ON MEASURES OF DEPENDENCE

By

A. RÉNYI (Budapest), member of the Academy

Introduction

In this paper we shall discuss and compare certain quantities which are used to measure the strength of dependence (or correlation, in the widest sense of this word) between two random variables. We formulate seven rather natural postulates which should be fulfilled by a suitable measure of dependence. The maximal correlation introduced by H. GEBELEIN [1] (for a more general treatment see [2]) fulfils all these postulates. As in our previous paper [2] we shall make use of the technique of conditional mean values, as developed by A. N. KOLMOGOROV [3], which is needed to define the different measures of dependence and to prove their properties, and the connections between them, under much more general conditions than this is usual in the literature.

In § 1 we introduce the definitions and notations to be used throughout the paper. § 2 contains the definitions and fundamental properties of the mentioned measures of dependence. § 3 contains the proof of the main theorem of the present paper (Theorem 2), according to which the maximal correlation can be attained, provided that the mean square contingency is finite.

§ 1. Definitions and notations

Let $[\Omega, \mathcal{A}, \mathbf{P}]$ be a probability space (see [3]), i. e. Ω an arbitrary non-empty set whose elements will be denoted by ω, $\mathcal{A}$ a σ-algebra of subsets of Ω whose elements will be denoted by capital letters A, B etc., and $\mathbf{P}=\mathbf{P}(A)$ a probability measure on $\mathcal{A}$. We shall denote random variables on $[\Omega, \mathcal{A}, \mathbf{P}]$ (i. e. real functions defined on Ω and measurable with respect to $\mathcal{A}$) by Greek letters ξ, η etc. If ξ is a random variable, we denote by $\mathbf{M}(\xi)$ its mean value and by $\mathbf{D}^2(\xi)$ its variance. If $\mathbf{M}(\xi)$ and $\mathbf{D}(\xi)$ exist and $\mathbf{D}(\xi)>0$, we put

$$\xi^* = \frac{\xi-\mathbf{M}(\xi)}{\mathbf{D}(\xi)} \tag{1}$$

and call the transformation by which ξ^* is obtained from ξ the standardization of ξ. If ξ is an arbitrary random variable, let $\mathcal{A}_\xi$ denote the least σ-algebra of subsets of Ω with respect to which ξ is measurable. If η is another random variable with finite mean value, we denote by $\mathbf{M}(\eta|\xi)$ the conditional mean value of η with respect to a given value of ξ; $\mathbf{M}(\eta|\xi)$ itself is a random variable which is measurable with respect to $\mathcal{A}_\xi$ and is such that for any $A \in \mathcal{A}_\xi$ we have

$$\int_A \mathbf{M}(\eta|\xi) d\mathbf{P} = \int_A \eta \, d\mathbf{P}; \tag{2}$$

of course, $\mathbf{M}(\eta|\xi)$ is unique only if we consider two random variables which are equal with probability 1 to be identical. In what follows we shall always take this for granted. The following two well-known properties (see [3]) of conditional mean values will often be used in the sequel: If $\mathbf{M}(\eta)$ exists, then

$$\mathbf{M}(\mathbf{M}(\eta|\xi)) = \mathbf{M}(\eta) \tag{3}$$

and

$$\mathbf{M}(g(\xi)\eta|\xi) = g(\xi)\mathbf{M}(\eta|\xi) \tag{4}$$

if $g(x)$ is a Borel-measurable real function of the real variable x. The curve $y = \mathbf{M}(\eta|\xi = x)$ is called the regression curve of η on ξ.

We shall denote the joint distribution of two random variables ξ and η by $\mathbf{Q}_{\xi,\eta}$, i. e. we put for any Borel subset C of the (x, y)-plane

$$\mathbf{Q}_{\xi,\eta}(C) = \mathbf{P}((\xi, \eta) \in C)$$

where $(\xi, \eta) \in C$ denotes the set of those $\omega \in \Omega$ for which the point with the coordinates $\xi(\omega), \eta(\omega)$ belongs to C. We denote, further, by $\mathbf{Q}_{\xi*\eta}$ the direct product of the distributions of ξ and η, i. e. we put for any two Borel subsets A and B of the real line

$$\mathbf{Q}_{\xi*\eta}(A*B) = \mathbf{P}(\xi \in A)\mathbf{P}(\eta \in B)$$

where $A*B$ denotes the direct product of the sets A and B, i. e. the set of all points (x, y) for which $x \in A$ and $y \in B$. The definition of $\mathbf{Q}_{\xi*\eta}$ is extended to any Borel subset C of the (x, y)-plane in the usual way (see e. g. [4]).

