ON BIVARIATE STOCHASTIC CONNECTION

by

PÉTER CSÁKI and JÁNOS FISCHER

Introduction

In this paper the problems of bivariate connection are discussed with the aid of Hilbert space theory and not by classical methods. This enables a more general treatment.

The maximal correlation is shown to be a highly adapt measure for the intensity of bivariate stochastic connection. The calculation of the maximal correlation leads to the determination of the eigenvalues of a pair of operators. Some characteristics of this pair of operators will be discussed.

In § 1 the main notions and symbols will be introduced and the conditional expected value will be considered, further some characteristics of the correlation ratio and the maximal correlation will be discussed. Moreover a generalized definition of the mean square contingency will be given. In § 2 this generalized definition will be proved to include the former definitions of this notion. Conditions under which the mentioned pair of operators forms a pair of integral operators will be given.

Finally, a method will be described for replacing any given distribution by a symmetric one apt to our purposes.

Some characteristics of maximal correlation will be treated with the aid of these results in our following paper.

§ 1. Basic notions

1.1. First of all, the notions, terminology and symbols used in this paper will be explained.

Let $(\Omega, \boldsymbol{S}, \mathbf{P})$ be a probability space, i. e. Ω a space of events, $\boldsymbol{S}$ a σ-algebra of its subsets and $\mathbf{P}$ a probability measure defined on $\boldsymbol{S}$ $(\mathbf{P}(\Omega) = 1)$.

Let ξ be any random variable (real measurable function defined on Ω). In the following, two random variables which coincide with probability 1 are considered as identical. The expected value of ξ — if it exists — is denoted by $\mathbf{M}(\xi)$, so $\mathbf{M}(\xi) = \int\limits_{\Omega} \xi \, d\, \mathbf{P}$. $\boldsymbol{S}_\xi$ denotes the smallest σ-algebra with respect to which ξ is measurable, i. e. such a σ-algebra of events $\mathcal{A}$ for which $\mathcal{A} = \{\xi \in B\}$, where B is any Borel set on the real line. It is pointed out by J. L. DOOB ([2] p. 603) that the random variable ζ is measurable with respect to $\boldsymbol{S}_\xi$ if and only if it is a function of ξ, i. e. if there exists a Borel measurable function $f(x)$ such that $\zeta = f(\xi)$.

$L^2 = L^2(\Omega, \boldsymbol{S}, \mathbf{P})$ denotes a space of random variables ζ for which $\mathbf{M}(\zeta^2)$ is finite. This space L^2 forms such a complete Hilbert space in which $(\zeta_1, \zeta_2) = \mathbf{M}(\zeta_1 \zeta_2)$ is the scalar product of $\zeta_1 \in L^2$ and $\zeta_2 \in L^2$, thus $\| \zeta \| = \sqrt{\mathbf{M}(\zeta^2)}$ is the norm of $\zeta \in L^2$. The standard deviation of $\zeta \in L^2$ is denoted by $\mathbf{D}(\zeta)$, i. e. $\mathbf{D}(\zeta) = \| \zeta - \mathbf{M}(\zeta) \|$. We call the random variable $\frac{\zeta - \mathbf{M}(\zeta)}{\mathbf{D}(\zeta)}$ where $\mathbf{D}(\zeta) \neq 0$ the standardized of $\zeta \in L^2$, and denote it by ζ^*; ζ is called standard if $\zeta = \zeta^*$ (i. e. if $\mathbf{M}(\zeta) = 0$, $\mathbf{D}(\zeta) = 1$). The correlation coefficient of $\zeta_1 \in L^2$, $\mathbf{D}(\zeta_1) \neq 0$ and $\zeta_2 \in L^2$, $\mathbf{D}(\zeta_2) \neq 0$ is denoted by $\mathbf{R}(\zeta_1, \zeta_2)$, thus $\mathbf{R}(\zeta_1, \zeta_2) = (\zeta_1^*, \zeta_2^*)$.

