)$ with arbitrary accuracy.

In consequence of this Theorem, (5.2) is equal to $\delta(L^2_{\xi,0}, L^2_{\eta,0})$. Accordingly, in 1.3 Postulate 1 asserts the closer the connection, the greater the maximal correlation.

Postulate 2 needs no commentary: it simply asserts, the maximal correlation being zero in the case of independence.

Postulate 3 expresses the case of "strict dependence". However, not only the case of ξ and/or η being a function of the other should be regarded as that of strict dependence, but also the more general type of dependence $f(\xi) = g(\eta)$ where f and g are non-constant Borel-measurable functions.
In this case

$$\delta(L^2_{\xi,0}, L^2_{\eta,0}) = 0\,. \tag{5.4}$$

Moreover, (5.4) may hold in the "irregular" cases, when there exist sequences of $f_n \in L^2_{\xi,0}$, $g_n \in L^2_{\eta,0}$ such that $||f_n|| = ||g_n|| = 1$ and $\lim\limits_{n=\infty} ||f_n—g_n|| = 0$, but no function of ξ is equal to a function of η. (See an example for this case in [10].) For sake of uniformity, these cases are also considered as those of "strict dependence".

A. Rényi [10] has given some Postulates for quantities measuring the intensity of stochastic dependence which are satisfied by the maximal correlation. If his Postulates are modified by substituting our Postulates 1 and 3 for these Postulates *F)* and *E)*, respectively, then the obtained system of postulates is satisfied by the maximal correlation only.

It is to be noted, that C. B. Bell [2] also dealt with the modification of A. Rényi's Postulates, but the maximal correlation does not satisfy his modified system.

§ 6. Particular cases and examples

In this section the notation $\xi = (\xi_1, \ldots, \xi_N)$ will be used for an N-dimensional and $\eta = (\eta_1, \ldots, \eta_M)$ for an M-dimensional random vector, where $N \leqq M$.

[2] Henceforth let $\mathbf{M}(\ )$ denote the expectation of the random variable in brackets.

6.1. Canonical distribution. The random vector (ξ, η) is said to be canonically distributed if the joint distribution function of it equals the product of the distribution functions of $(\xi_1, \eta_1), \ldots, (\xi_N, \eta_N)$ and of $\eta_{N+1}, \ldots, \eta_M$, i.e

$$(6.1)\ H(x_1, \ldots, x_N, y_1, \ldots, y_M) = H_1(x_1, y_1) \ldots H_N(x_N, y_N)\, H_{N+1}(y_{N+1}) \ldots H_M(y_M)\,.$$

Lemma 6.1. *For canonical distributions* $(\xi_{i_1}, \ldots, \xi_{i_k}, \eta_{j_1}, \ldots, \eta_{j_l})$ *is independent of* $(\xi_{i'_1}, \ldots, \xi_{i'_{k'}}, \eta_{j'_1}, \ldots, \eta_{j'_{l'}})$, *whenever all the indices* i' *and* j' *differ from all the indices* i *and* j. *Hence if* $f_1 \in L^2_{\xi_1}, \ldots, f_N \in L^2_{\xi_N}, g_1 \in L^2_{\eta_1}, \ldots, g_n \in L^2_{\eta_M}$ *then the products* $f_1 \ldots f_N \in L^2_\xi$ *and* $g_1 \ldots g_M \in L^2_\eta$, *further* $\|f_1 \ldots f_N\| = \|f_1\| \ldots \|f_N\|$, $\|g_1 \ldots g_M\| = \|g_1\| \ldots \|g_M\|$ *and* $(f_1 \cdots f_N, g_1, \ldots g_M) = (f_1, g_1) \ldots (f_N, g_N) \times \times \mathbf{M}(g_{N+1}) \ldots \mathbf{M}(g_M)$. *If* $\{f_{in_i}\}$ *and* $\{g_{jm_j}\}$ *are complete orthonormal systems in* $L^2_{\xi_i}$ *resp.* $L^2_{\eta_j}$ $(i = 1, \ldots, N;\ j = 1, \ldots, M)$ *then the sets of all the possible products* $\{f_{1n_1} \cdots f_{Nn_N}\}$ *and* $\{g_{1m_1} \ldots g_{Mm_M}\}$ *are complete orthonormal systems in* L^2_ξ, *resp.* L^2_η.

Proof. The statements may be verified directly from the definition of the canonical distribution.

Lemma 6.2. *For canonical distributions*

$$(6.2)\qquad \boldsymbol{P}_\eta f_1 \ldots f_N = \boldsymbol{P}_{\eta_1} f_1 \ldots \boldsymbol{P}_{\eta_N} f_N,$$

$$(6.3)\qquad \boldsymbol{P}_\xi g_1 \ldots g_M = \boldsymbol{P}_{\xi_1} g_1 \ldots \boldsymbol{P}_{\xi_N} g_N \mathbf{M}(g_{N+1}) \ldots \mathbf{M}(g_M),$$

where $f_i \in L^2_{\xi_i}$ *and* $g_j \in L^2_{\eta_j}$ $(i = 1, \ldots, N;\ j = 1, \ldots, M)$.

Proof. Let $\{g_{jm_j}\}$ be complete orthonormal systems in $L^2_{\eta_j}$ $(j = 1, \ldots, M)$ and $g = \sum\limits_{m_1, \ldots, m_M} \beta_{m \ldots m_M} g_{1m_1} \ldots g_{Mm_M}$ an arbitrary element of L^2_η. Thus

$$(\boldsymbol{P}_{\eta_1} f_1 \ldots \boldsymbol{P}_{\eta_N} f_N, g) = \sum_{m_1, \ldots, m_M} \beta_{m_1 \ldots m_M} (\boldsymbol{P}_{\eta_1} f_1 \ldots \boldsymbol{P}_{\eta_N} f_N, g_{1m_1} \ldots g_{Mm_M}) =$$

$$= \sum_{m_1, \ldots, m_M} \beta_{m_1 \ldots m_M} (f_1, g_{1m_1}) \ldots (f_N, g_{Nm_N})\, \mathbf{M}(g_{N+1, m_{N+1}}) \ldots \mathbf{M}(g_{Mm_M}) =$$

$$= (f_1 \cdots f_N, g)$$

which proves (6.2). The proof of (6.3) is analogous.