§ 2. Definitions and fundamental properties of measures of dependence

Let ξ and η be random variables on a probability space $[\Omega, \mathcal{A}, \mathbf{P}]$, neither of them being constant with probability 1. In almost every field of application of statistics one encounters often the problem that one has to characterize by a numerical value the strength of dependence between ξ and η. Of course, such a value serves only for comparison, and thus its range is arbitrary. It is natural to choose the range $[0, 1]$ and to make correspond the value 1 to strict dependence and thus 0 to independence. With these conventions the following set of postulates for an appropriate measure of dependence, which shall be denoted by $\delta(\xi, \eta)$, seems to be natural:

A) *$\delta(\xi, \eta)$ is defined for any pair of random variables ξ and η, neither of them being constant with probability* 1.

B) $\delta(\xi, \eta) = \delta(\eta, \xi)$.

C) $0 \leqq \delta(\xi, \eta) \leqq 1$.

D) *$\delta(\xi, \eta) = 0$ if and only if ξ and η are independent.*

E) *$\delta(\xi, \eta) = 1$ if*[1] *there is a strict dependence between ξ and η, i. e. either $\xi = g(\eta)$ or $\eta = f(\xi)$ where $g(x)$ and $f(x)$ are Borel-measurable functions.*

F) *If the Borel-measurable functions $f(x)$ and $g(x)$ map the real axis in a one-to-one way onto itself, $\delta(f(\xi), g(\eta)) = \delta(\xi, \eta)$.*

G) *If the joint distribution of ξ and η is normal, then $\delta(\xi, \eta) = |\mathbf{R}(\xi, \eta)|$ where $\mathbf{R}(\xi, \eta)$ is the correlation coefficient of ξ and η.*

Let us now consider the most frequently used measures of dependence and see which of the above properties they possess.[2]

1. *The correlation coefficient.* The correlation coefficient $\mathbf{R}(\xi, \eta)$ is defined, provided that $\mathbf{D}(\xi)$ and $\mathbf{D}(\eta)$ are finite and positive, by

$$\mathbf{R}(\xi, \eta) = \frac{\mathbf{M}(\xi\eta) - \mathbf{M}(\xi)\mathbf{M}(\eta)}{\mathbf{D}(\xi)\mathbf{D}(\eta)} = \mathbf{M}(\xi^*, \eta^*). \tag{5}$$

It has the range $[-1, +1]$, thus only its absolute value satisfies postulate C). $|\mathbf{R}(\xi, \eta)|$ has the properties B) and C) (and, of course, G)), but it does not have the other properties. As a matter of fact, it is defined only if $\mathbf{D}(\xi)$ and $\mathbf{D}(\eta)$ are finite and positive, it may vanish also if ξ and η are not independent; moreover, it may vanish in spite of a functional dependence between ξ and η; for example, if ξ is uniformly distributed in $(-1, +1)$ and $\eta = 5\xi^3 - 3\xi$, we have $\mathbf{R}(\xi, \eta) = 0$; $|\mathbf{R}(\xi, \eta)|$ is equal to 1 if and only if there is a linear relation between ξ and η.

2. *The correlation ratios.* The correlation ratio $\Theta_\xi(\eta)$ of η on ξ is defined (see [3]) by

$$\Theta_\xi(\eta) = \frac{\mathbf{D}(\mathbf{M}(\eta \mid \xi))}{\mathbf{D}(\eta)}, \tag{6}$$

provided that $\mathbf{D}(\eta)$ exists and is positive. It has the range $[0, 1]$ but it is

[1] It seems at the first sight natural to postulate that $\delta(\xi, \eta) = 1$ *only* if there is a strict dependence of the mentioned type between ξ and η, but this condition is rather restrictive, and it is better to leave it out.