Let us denote the distribution function of any random variable ξ by $F(x)$, further

$$L_F^2 = \left\{ f(x) : \int_{-\infty}^{\infty} f^2(x)\, dF(x) < \infty \right\}.$$

L_F^2 is a complete and separable Hilbert space in which the scalar product is defined by $\big(f_1(x), f_2(x)\big) = \int_{-\infty}^{\infty} f_1(x) f_2(x)\, dF(x)$. In case of ξ being a discrete random variable which can take on n different values, the space L_F^2 is an n-dimensional Euclidean space.

Let be $f(x) \in L_F^2$. Then the random variable $f = f(\xi)$ is such that $f \in L^2$. The space of random variables of this form are denoted by L_ξ^2. Obviously $L_\xi^2 = L^2(\Omega, \boldsymbol{S}_\xi, \mathbf{P})$. Between the elements of the space L_F^2 and L_ξ^2 there exists a one-to-one correspondence preserving the scalar product — and so the norm as well. Consequently, L_ξ^2 is a complete and separable Hilbert space and it is of finite dimension if ξ takes on only a finite number of values. All what has been stated for L_F^2 is valid for L_ξ^2, too.

1.2. Every random variable $\zeta \in L^2$ can be uniquely decomposed in the form $\zeta = \zeta' + \zeta''$ where $\zeta' \in L_\xi^2$ and ζ'' is orthogonal to any element of L_ξ^2. Accordingly, for ζ there exists such a unique $\zeta' \in L_\xi^2$ that $(\zeta', f) = (\zeta, f)$ for any $f \in L_\xi^2$, i. e. $\zeta - \zeta'$ is orthogonal to L_ξ^2. For ζ', $\min \| \zeta - f \| = \| \zeta - \zeta' \|$ holds. The operator transforming ζ into its orthogonal projection on L_ξ^2 (i. e. into ζ') is denoted by $\mathbf{A}_\xi$. Therefore

$$(\mathbf{A}_\xi \zeta, f) = (\zeta, f) \text{ whenever } \zeta \in L^2,\ f \in L_\xi^2. \tag{1.1}$$

Later we shall see for any $\zeta \in L^2$, that $\mathbf{A}_\xi \zeta$ is the regression curve of ζ on ξ (i. e. the conditional expected value of ζ on ξ).

The conditional expected value is defined by KOLMOGOROV in the following way: the conditional expected value of the integrable random variable ζ on the conditioning random variable ξ is such an $\boldsymbol{S}_\xi$-measurable random variable $\mathbf{M}(\zeta \mid \xi)$ for which

$$\int_{\mathcal{A}} \mathbf{M}(\zeta \mid \xi)\, d\mathbf{P} = \int_{\mathcal{A}} \zeta\, d\mathbf{P} \quad \text{for any } \mathcal{A} \in \boldsymbol{S}_\xi. \tag{1.2}$$

$\mathbf{M}(\zeta \mid \xi)$ is with probability 1 uniquely determined by the Radon—Nikodym theorem, further

$$\mathbf{M}(\mathbf{M}(\zeta \mid \xi)) = \mathbf{M}(\zeta)$$

and for any S_ξ-measurable random variable f such that ζf is integrable,

$$\mathbf{M}(\zeta f \mid \xi) = f\,\mathbf{M}(\zeta \mid \xi)$$

holds. Consequently,

$$\int_\Omega f\,\mathbf{M}(\zeta \mid \xi)\,d\mathbf{P} = \int_\Omega \zeta f\,d\mathbf{P}\,. \tag{1.3}$$