Theorem 6.1. *Let* $f_{i_1}, g_{i_1}, \ldots, f_{i_k}, g_{i_k}$ *be pairs of eigenfunctions of* $\boldsymbol{P}^{\xi_{i_1}}_{\eta_{i_1}}, \boldsymbol{P}^{\eta_{i_1}}_{\xi_{i_1}}, \ldots, \boldsymbol{P}^{\xi_{i_k}}_{\eta_{i_k}}, \boldsymbol{P}^{\eta_{i_k}}_{\xi_{i_k}}$ *and belong to the respective eigenvalues* $\lambda_{i_1}, \ldots, \lambda_{i_k}$ *of a canonical distribution. Then the pair of products* $f_{i_1} \ldots f_{i_k}$, $g_{i_1} \ldots g_{i_k}$ *forms a pair of eigenfunctions of* $\boldsymbol{P}^\xi_\eta, \boldsymbol{P}^\eta_\xi$ *and belongs to the product* $\lambda_{i_1} \ldots \lambda_{i_k}$ *for any* k^{th} *order combination* $(i_1, \ldots, i_k)$ *of the elements* $1, \ldots, N$.

Proof. According to Lemma (6.2),

$$\boldsymbol{P}^\xi_\eta f_{i_1} \ldots f_{i_k} = \boldsymbol{P}^{\xi_{i_1}}_{\eta_{i_1}} f_{i_1} \ldots \boldsymbol{P}^{\xi_{i_k}}_{\eta_{i_k}} f_{i_k} = \lambda_{i_1} \ldots \lambda_{i_k} g_{i_1} \ldots g_{i_k},$$

$$\boldsymbol{P}^\eta_\xi g_{i_1} \ldots g_{i_k} = \boldsymbol{P}^{\eta_{i_1}}_{\xi_{i_1}} g_{i_1} \ldots \boldsymbol{P}^{\eta_{i_k}}_{\xi_{i_k}} g_{i_k} = \lambda_{i_1} \ldots \lambda_{i_k} f_{i_1} \ldots f_{i_k}.$$

Theorem 6.2. *For canonical distributions*

$$(6.4)\qquad \mathbf{S}(\xi, \eta) = \max_{i=1, \ldots, N} \mathbf{S}(\xi_i, \eta_i).$$

Proof. Let $\{f_{in_i}\}$ and $\{g_{jn_j}\}$ be complete orthonormal systems in $L^2_{\xi_i}$ resp. $L^2_{\eta_j}$, such that $f_{i0} = g_{j0} = \chi_\Omega$ $(i = 1, \ldots, N;\ j = 1, \ldots, M)$. Let further $f = \sum_{n_1+\ldots+n_N>0} \alpha_{n_1} \cdots {}_{n_N} f_{1n_1} \cdots f_{Nn_N} \in L^2_{\xi,0}$ and $g = \sum_{m_1+\ldots+m_M>0} \beta_{m_1} \cdots {}_{m_M} g_{1m_1} \cdots g_{Mm_M} \in L^2_{\eta,0}$ be arbitrary normed random variables. In this case one obtains

$$|(f, g)| = \Big| \sum_{n_1+\ldots+n_N>0} \sum_{m_1+\ldots+m_M>0} \alpha_{n_1\ldots n_N} \beta_{m_1\ldots m_M} (f_{1n_1}, g_{1m_1}) \times \cdots \times$$

$$\times (f_{Nn_N}, g_{Nm_N})\, \mathbf{M}(g_{N+1, m_{N+1}}) \cdots \mathbf{M}(g_{Mm_M}) \Big| \leqq$$

$$\leqq \sum_{\substack{n_1, \ldots, n_N \\ n_1>0}} \sum_{\substack{m_1, \ldots, m_N \\ m_1>0}} |\alpha_{n_1\ldots n_N} \beta_{m_1\ldots m_N 0\ldots 0} (f_{1n_1}, g_{1m_1}) \cdots (f_{Nn_N}, g_{Nm_N})| + \cdots +$$

$$+ \sum_{n_N>0} \sum_{m_N>0} |\alpha_{0\ldots 0 n_N} \beta_{0\ldots 0 m_N 0\ldots 0} (f_{Nn_N}, g_{Nm_N})| \leqq$$

$$\leqq \mathbf{S}(\xi_1, \eta_1) \sum_{\substack{n_1, \ldots, n_N \\ n_1>0}} \sum_{\substack{m_1, \ldots, m_N \\ m_1>0}} |\alpha_{n_1\ldots n_N} \beta_{m_1\ldots m_N 0\ldots 0}| + \cdots +$$

$$+ \mathbf{S}(\xi_N, \eta_N) \sum_{n_N>0} \sum_{m_N>0} |\alpha_{0\ldots 0 n_N} \beta_{0\ldots 0 m_N 0\ldots 0}| \leqq$$

$$\leqq \max_{i=1,\ldots,N} \mathbf{S}(\xi_i, \eta_i) \sum_{n_1+\ldots+n_N>0} \sum_{m_1+\ldots+m_N>0} |\alpha_{n_1\ldots n_N} \beta_{m_1\ldots m_N 0\ldots 0}| \leqq \max_{i=1,\ldots,N} \mathbf{S}(\xi_i, \eta_i)\,.$$

On the other hand, $\mathbf{S}(\xi_i, \eta_i) \leqq \mathbf{S}(\xi, \eta)$ $(i = 1, \ldots, N)$ is trivial, thus (6.4) is true.