[2] To make the present paper self-contained we repeat some of the results of [2].

not symmetric. If we consider instead of $\boldsymbol{\Theta}_\xi(\eta)$ the quantity

$$\boldsymbol{\Theta}(\xi, \eta) = \max\left(\boldsymbol{\Theta}_\xi(\eta), \boldsymbol{\Theta}_\eta(\xi)\right), \tag{7}$$

it still does not satisfy A), D) and F). In fact, $\boldsymbol{\Theta}(\xi, \eta)$ is defined only if ξ and η have finite variances. As regards D), $\boldsymbol{\Theta}(\xi, \eta)$ vanishes e. g. if the random point (ξ, η) is uniformly distributed in a circle. $\boldsymbol{\Theta}(\xi, \eta)$ satisfies, however, postulate E).

As well known, postulate G) is also fulfilled for $\boldsymbol{\Theta}_\xi(\eta)$ and $\boldsymbol{\Theta}(\xi, \eta)$. It is easy to show, further, (see [2]) that

$$\boldsymbol{\Theta}_\xi(\eta) = \sup_{(f)} \mathbf{R}(f(\xi), \eta) \tag{8}$$

where $f = f(x)$ runs over all Borel-measurable functions for which $\mathbf{R}(f(\xi), \eta)$ is defined, i. e. for which $f(\xi)$ has finite positive variance. There can be always found a function $f_0(x)$ such that

$$\boldsymbol{\Theta}_\xi(\eta) = \mathbf{R}(f_0(\xi), \eta). \tag{9}$$

In fact, we may, without restricting the generality, consider only such functions f for which $\mathbf{M}(f(\xi)) = 0$ and $\mathbf{D}(f(\xi)) = 1$ and suppose $\mathbf{M}(\eta) = 0$, $\mathbf{D}(\eta) = 1$; in this case we have by the inequality of Schwarz and by virtue of (3) and (4)

$$\mathbf{R}(f(\xi), \eta) = \mathbf{M}(f(\xi)\eta) = \mathbf{M}(f(\xi)\mathbf{M}(\eta\,|\,\xi)) \leqq \mathbf{D}(\mathbf{M}(\eta\,|\,\xi))$$

with equality standing if and only if $f(\xi) = f_0(\xi) = \dfrac{\mathbf{M}(\eta\,|\,\xi)}{\mathbf{D}(\mathbf{M}(\eta\,|\,\xi))}$ for which choice of f_0 therefore (9) holds.

3. *The maximal correlation.* In view of (8) it is quite natural to consider the quantity

$$\mathbf{S}(\xi, \eta) = \sup_{f, g} \mathbf{R}(f(\xi), g(\eta)) \tag{10}$$

where $f(x)$ and $g(x)$ run over all Borel-measurable functions such that $\mathbf{R}(f(\xi), g(\eta))$ has a sense, i. e. such that $f(\xi)$ and $g(\eta)$ have finite und positive variance. The quantity (10) has been introduced for discrete and absolutely continuous distributions, respectively, by H. GEBELEIN [1] (see also [5]) and called by him the *maximal correlation* (Maximal-Korrelation) of ξ and η. A more general treatment is given in [2]. It is easy to show that $\mathbf{S}(\xi, \eta)$ has all the properties A) to G) listed above. That $\mathbf{S}(\xi, \eta)$ has the properties A), B), C), E) and F) is evident.[3] That G) is also satisfied for

[3] $\mathbf{S}(\xi, \eta)$ may be equal to 1 not only if $\eta = f(\xi)$ or $\xi = g(\eta)$. In fact, if $f(\xi) = g(\eta)$ where f and g are Borel-measurable functions, then $\mathbf{S}(\xi, \eta) = 1$.