Theorem 6.3. *For canonical distributions*

$$1 + \mathbf{C}^2(\xi, \eta) = [1 + \mathbf{C}^2(\xi_1, \eta_1)] \cdots [1 + \mathbf{C}^2(\xi_N, \eta_N)] \tag{6.5}$$

where $\mathbf{C}(\zeta_1, \zeta_2) = |||\, P^{\zeta_1}_{\zeta_2} \,|||$ *is the mean-square contingency of (vector) random variables* ζ_1 *and* ζ_2.

Proof. If $\{f_{in_i}\}$ and $\{g_{jn_j}\}$ are complete orthonormal systems in L^2_ξ resp. L^2_η such that $f_{i0} = g_{j0} = \chi_\Omega$ $(i = 1, \ldots, N;\ j = 1, \ldots, M)$ then from (1.15)

$$1 + \mathbf{C}^2(\xi, \eta) = \sum_{\substack{n_1, \ldots, n_N \\ m_1, \ldots, m_M}} (f_{1n_1} \cdots f_{Nn_N}, g_{1m_1} \cdots g_{Mm_M})^2 =$$

$$= \sum_{n_1, m_1} (f_{1n_1}, g_{1m_1})^2 \cdots \sum_{n_N, m_N} (f_{Nm_N}, g_{Nm_N})^2 = [1 + \mathbf{C}^2(\xi_1, \eta_1)] \cdots [1 + \mathbf{C}^2(\xi_N, \eta_N)]\,.$$

Theorem 6.4. *If the eigenvalues* λ_{in_i} *for* $n_i > 0$ *belong to pairs of eigenfunctions of* $P^{\xi_i}_{\eta_i}$, $P^{\eta_i}_{\xi_i}$ $(i = 1, \ldots, N)$ *furnishing complete orthogonal systems in* $L^2_{\xi_i, 0}$ *and* $L^2_{\eta_i, 0}$, *respectively, further if* $\lambda_{i0} = 1$ *then the sequence* $\{\lambda_{1n_1} \ldots \lambda_{Nn_N}\}$ *contains all the non-zero eigenvalues of* P^ξ_η, P^η_ξ.

Proof. In consequence of (1.4) and (6.5)

$$1 + \sum_{n_1+\ldots+n_N>0} \lambda^2_{1n_1} \cdots \lambda^2_{Nn_N} = \big(1 + \sum_{n_1>0} \lambda^2_{1n_1}\big) \cdots \big(1 + \sum_{n_N>0} \lambda^2_{Nn_N}\big) =$$

$$= [1 + \mathbf{C}^2(\xi_1, \eta_1)] \cdots [1 + \mathbf{C}^2(\xi_N, \eta_N)] = 1 + \mathbf{C}^2(\xi, \eta)\,.$$

Regarding Theorem 6.1, this proves the statement.

6.2. Normal distribution. Let the joint distribution of (ξ, η) be an $N + M$-dimensional Gaussian one with the covariance matrix

$$\Sigma = \begin{bmatrix} \Sigma_{11} & \Sigma_{12} \\ \Sigma_{12}' & \Sigma_{22} \end{bmatrix}, \tag{6.6}$$

where Σ_{11} is the covariance matrix of ξ, Σ_{22} that of η; the elements of Σ_{12} are the covariances between the components of ξ and those of η. (Σ_{12}' means the transposed of Σ_{12}.)

If ξ is degenerately distributed i.e. the determinant of Σ_{11} vanishes then some components of ξ can be expressed in terms of the others. Omitting these components, the space L_ξ^2 remains unchanged. Thus a non-degenerate distribution of ξ, and similarly of η, may be supposed without restricting generality.

There can be made use of the known fact that a non-degenerate $N + M$-dimensional normal random vector (ξ, η) can be transformed by non-degenerate linear transformations of its N-dimensional component ξ and of its M-dimensional component η so that the obtained random vector (ξ', η') be of $N + M$-dimensional zero-vector expectation and of covariance matrix

$$\begin{bmatrix} 1 & \dots & 0 & \varrho_1 & \dots & 0 & 0 & \dots & 0 \\ \vdots & \ddots & & \vdots & \ddots & & \vdots & & \\ 0 & \dots & 1 & 0 & \dots & \varrho_N & 0 & \dots & 0 \\ \varrho_1 & \dots & 0 & 1 & \dots & 0 & 0 & \dots & 0 \\ \vdots & \ddots & & \vdots & \ddots & & \vdots & & \\ 0 & \dots & \varrho_N & 0 & \dots & 1 & 0 & \dots & 0 \\ 0 & \dots & 0 & 0 & \dots & 0 & 1 & \dots & 0 \\ \vdots & & & \vdots & & & \vdots & \ddots & \\ 0 & \dots & 0 & 0 & \dots & 0 & 0 & \dots & 1 \end{bmatrix} \qquad \varrho_1 \geqq \dots \geqq \varrho_N \geqq 0. \tag{6.7}$$

This means that the components of ξ' and η' are standard normal random variables and that their joint distribution is canonical (see H. Hotelling [6]). The correlation coefficients $\varrho_1, \dots, \varrho_N$ are called the canonical correlations of ξ and η. Thus, $L_{\xi'}^2 = L_\xi^2$ and $L_{\eta'}^2 = L_\eta^2$ holding, ξ, η can be taken as canonically distributed with standard components without any restriction of generality.

Theorem 6.5. *The maximal correlation of normally distributed random vector variables is equal to their greatest canonical correlation.*

Proof. The normally distributed vector (ξ, η) with covariance matrix (6.7) — according to $\mathbf{S}(\xi_i, \eta_i) = \varrho_i$ $(i = 1, \dots, N)$ and Theorem 6.2 — has

$$\mathbf{S}(\xi, \eta) = \varrho_1,$$

from which the statement follows.

When transforming the normal distribution with covariance matrix (6.6) into a canonical distribution, it is to see that $\mathbf{S}(\xi, \eta)$ is the greatest root of the equation

$$\det \begin{bmatrix} \varrho \Sigma_{11} & \Sigma_{12} \\ \Sigma'_{12} & \varrho \Sigma_{22} \end{bmatrix} = 0$$

and that the pair of linear eigenfunctions belonging to $\mathbf{S}(\xi, \eta)$ is

$$f = \sum_{i=1}^{N} \alpha_i(\xi_i - \mathbf{M}(\xi_i)), \quad g = \sum_{i=1}^{M} \beta_j(\eta_j - \mathbf{M}(\eta_j)),$$

where $\alpha = (\alpha_1, \ldots, \alpha_N)$ and $\beta = (\beta_1, \ldots, \beta_M)$ are the vectors satisfying the equations

$$\Sigma'_{12}\, \alpha = \mathbf{S}(\xi, \eta)\, \Sigma_{22}\, \beta$$

$$\Sigma_{12}\, \beta = \mathbf{S}(\xi, \eta)\, \Sigma_{11}\, \alpha.$$

6.3. Multinominal distribution. Let the joint distribution of (ξ, η) be

$$p_{i_1 \ldots i_N j_1 \ldots j_M} = \frac{L!}{i_1! \ldots i_N!\, j_1! \ldots j_M! (L - I - J)!} p_1^{i_1} \ldots p_N^{i_N} q_1^{j_1} \ldots q_M^{j_M} (1 - P - Q)^{L-I-J},$$

where $i_1, \ldots, i_N, j_1, \ldots, j_M$ are non-negative integers, L is a positive one and $p_1, \ldots, p_N, q_1, \ldots, q_M$ are non-negative numbers, further

$$I = i_1 + \ldots + i_N, \quad J = j_1 + \ldots + j_M, \quad I + J \leq L,$$

$$P = p_1 + \ldots + p_N, \quad Q = q_1 + \ldots + q_M, \quad P + Q \leq 1.$$

The marginal and the conditional probabilities are

$$p_{i_1 \ldots i_N \cdot} = \frac{L!}{i_1! \ldots i_N! (L - I)!} p_1^{i_1} \ldots p_N^{i_N} (1 - P)^{L-I} \qquad I \leq L,$$

$$p_{\cdot j_1 \ldots j_M} = \frac{L!}{j_1! \ldots j_M! (L - J)!} q_1^{j_1} \ldots q_M^{j_M} (1 - Q)^{L-J} \qquad J \leq L;$$

$$p_{i_1 \ldots i_N \cdot | j_1 \ldots j_M} = \frac{(L-J)!}{i_1! \ldots i_N! (L - J - I)!} \left(\frac{p_1}{1-Q}\right)^{i_1} \ldots \left(\frac{p_N}{1-Q}\right)^{i_N} \left(1 - \frac{P}{1-Q}\right)^{L-J-I}$$

$$I \leq L - J; \; J \leq L,$$

$$p_{\cdot j_1 \ldots j_M | i_1 \ldots i_N} = \frac{(L-I)!}{j_1! \ldots j_M! (L - I - J)!} \left(\frac{q_1}{1-P}\right)^{j_1} \ldots \left(\frac{q_M}{1-P}\right)^{j_M} \left(1 - \frac{Q}{1-P}\right)^{L-I-J}$$

$$J \leq L - I; \; I \leq L.$$

In order to calculate the maximal correlation for such distributions the formule

$$\sum_{i=0}^{L} i^n \binom{L}{i} p^i (1-p)^{L-i} = \sum_{j=1}^{n} a_{jn}\, p^j \frac{L!}{(L-j)!} \qquad n = 1, 2, \ldots \tag{6.8}$$

is to be applied where L is a positive integer, $0 \leq p \leq 1$ and a_{jn} are appropriate coefficients with $a_{nn} = 1$.

Theorem 6.6. *The non-zero eigenvalues of the pair of operators* $\mathbf{P}^{\xi}_{\eta}$, $\mathbf{P}^{\eta}_{\xi}$ *are*

$$\lambda_n = \left[\frac{PQ}{(1-P)(1-Q)}\right]^{\frac{n}{2}} \qquad n = 1, \ldots, L \tag{6.9}$$

and the maximal correlation

$$\mathbf{S}(\xi, \eta) = \left[\frac{PQ}{(1-P)(1-Q)}\right]^{\frac{1}{2}}. \tag{6.10}$$

Proof. In consequence of

$$\sum_{i_1+\ldots+i_N \leq L-J} (i_1 + \ldots + i_N)^n p_{i_1\ldots i_N \cdot | j_1 \ldots j_M} =$$

$$= \sum_{I=0}^{L-J} I^n \binom{L-J}{I} \left(1 - \frac{P}{1-Q}\right)^{L-J-I} \sum_{i_1+\ldots+i_N=I} \frac{I!}{i_1! \ldots i_N!} \left(\frac{p_1}{1-Q}\right)^{i_1} \cdots \left(\frac{p_N}{1-Q}\right)^{i_N} =$$

$$= \sum_{I=0}^{L-J} I^n \binom{L-J}{I} \left(\frac{P}{1-Q}\right)^I \left(1 - \frac{P}{1-Q}\right)^{L-J-I} = \sum_{j=1}^{n} a_{jn} \left(\frac{P}{1-Q}\right)^j \frac{(L-J)!}{(L-J-j)!}$$