$\mathbf{S}(\xi, \eta)$ has been shown in [1]. To show that it has the property D) choose

$$f(x)=\begin{cases}1 & \text{for} \quad x<a,\\ 0 & \text{for} \quad x \geqq a\end{cases} \quad \text{and} \quad g(x)=\begin{cases}1 & \text{for} \quad x<b,\\ 0 & \text{for} \quad x \geqq b.\end{cases}$$

In contrary to (9), there does not always exist such functions $f_0(x)$ and $g_0(x)$ that

$$\mathbf{S}(\xi, \eta)=\mathbf{R}(f_0(\xi), g_0(\eta)). \tag{11}$$

If (11) holds for some f_0 and g_0, we shall say that *the maximal correlation of ξ and η can be attained.* An example of two random variables for which the maximal correlation can not be attained has been found by J. CZIPSZER to whom the author is obliged for kindly communicating his example.

Another example constructed by the same idea is the following: Let the random point (ξ, η) be uniformly distributed in the domain G shown on Fig. 1, bounded by two curves meeting each other at the origin where they have the straight line $y=x$ for their common tangent. Choose

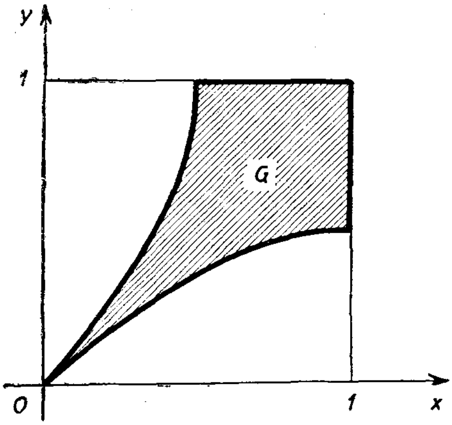

Fig. 1

$$f(x)=g(x)=\begin{cases}1 & \text{for} \quad 0 \leqq x \leqq \varepsilon,\\ 0 & \text{for} \quad x>\varepsilon\end{cases}$$

and let $\varepsilon \rightarrow 0$; it follows that $\mathbf{S}(\xi, \eta)=1$, while, evidently, there do not exist such functions f_0, g_0 that (11) should hold, as (11) would imply $f_0(\xi) \equiv g_0(\eta)$, which is clearly impossible as it would imply $\iint_G |f_0(x)-g_0(y)| dx dy=0$ and thus that $f_0(x)$ and $g_0(x)$ both are equal to the same constant, in which case, however, $\mathbf{R}(f_0(\xi), g_0(\eta))$ is not defined.

Sufficient conditions under which the maximal correlation can be attained will be given in § 3.

4. *The mean square contingency.* Let us call as in [2] the dependence between the random variables ξ and η *regular* if their joint distribution $\mathbf{Q}_{\xi,\eta}$ is absolutely continuous with respect to the direct product $\mathbf{Q}_{\xi * \eta}$ of their distributions. If the dependence between ξ and η is regular, then according to the theorem of Radon—Nikodym (see [4]), there exists a Borel-measurable function $k(x, y)$ such that if $F(x)$ and $G(y)$ denote the distribution functions

of ξ and η, respectively, we have for any Borel set C of the (x, y)-plane

$$\mathbf{Q}_{\xi,\eta}(C)=\iint_C k(x, y)\,dF(x)\,dG(y). \tag{12}$$

Put

$$\mathbf{C}(\xi,\eta)=\left(\int_{-\infty}^{+\infty}\int_{-\infty}^{+\infty}(k(x, y)-1)^2\,dF(x)\,dG(y)\right)^{\frac{1}{2}}. \tag{13}$$

We call $\mathbf{C}(\xi,\eta)$ the *mean square contingency* of ξ and η. If the joint distribution of ξ and η is either discrete or absolutely continuous, the above definition introduced in [2] reduces to the usual one (see e. g. [6]). $\mathbf{C}(\xi,\eta)$ may be considered also as a measure of the dependence between ξ and η. It satisfies B), D) and F) but none of A), C) and E). Its range is evidently $[0, +\infty]$; this can of course be transformed into $[0, 1]$ by considering e. g. instead of $\mathbf{C}(\xi,\eta)$ the quantity $\Gamma(\xi,\eta)=\dfrac{\mathbf{C}(\xi,\eta)}{\sqrt{1+\mathbf{C}^2(\xi,\eta)}}$ in which case G) is also satisfied; but A) and E) remain unfulfilled.