the equality

$$\mathbf{P}^{\xi}_{\eta}(\varphi_n - \mathbf{M}(\varphi_n)) = \sum_{j=1}^{n} a_{jn} \left(\frac{P}{1-Q}\right)^j (L - \psi_1) \ldots (L - \psi_1 - j + 1) - \mathbf{M}(\varphi_n)$$
$$n = 1, \ldots, L$$

and analogously the equality

$$\mathbf{P}^{\eta}_{\xi}(\psi_n - \mathbf{M}(\psi_n)) = \sum_{j=1}^{n} a_{jn} \left(\frac{Q}{1-P}\right)^j (L - \varphi_1) \ldots (L - \varphi_1 - j + 1) - \mathbf{M}(\psi_n)$$
$$n = 1, \ldots, L$$

hold, where

$$\varphi_n = (\xi_1 + \ldots + \xi_N)^n \quad \text{and} \quad \psi_n = (\eta_1 + \ldots + \eta_M)^n \quad n = 1, \ldots, L.$$

Here a generalized form of Theorem 3 of [4] may be applied. Viz. as it is obvious, this Theorem may be generalized to any pairs of Hilbert spaces and any pair of operators adjoint to each other: the proof is analogous. Thus (6.9) supplies eigenvalues.

With the aid of identity (1.4) it is proved that (6.9) provides all the non-zero eigenvalues:

$$1 + \mathbf{C}^2(\xi, \eta) = \sum_{i_1+\ldots+i_N+j_1+\ldots+j_M \leq L} \frac{p^2_{i_1\ldots i_N j_1 \ldots j_M}}{p_{i_1\ldots i_N \cdot}\, p_{\cdot j_1 \ldots j_M}} =$$

$$= \sum_{I+J \leq L} \frac{(L-I)!\,(L-J)!}{I!\,J!\,(L-I-J)!^2} \left(1 - \frac{P}{1-Q}\right)^{L-I-J} \left(1 - \frac{Q}{1-P}\right)^{L-I-J} \times$$

$$\times \sum_{\substack{i_1+\ldots+i_N=I \\ j_1+\ldots+j_M=J}} \frac{I!\,J!}{i_1!\ldots i_N!\,j_1!\ldots j_M!} \left(\frac{p_1}{1-Q}\right)^{i_1}\cdots\left(\frac{p_N}{1-Q}\right)^{i_N}\left(\frac{q_1}{1-P}\right)^{j_1}\cdots\left(\frac{q_M}{1-P}\right)^{j_M} =$$

$$= \sum_{I+J\leq L} \frac{(L-I)!\,(L-J)!}{I!\,J!\,(L-I-J)!^2}\left(\frac{P}{1-Q}\right)^{L}\left(1-\frac{P}{1-Q}\right)^{L-I-J}\left(\frac{Q}{1-P}\right)^{J}\times$$

$$\times\left(1-\frac{Q}{1-P}\right)^{L-I-J} = \sum_{\substack{I+J\leq L \\ \leq k\leq L-I-J \\ 0\leq l\leq L-I-J}} \frac{(L-I)!\,(L-J)!}{I!\,J!\,(L-I-J-k)!\,(L-I-J-l)!\,k!\,l!}\times$$

$$\times(-1)^{k+l}\left(\frac{P}{1-Q}\right)^{I+k}\left(\frac{Q}{1-P}\right)^{J+l} = \sum_{m=0}^{L}\sum_{n=0}^{L}\left(\frac{P}{1-Q}\right)^{m}\left(\frac{Q}{1-P}\right)^{n}\sum_{I=0}^{m}\binom{L-n}{I}\times$$

$$\times\binom{-L+m-1}{m-I}\sum_{J=0}^{n}\binom{L-m}{J}\binom{-L+n-1}{n-J} = \sum_{m=0}^{L}\sum_{n=0}^{L}(-1)^{m+n}\binom{n}{m}\binom{m}{n}\times$$

$$\times\left(\frac{P}{1-Q}\right)^{m}\left(\frac{Q}{1-P}\right)^{n} = \sum_{n=0}^{L}\left[\frac{PQ}{(1-P)(1-Q)}\right]^{n} = 1+\sum_{n=1}^{L}\lambda_n^2\,.$$

Therefore (6.10) holds, too.

6.4. Multihypergeometric distribution. Let the joint distribution of $(\xi,\ \eta)$ be

$$p_{i_1\ldots i_N j_1\ldots j_M} = \frac{\binom{A_1}{i_1}\cdots\binom{A_N}{i_N}\binom{B_1}{j_1}\cdots\binom{B_M}{j_M}\binom{L-A-B}{k-I-J}}{\binom{L}{k}}\,,$$

where $i_1, \ldots, i_N,\ j_1, \ldots, j_M$ are non-negative integers, $A_1, \ldots, A_N,\ B_1, \ldots B_M$, L and k are positive ones, further

$$I = i_1 + \ldots + i_N\ ,\quad J = j_1 + \ldots + j_M\ ,\quad I + J \leq k\,,$$

$$A = A_1 + \ldots + A_N\,,\quad B = B_1 + \ldots + B_M\,,\quad A + B \leq L$$

and $k \leq L$.

The marginal and the conditional probabilities are

$$p_{i_1\ldots i_N\cdot} = \frac{\binom{A_1}{i_1}\cdots\binom{A_N}{i_N}\binom{L-A}{k-I}}{\binom{L}{k}} \qquad I \leq k\,,$$

$$p_{\cdot j_1 \ldots j_M} = \frac{\binom{B_1}{j_1} \cdots \binom{B_M}{j_M} \binom{L-B}{k-J}}{\binom{L}{k}} \qquad J \leqq k\,,$$

$$p_{i_1 \ldots i_N \cdot | j_1 \ldots j_M} = \frac{\binom{A_1}{i_1} \cdots \binom{A_N}{i_N} \binom{L-B-A}{k-J-I}}{\binom{L-B}{k-J}} \qquad I \leqq k-J;\ J \leqq k\,,$$

$$p_{\cdot j_1 \ldots j_M | i_1 \ldots i_N} = \frac{\binom{B_1}{j_1} \cdots \binom{B_M}{j_M} \binom{L-A-B}{k-I-J}}{\binom{L-A}{k-I}} \qquad J \leqq k-I;\ I \leqq k\,.$$

Analogously to (6.8) in the preceding example, in the present case

$$\sum_{i=0}^{k} i^n \frac{\binom{L_1}{i}\binom{L-L_1}{k-i}}{\binom{L}{k}} = \sum_{j=1}^{n} a_{jn} \frac{\binom{L_1}{j}}{\binom{L}{j}} \frac{k!}{(k-j)!} \qquad n = 1, 2, \ldots$$

is needed where $k \leqq L$ are positive integers and a_{jn} are appropriate coefficients with $a_{nn} = 1$.

Theorem 6.7. *The non-zero eigenvalues of the pair of operators* $\mathbf{P}_\eta^\xi$, $\mathbf{P}_\xi^\eta$ *are*

$$\lambda_n = \left[\frac{\binom{A}{n}\binom{B}{n}}{\binom{L-A}{n}\binom{L-B}{n}}\right]^{\frac{1}{2}} \qquad n = 1, \ldots, k \tag{6.11}$$

and the maximal correlation

$$\mathbf{S}(\xi, \eta) = \left[\frac{AB}{(L-A)(L-B)}\right]^{\frac{1}{2}}. \tag{6.12}$$

Proof. In consequence of

$$\sum_{i_1+\ldots+i_N \leqq k-J} (i_1 + \ldots + i_N)^n\, p_{i_1 \ldots i_N \cdot | j_1 \ldots j_M} =$$

$$= \sum_{I=0}^{k-J} I^n \frac{\binom{L-B-A}{k-J-I}}{\binom{L-B}{k-J}} \sum_{i_1+\ldots+i_N=I} \binom{A_1}{i_1} \cdots \binom{A_N}{i_N} = \sum_{j=1}^{n} a_{jn} \frac{\binom{A}{j}}{\binom{L-B}{j}} \frac{(k-J)!}{(k-J-j)!}\,,$$

the equality

$$\boldsymbol{P}^{\xi}_{\eta}(\varphi_n - \mathbf{M}(\varphi_n)) = \sum_{j=1}^{n} a_{jn} \frac{\binom{A}{j}}{\binom{L-B}{j}} (k-\psi_1)\ldots(k-\psi_1-j+1) - \mathbf{M}(\varphi_n) \qquad n = 1, \ldots, k$$

and analogously the equality

$$\boldsymbol{P}^{\eta}_{\xi}(\psi_n - \mathbf{M}(\psi_n)) = \sum_{j=1}^{n} a_{jn} \frac{\binom{B}{J}}{\binom{L-A}{j}} (k-\varphi_1)\ldots(k-\varphi_1-j+1) - \mathbf{M}(\psi_n) \qquad n = 1, \ldots, k\,,$$

hold, where

$$\varphi_n = (\xi_1 + \ldots + \xi_N)^n \text{ and } \psi_n = (\eta_1 + \ldots + \eta_M)^n \qquad n = 1, \ldots, k\,.$$

Analogously to the previous example, here (6.11) provides eigenvalues. Finally, we prove that these are all the non-zero eigenvalues:

$$\sum_{i_1+\ldots+i_N+j_1+\ldots+j_M\leq k} \frac{p^2_{i_1\ldots i_N j_1\ldots j_M}}{p_{i_1\ldots i_N\cdot}\, p_{\cdot j_1\ldots j_M}} =$$

$$= \sum_{I+J\leq k} \frac{\binom{L-A-B}{k-I-J}^2}{\binom{L-A}{k-I}\binom{L-B}{k-J}} \sum_{\substack{i_1+\ldots+i_N=I\\ j_1+\ldots+j_M=J}} \binom{A_1}{i_1}\cdots\binom{A_N}{i_N}\binom{B_1}{j_1}\cdots\binom{B_M}{j_M} =$$

$$= \sum_{I+J\leq k} \frac{\binom{A}{I}\binom{B}{J}\binom{L-A-B}{k-I-J}^2}{\binom{L-A}{k-I}\binom{L-B}{k-J}} = 1 + \sum_{n=1}^{k} \lambda_n^2.$$

The last equality is the consequence of the fact, that neither side depends on N and M and that the left side of it is equal to the left side of the first equality for $N = M = 1$. Namely, in this particular case the left side of the last equality is equal to $1 + \sum_{n=1}^{k} \lambda_n^2$, since both the dimensions of $L^2_{\xi,0}$ and $L^2_{\eta,0}$ equal k, the number of possible pairs of eigenfunctions.