It should be mentioned that the following series of inequalities is valid (see [2]):

$$0\leq|\mathbf{R}(\xi,\eta)|\leq\min(\Theta_\xi(\eta),\Theta_\eta(\xi))\leq\Theta(\xi,\eta)\leq\mathbf{S}(\xi,\eta)\leq\mathbf{C}(\xi,\eta). \tag{14}$$

Only the last inequality remains to be verified. If

$$\mathbf{M}(f(\xi))=\mathbf{M}(g(\eta))=0 \quad \text{and} \quad \mathbf{D}(f(\xi))=\mathbf{D}(g(\eta))=1,$$

we have

$$\mathbf{R}(f(\xi), g(\eta))=\int_{-\infty}^{+\infty}\int_{-\infty}^{+\infty}f(x)g(y)(k(x, y)-1)\,dF(x)\,dG(y).$$

It follows by the inequality of Schwarz that $\mathbf{R}(f(\xi), g(\eta))\leq\mathbf{C}(\xi,\eta)$. As this holds for any f and g, the inequality $\mathbf{S}(\xi,\eta)\leq\mathbf{C}(\xi,\eta)$ follows immediately.

5. *Some further remarks on the maximal correlation and on other measures of dependence.* The quantity $\Theta(\xi,\eta)$ introduced above has the property that it is equal to 1 only if $\eta=f(\xi)$ or $\xi=g(\eta)$. It follows that $\frac{1}{2}(\mathbf{S}(\xi,\eta)+ +\Theta(\xi,\eta))$ has the same property, and it has besides that all properties B) to E) and G), but A) and F) are not fulfilled. Of course, there is some arbitrariness in taking just the arithmetic mean, as many other mean values would do the same. It is more reasonable to say that though the maximal correlation is in many respect superior to other measures of dependence, it does not make superfluous to consider other measures, e. g. the correlation ratios, too. Of course, the drawback of the maximal correlation is that it is

often difficult to calculate it. It should be, however, added that this is not so difficult a task as it may seem at the first sight. We shall return to this question in the next § where methods for calculating $\mathbf{S}(\xi, \eta)$ will be mentioned. I should mention here only that there are already many cases known in which $\mathbf{S}(\xi, \eta)$ can be effectively calculated. For instance, P. BÁRTFAI has calculated $\mathbf{S}(\xi, \eta)$ for the uniform distribution in a circle, and found $\mathbf{S}(\xi, \eta)=\frac{1}{3}$. P. CSÁKI and J. FISCHER determined the value of $\mathbf{S}(\xi, \eta)$ for the case when the point (ξ, η) is uniformly distributed in the domain $|x|^p+|y|^p \leqq 1$ where $p>0$ and have shown that $\mathbf{S}(\xi, \eta)=\frac{1}{p+1}$. (This includes the result of BÁRTFAI for $p=2$.)

A measure of dependence based on information-theoretical considerations has been recommended recently by E. H. LINFOOT [7], namely the quantity $\mathbf{L}(\xi, \eta)=(1-e^{-2\mathbf{I}(\xi, \eta)})^{\frac{1}{2}}$ where $\mathbf{I}(\xi, \eta)$ is the amount of information which ξ and η, resp., contain with respect to the other (see e. g. [8]). If the dependence between ξ and η is regular, $\mathbf{I}(\xi, \eta)$ can be written in the form

$$\mathbf{I}(\xi, \eta)=\int_{-\infty}^{+\infty}\int_{-\infty}^{+\infty} k(x, y) \log k(x, y) dF(x) dG(y).$$

This quantity has all the properties A), B), C), D), E), F) and G). (LINFOOT has chosen the particular formula by which $\mathbf{L}(\xi, \eta)$ is calculated from $\mathbf{I}(\xi, \eta)$ to ensure that postulate G) should be fulfilled.)