6.5. The distribution of (ξ, η) having the joint frequency function

$$h(x_1,\ldots, x_N, y_1,\ldots, y_M) = \begin{cases} \dfrac{1}{t} & \text{if } (x_1^2 + \ldots + x_N^2)^{\frac{1}{2}} + (y_1^2 + \ldots + y_M^2)^{\frac{1}{2}} \leq 1 \\ 0 & \text{otherwise}, \end{cases}$$

where $p > 0$, $q > 0$ and

$$t = \int\cdots\int_{(x_1^2+\ldots+x_N^2)^{\frac{p}{2}} + (y_1^2+\ldots+y_M^2)^{\frac{q}{2}} \leq 1} dx_1 \ldots dx_N\, dy_1 \ldots dy_M\,.$$

The marginal and the conditional frequency functions are

$$h_{1\dots N\cdot}(x_1, \dots, x_N) = \frac{\pi^{\frac{M}{2}}}{t\,\Gamma\left(\frac{M}{2}+1\right)}\left[1-(x_1^2+\dots+x_N^2)^{\frac{p}{2}}\right]^{\frac{M}{q}};\; x_1^2+\dots+x_N^2 \leqq 1,$$

$$h_{\cdot 1\dots M}(y_1, \dots, y_M) = \frac{\pi^{\frac{N}{2}}}{t\,\Gamma\left(\frac{N}{2}+1\right)}\left[1-(y_1^2+\dots+y_M^2)^{\frac{q}{2}}\right]^{\frac{N}{p}};\; y_1^2+\dots+y_M^2 \leqq 1;$$

$$h_{1\dots N\cdot|1\dots M}(x_1, \dots, x_N \,|\, y_1, \dots, y_M) = \frac{\Gamma\left(\frac{N}{2}+1\right)}{\pi^{\frac{N}{2}}}\left[1-(y_1^2+\dots+y_M^2)^{\frac{q}{2}}\right]^{-\frac{N}{p}};$$

$$x_1^2+\dots+x_N^2 \leqq \left[1-(y_1^2+\dots+y_N^2)^{\frac{q}{2}}\right]^{\frac{2}{p}},\; y_1^2+\dots+y_M^2 \leqq 1,$$

$$h_{\cdot 1\dots M|1\dots N}(y_1, \dots, y_M \,|\, x_1, \dots, x_N) = \frac{\Gamma\left(\frac{M}{2}+1\right)}{\pi^{\frac{M}{2}}}\left[1-(x_1^2+\dots+x_N^2)^{\frac{q}{2}}\right]^{-\frac{M}{q}};$$

$$y_1^2+\dots+y_M^2 \leqq \left[1-(x_1^2+\dots+x_N^2)^{\frac{p}{2}}\right]^{\frac{2}{q}},\; x_1^2+\dots+x_N^2 \leqq 1.$$

Theorem 6.8. *The non-zero eigenvalues of the pair of operators $\boldsymbol{P}_\eta^\xi$, $\boldsymbol{P}_\xi^\eta$ are*

$$\lambda_n = \left[\frac{NM}{(pn+N)(qn+M)}\right]^{\frac{1}{2}} \qquad n = 1, 2, \dots \tag{6.13}$$

and the maximal correlation

$$\mathbf{S}(\xi, \eta) = \left[\frac{NM}{(p+N)(q+M)}\right]^{\frac{1}{2}}. \tag{6.14}$$

Proof. In consequence of

$$\underset{x_1^2+\dots+x_N^2 \leqq [1-(y_1^2+\dots+y_M^2)^{\frac{q}{2}}]^{\frac{2}{p}}}{\int\cdots\int} (x_1^2+\dots+x_N^2)^{\frac{pn}{2}} h_{1\dots N\cdot|1\dots M}(x_1, \dots, x_N|y_1, \dots, y_M)\times$$

$$\times\, dx_1 \dots dx_N = \frac{N}{pn+N}\left[1-(y_1^2+\dots+y_M^2)^{\frac{q}{2}}\right]^n,$$

the equality

$$\boldsymbol{P}_\eta^\xi(\varphi_n - \mathbf{M}(\varphi_n)) = \frac{N}{pn+N}(1-\psi_1)^n - \mathbf{M}(\varphi_n) \qquad n = 1, 2, \dots$$

and analogously the equality

$$\boldsymbol{P}_\xi^\eta(\psi_n - \mathbf{M}(\psi_n)) = \frac{M}{qn+M}(1-\varphi_1)^n - \mathbf{M}(\psi_n) \qquad n = 1, 2, \dots,$$

hold, where

$$\varphi_n = (\xi_1^2 + \ldots + \xi_N^2)^{\frac{pn}{2}} \quad \text{and} \quad \psi_n = (\eta_1^2 + \ldots + \eta_M^2)^{\frac{qn}{2}} \quad n = 1, 2, \ldots$$

As in the above examples, here (6.13) are eigenvalues and (6. 14) holds, for (6. 13) gives all the non-zero eigenvalues:

$$\int\limits_{(x_1^2+\ldots+x_N^2)^{\frac{p}{2}}+(y_1^2+\ldots+y_M^2)^{\frac{q}{2}}\leq 1}\cdots\int \frac{h^2(x_1, \ldots, x_M, y_1, \ldots, y_M)}{h_{1\ldots N\cdot}(x_1, \ldots, x_N)\, h_{\cdot 1\ldots M}(y_1, \ldots, y_M)} \times$$

$$\times\, dx_1 \ldots dx_N\, dy_1 \ldots dy_M = \frac{M\,\Gamma\left(\frac{N}{2}+1\right)}{\pi^{\frac{N}{2}}} \int\limits_{x_1^2+\ldots+x_N^2\leq 1}\cdots\int \left[1-(x_1^2+\ldots+x_N^2)^{\frac{p}{2}}\right]^{-\frac{M}{q}} \times$$

$$\times \int\limits_0^{[1-(x_1^2+\ldots+x_N^2)^{\frac{p}{2}}]^{\frac{1}{q}}} (1-\sigma^q)^{-\frac{N}{p}}\, \sigma^{M-1}\, d\sigma\, dx_1 \ldots dx_N =$$

$$= \frac{M\,\Gamma\left(\frac{N}{2}+1\right)}{\pi^{\frac{N}{2}}} \sum_{n=0}^{\infty} (-1)^n \binom{-\frac{N}{P}}{n} \frac{1}{qn+M} \times$$

$$\times \int\limits_{x_1^2+\ldots+x_N^2\leq 1}\cdots\int \left[1-(x_1^2+\ldots+x_N^2)^{\frac{p}{2}}\right]^n dx_1 \ldots dx_N =$$

$$= NM \sum_{n=0}^{\infty} (-1)^n \binom{-\frac{N}{P}}{n} \frac{1}{qn+M} \int\limits_0^1 (1-\varrho^p)^n\, \varrho^{N-1}\, d\varrho =$$

$$= \sum_{n=0}^{\infty} \frac{N}{pn+N}\,\frac{M}{qn+M} = 1 + \sum_{n=1}^{\infty} \lambda_n^2.$$

The integrals were calculated by introducing spherical coordinates.