§ 3. Conditions under which the maximal correlation can be attained

It is easy to see that if (11) holds, i. e. the maximal correlation of ξ and η is attained for $f_0(\xi)$ and $g_0(\eta)$, then we have, provided that $\mathbf{M}(f_0(\xi))=\mathbf{M}(g_0(\eta))=0$ and $\mathbf{D}(f_0(\xi))=\mathbf{D}(g_0(\eta))=1$, putting for the sake of brevity $S=\mathbf{S}(\xi, \eta)$,

$$\mathbf{M}(f_0(\xi) \mid \eta)=S g_0(\eta) \tag{15}$$

and

$$\mathbf{M}(g_0(\eta) \mid \xi)=S f_0(\xi). \tag{16}$$

As a matter of fact, if (11) holds, then by the same argument as used in proving that (9) holds for $f_0(\xi)=\frac{\mathbf{M}(\eta \mid \xi)}{\mathbf{D}(\mathbf{M}(\eta \mid \xi))}$, one gets (15) and (16). Thus $f_0(x)$ and $g_0(x)$ satisfying (11) can — if they exist — be found as solutions of the system of equations (15) and (16). These equations can be replaced by the equations

(17) $$\mathbf{M}(\mathbf{M}(f_0(\xi)\,|\,\eta)\,|\,\xi) = S^2 f_0(\xi)$$

and

(18) $$\mathbf{M}(\mathbf{M}(g_0(\eta)\,|\,\xi)\,|\,\eta) = S^2 g_0(\eta).$$

It is sufficient to determine f_0 from (17); if f_0 is known, g_0 can be obtained from (15).

Let $\mathfrak{L}_\xi^2$ denote the Hilbert space of all random variables of the form $f(\xi)$ for which $\mathbf{M}(f(\xi)) = 0$ and $\mathbf{D}(f(\xi))$ is finite, and similarly $\mathfrak{L}_\eta^2$ the Hilbert space of all those random variables $g(\eta)$ for which $\mathbf{M}(g(\eta)) = 0$ and $\mathbf{D}(g(\eta))$ is finite. Let us put for any $f = f(\xi) \in \mathfrak{L}_\xi^2$

(19) $$Af = \mathbf{M}(\mathbf{M}(f(\xi)\,|\,\eta)\,|\,\xi),$$

then (17) can be written in the form

(20) $$Af_0 = S^2 f_0.$$

Let us define the inner product (f_1, f_2) for $f_1 = f_1(\xi) \in \mathfrak{L}_\xi^2, f_2 = f_2(\xi) \in \mathfrak{L}_\xi^2$ by

(21) $$(f_1, f_2) = \mathbf{M}(f_1(\xi) f_2(\xi))$$

and put for $f \in \mathfrak{L}_\xi^2$

(22) $$\|f\| = (f, f)^{\frac{1}{2}} = \mathbf{D}(f(\xi)).$$

Let us investigate the transformation A. As for any random variable ζ we have $\mathbf{M}(\mathbf{M}^2(\zeta\,|\,\xi)) = \mathbf{D}^2(\zeta)\Theta_\xi^2(\zeta) \leqq \mathbf{D}^2(\zeta)$, one obtains $\|Af\| \leqq \|f\|$.

Thus it follows that Af is a bounded linear transformation of the Hilbert space $\mathfrak{L}_\xi^2$. We shall show that A is self-adjoint, moreover that it is positive definite.

As a matter of fact, one gets for $f_1 \in \mathfrak{L}_\xi^2$, $f_2 \in \mathfrak{L}_\xi^2$, by using (3) and (4) repeatedly,

(23) $$(Af_1, f_2) = \mathbf{M}(\mathbf{M}(f_1(\xi)\,|\,\eta)\,\mathbf{M}(f_2(\xi)\,|\,\eta)).$$

Interchanging f_1 and f_2 it follows that

(24) $$(Af_1, f_2) = (f_1, Af_2)$$

which shows that A is a bounded self-adjoint transformation of $\mathfrak{L}_\xi^2$. It follows also from (23) that

(25) $$(Af, f) = \mathbf{M}(\mathbf{M}^2(f(\xi)\,|\,\eta)) \geqq 0$$

and thus A is positive definite.

Now clearly for any $f \in \mathfrak{L}^2_\xi$ and $g \in \mathfrak{L}^2_\eta$ with $\mathbf{D}(f(\xi)) = \mathbf{D}(g(\eta)) = 1$ we have by (3) and the Schwarz inequality

$$\mathbf{M}^2(f(\xi)g(\eta)) = \mathbf{M}^2(\mathbf{M}(f(\xi)\,|\,\eta)g(\eta)) \leqq \mathbf{M}(\mathbf{M}^2(f(\xi)\,|\,\eta)) = (Af, f),$$

and thus putting

$$\lambda = \sup_{\substack{\|f\|=1 \\ f \in \mathfrak{L}^2_\xi}} (Af, f) \tag{26}$$

we have

$$S^2 \leqq \lambda. \tag{27}$$

On the other hand, if $f \in \mathfrak{L}^2_\xi$ and $\|f\| = 1$, putting $g(\eta) = \mathbf{M}(f(\xi)\,|\,\eta)$ we have $\mathbf{D}^2(g(\eta)) = \mathbf{M}(f(\xi)g(\eta)) \leqq S\mathbf{D}(g(\eta))$ which implies $\mathbf{D}(g(\eta)) \leqq S$; thus it follows

$$(Af, f) = \mathbf{M}(g(\eta)\mathbf{M}(f(\xi)\,|\,\eta)) = \mathbf{M}(f(\xi)g(\eta)) \leqq S\mathbf{D}(g(\eta)) \leqq S^2$$

and therefore

$$\lambda \leqq S^2. \tag{28}$$

Thus we have from (27) and (28)

$$S^2 = \lambda = \sup_{\substack{f \in \mathfrak{L}^2_\xi \\ \|f\|=1}} (Af, f). \tag{29}$$

It is known from the theory of bounded self-adjoint transformations that in case A is completely continuous (see e. g. [9]), then $\lambda = \sup\limits_{\|f\|=1} (Af, f)$ is the greatest eigenvalue of A and there exists an eigenfunction belonging to the eigenvalue λ.

Thus we have proved the following

THEOREM 1. *If the transformation A defined by* (19) *is completely continuous, then the maximal correlation of ξ and η is attained for $f_0(\xi)$ and $g_0(\eta)$ where f_0 is an eigenfunction belonging to the greatest eigenvalue $S^2 = \mathbf{S}^2(\xi, \eta)$ of A and* $g_0(\eta) = \dfrac{1}{S}\mathbf{M}(f_0(\xi)\,|\,\eta)$.

The condition that A should be completely continuous is not easy to verify in concrete cases. Therefore the following theorem is useful:

THEOREM 2. *If the dependence between ξ and η is regular and the mean square contingency $\mathbf{C}(\xi, \eta)$ is finite, then the transformation A is completely continuous and thus the maximal correlation of ξ and η can be attained.*

PROOF OF THEOREM 2. We have by supposition, denoting by $F(x)$ and $G(y)$ the distribution functions of ξ and η, resp.,

$$\mathbf{Q}_{\xi,\eta}(C) = \iint\limits_C k(x, y)\,dF(x)\,dG(y)$$

where

$$(30) \qquad \int_{-\infty}^{+\infty}\int_{-\infty}^{+\infty} k^2(x, y)\,dF(x)\,dG(y) = 1 + \mathbf{C}^2(\xi, \eta)$$

is finite.

Evidently, in proving that A is completely continuous, we can operate on another Hilbert space which is isomorphic to $\mathfrak{L}_\xi^2$ considered above (i. e. choose another realization of the same abstract Hilbert space). Now $\mathfrak{L}_\xi^2$ is clearly isomorphic to the Hilbert space L_F^2 of all measurable functions $f(x)$ for which

$$\int_{-\infty}^{+\infty} f(x)\,dF(x) = 0 \quad \text{and} \quad \int_{-\infty}^{+\infty} f^2(x)\,dF(x) < +\infty,$$

provided that $f(x)$ corresponds to $f(\xi)$, and for $f_1 \in L_F^2, f_2 \in L_F^2$ the inner product (f_1, f_2) is defined by

$$(f_1, f_2) = \int_{-\infty}^{+\infty} f_1(x) f_2(x)\,dF(x).$$

In this Hilbert space we have

$$Af(x) = \int_{-\infty}^{+\infty} f(u)\Big(\int_{-\infty}^{+\infty} k(u, v)k(x, v)\,dG(v)\Big)\,dF(u)$$

(see [9]). Now it is well known that an integral operator is completely continuous if the square of its kernel is integrable. Thus A is completely continuous provided that

$$(31) \qquad \Delta = \int_{-\infty}^{+\infty}\int_{-\infty}^{+\infty}\Big(\int_{-\infty}^{+\infty} k(u, v)k(x, v)\,dG(v)\Big)^2 dF(u)\,dF(x)$$

exists.