(Received December 6, 1962.)

REFERENCES

[1] Bahadur, R. R.: "Measurable subspaces and subalgebras". *Proc. Amer. Math. Soc* **6** (1955) 565—570.

[2] Bell, C. B.: "Mutual information and maximal correlation as measures of dependence". *Ann. Math. Stat.* **33** (1962) 587—595.

[3] Csáki, P.—Fischer, J.: "On bivariate stochastic connection". *Publ. Math. Inst. Hung. Acad. Sci.* **5** (1960) 311—323.

[4] Csáki, P.—Fischer, J.: "Contributions to the problem of maximal correlation". *Publ. Math. Inst. Hung. Acad. Sci.* **5** (1960) 325—337.

[5] Gebelein, H.: "Das statistische Problem der Korrelation als Variations- und Eigenwertproblem und sein Zusammenhang mit der Ausgleichsrechnung". *Zeitschrift für angewandte Mathematik und Mechanik* **21** (1941) 364—379.

[6] Hotelling, H.: "Relations between two sets of variates". *Biometrika* **28** (1936) 321—377.

[7] Канторович, Л. В. — Акилов, Г. П.: *Функциональный анализ в нормирсванных пространствах*. Гос. Издат. Физ. Мат. Лит., Москва, 1959.

[8] Колмогоров, А. Н. — Розанов, Ю. А.: "Об условиях сильного перемешивания гауссовского стационарного процесса." *Теория вероятностей и ее применения* **5** (1960) 222—227.

[9] Kramer, H. P.: "Maximal correlation and uncertainty". *Abstract* 576-213, *Notices Amer. Math. Soc.* **7** (1960) 977—978.

[10] Rényi, A.: "On measures of dependence". *Acta Math. Acad. Sci. Hung.* **10** (1959) 441—451.

[11] Umegaki, H.: "Conditional expectation in an operator algebra". *Tohoku Mathematical Journal* **6** (1954) 177—181.

ОБ ОБЩЕМ ПОНЯТИИ МАКСИМАЛЬНОЙ КОРРЕЛЯЦИИ

P. CSÁKI и J. FISCHER

Резюме

Авторы рассматривают некоторые свойства обобщенной максимальной корреляции. Цель статьи — обобщение, соответственно дополнение некоторых результатов, полученных авторами а также другими исследователями данной области.

В § 1 рассматривается понятие $\mathbf{S}(H_1, H_2)$, характеризующее взаимное расположение подпространств H_1, H_2 в общем Гильбертовом пространстве, являющемся в случае L^2 равным максимальной корреляцией.

Для этой цели авторы вводят в разделе 1.3 постулаты, основывающиеся на величине $\delta(H_1, H_2)$ определенной соотношением (1.7). Величина $\mathbf{S}(H_1, H_2)$, определенная соотношением (1.8) удовлетворяет этим постулатам. Приведенные здесь теоремы являются также обобщениями известных результатов, относящихся к максимальной корреляции (скалярных) случайных величин, а теорема 1.6 дает новое определение величины $\mathbf{S}(H_1, H_2)$.

В § 2 авторы показывают, что не существует такой не тождественно исчезающей функцией величины $\mathbf{S}(H_1, H_2)$, произведенные которой множества $V(x, \varepsilon)$ могли бы служить базисом топологии в множестве подпространств.

В § 3 дается толкование понятия максимальной корреляции в некоммутативной теории вероятностей.

В § 4 авторы рассматривают максимальную корреляцию между σ-алгебрами, обобщая при этом несколько их более ранних результатов (теоремы 4.1 и 4.4). В качестве частного случая рассматриваются максимальные корреляции между стохастическими процессами а также между векторными случайными величинами.

В § 5 авторы исследуют вопрос, в какой мере максимальная корреляция, как число измеряющее стохастическую связь, выражает наглядное содержание понятия этой связи. Они характеризуют интенсивность стохастической связи между двумья случайными величинами возможной

малостью среднеквадратической ошибки одной из случайных величин, расчитанной от ее ожидаемой условной величины по отношению другой случайной величине при взаимно-однозначном преобразовании случайных величин.

В этой связи они доказывают, что максимальная корреляция двух случайных величин равна сюпремуму корреляционных коэффициентов их взаимнооднозначных функций (теорема 5.1), значит она тем больше чем теснее связь в вышеуказанном смысле.

В § 6 авторы определяют значение максимальной корреляции в нескольких частных случаях и примерах. Они показывают, что максимальную корреляцию векторных случайных величин с каноническим распределением дает наибольшая максимальная корреляция между их компонентами (теорема 6.2). Отсюда они водят, что максимальная корреляция векторных случайных величин с нормальным распределением равна их наибольшей канонической корреляции (теорема 6.5). В случае полиномиального распределения максимальная корреляция даётся формулой (6.10), а в случае полигипергеометрического распределения формулой (6.12). Максимальная корреляция приведенного в разделе 6.5 распределения даётся формулой (6.14).