As by the inequality of Schwarz

$$\Big(\int_{-\infty}^{+\infty} k(u, v)k(x, v)\,dG(v)\Big)^2 \leqq \int_{-\infty}^{+\infty} k^2(u, v)\,dG(v) \cdot \int_{-\infty}^{+\infty} k^2(x, v)\,dG(v),$$

it follows by (30) that

$$(32) \qquad \Delta \leqq (1 + \mathbf{C}^2(\xi, \eta))^2,$$

which proves Theorem 2.

It should be mentioned that the condition that $\mathbf{C}(\xi, \eta)$ is finite is not necessary.

A simple example is furnished by the case when $\eta = \xi$; in this case, of course, $\mathbf{S}(\xi, \eta) = 1$ and the maximal correlation is trivially attained for

$f_0(x)=g_0(x)=x$, but the dependence between ξ and η is not regular and the mean square contingency is not defined.

It should be added that Theorem 2 could also be proved by considering the bilinear functional $B(f, g)$ defined for $f=f(x)\in L_F^2$ and $g=g(y)\in L_G^2$ by

$$B(f,g)=\int_{-\infty}^{+\infty}\int_{-\infty}^{+\infty} f(x)g(y)k(x,y)dF(x)dG(y) \tag{33}$$

and proving directly that if $\mathbf{C}(\xi, \eta)$ is finite, B is completely continuous in the sense of Hilbert [10], i. e. that if $f_n\in L_F^2, g_n\in L_G^1, \|f_n\|\leqq 1, \|g_n\|\leqq 1$ $(n=1,2,\ldots)$, further f_n and g_n converge weakly to f and g in L_F^2 and L_G^2, resp., then $\lim\limits_{n\to+\infty} B(f_n, g_n)=B(f,g)$. The proof is essentially the same as that for the case of a bilinear form $\sum\limits_{j=1}^{\infty}\sum\limits_{k=1}^{\infty} c_{jk}x_j y_k$ of infinitely many variables.

(Received 8 September 1959)

References

[1] H. Gebelein, Das statistische Problem der Korrelation als Variations- und Eigenwertproblem und sein Zusammenhang mit der Ausgleichungsrechnung, *Zeitschrift für angew. Math. und Mech.*, **21** (1941), pp. 364—379.

[2] A. Rényi, New version of the probabilistic generalization of the large sieve, *Acta Math. Acad. Sci. Hung.*, **10** (1959), pp. 218—226.

[3] A. Kolmogoroff, *Grundbegriffe der Wahrscheinlichkeitsrechnung*, Ergebnisse der Math. u. Grenzgebiete (Berlin, 1933).

[4] P. Halmos, *Measure theory* (New York, 1952).

[5] A. O. Hirschfeld, A connection between correlation and contingency, *Proc. Cambridge Phil. Soc.*, **31**, pp. 520—524.

[6] H. Cramér, *Mathematical methods of statistics* (Princeton, 1946).

[7] E. H. Linfoot, An informational measure of correlation, *Information and Control*, **1** (1957), pp. 85—89.

[8] L. M. Gelfand, A. M. Jaglom and A. N. Kolmogoroff, *Zur allgemeinen Definition der Information*, Arbeiten zur Informationstheorie. II (Berlin, 1958), pp. 57—60.

[9] F. Riesz et B. Sz.-Nagy, *Leçons d'analyse fonctionnelle*, (Budapest, 1955).

[10] D. Hilbert, *Grundzüge einer allgemeinen Theorie der linearen Integralgleichungen* (Leipzig, 1912).