We can point out (see also R. R. BAHADUR [1]) that *if* $\zeta \in L^2$ *then* $\mathbf{A}_\xi \zeta = \mathbf{M}(\zeta \mid \xi)$ *with probability 1, i. e. the regression (conditional expected value) of* ζ *on* ξ *coincides with the orthogonal projection of* ζ *on* L^2_ξ. Namely, let $\mathcal{A} \in S_\xi$ — for which there exists a Borel set B such that $\mathcal{A} = \{\xi \in B\}$ — then especially for

$$f = \chi_{\mathcal{A}} = \begin{cases} 1 & \xi \in B \\ 0 & \xi \notin B\,; \end{cases}$$

$f \in L^2_\xi$ and from (1.1)

$$\int_{\mathcal{A}} \mathbf{A}_\xi \zeta\,d\mathbf{P} = \int_{\mathcal{A}} \zeta\,d\mathbf{P}$$

wherefrom $\mathbf{A}_\xi \zeta = \mathbf{M}(\zeta \mid \xi)$ with probability 1.

1.3. The correlation ratio of the random variable $\zeta \in L^2$ on ξ is according to the above

$$\theta_\xi(\zeta) = \frac{\mathbf{D}(\mathbf{M}(\zeta \mid \xi))}{\mathbf{D}(\zeta)} = \frac{\|\mathbf{M}(\zeta - \mathbf{M}(\zeta) \mid \xi\|}{\|\zeta - \mathbf{M}(\zeta)\|} = \|\mathbf{A}_\xi \zeta^*\| \tag{1.4}$$

from which

$$\theta^2_\xi(\zeta) = (\mathbf{A}_\xi \zeta^*, \mathbf{A}_\xi \zeta^*) = (\zeta^*, \mathbf{A}_\xi \zeta^*)\,. \tag{1.5}$$

Dividing (1.5) by (1.4) we have

$$\theta_\xi(\zeta) = \frac{(\zeta^*, \mathbf{A}_\xi \zeta^*)}{\|\mathbf{A}_\xi \zeta^*\|} = \mathbf{R}(\zeta, \mathbf{A}_\xi \zeta)\,. \tag{1.6}$$

If $\zeta \in L^2$ and $f \in L^2_\xi$,

$$\|\mathbf{A}_\xi \zeta^* - (f^*, \zeta^*) f^*\|^2 = \|\mathbf{A}_\xi \zeta^*\|^2 - (f^*, \zeta^*)^2$$

hence

$$\theta^2_\xi(\zeta) = (f^*, \zeta^*)^2 + \|\mathbf{A}_\xi \zeta^* - (f^*, \zeta^*) f^*\|^2 \qquad \text{(see Fig. 1)}. \tag{1.7}$$

It is evident, that in case $f \in L^2_\xi$ we have $L^2_f \subset L^2_\xi$, further $L^2_f = L^2_\xi$ if and only if $f(x)$ is univalent, i. e. if it has an inverse function. Hence as for $\zeta \in L^2$ the relation $\mathbf{A}_f \mathbf{A}_\xi \zeta = \mathbf{A}_f \zeta$ is valid,

$$\|\mathbf{A}_f \zeta\| \leq \|\mathbf{A}_\xi \zeta\| \tag{1.8}$$

follows, where equality holds if and only if $\mathbf{A}_\xi \zeta$ is a function of f (in this case $\mathbf{A}_f \zeta = \mathbf{A}_\xi \zeta$ holds, too).

1.4. As $\zeta_c \equiv c \in L^2$ for any real number c, consequently the subspace orthogonal to ζ_c may be considered and denoted by L_0^2. This is the space of random variables with expected values zero and finite standard deviations. Analogously, $L_{F,0}^2$ for such a subspace of L_F^2 and $L_{\xi,0}^2$ for L_ξ^2.

Let us consider ξ and η, an arbitrary pair of random variables. *In 1.4. and 1.5 the domains of the operators* $\mathbf{A}_\eta$ *and* $\mathbf{A}_\xi$ *are restricted to the spaces* $L_{\xi,0}^2$ *and* $L_{\eta,0}^2$, *respectively.*

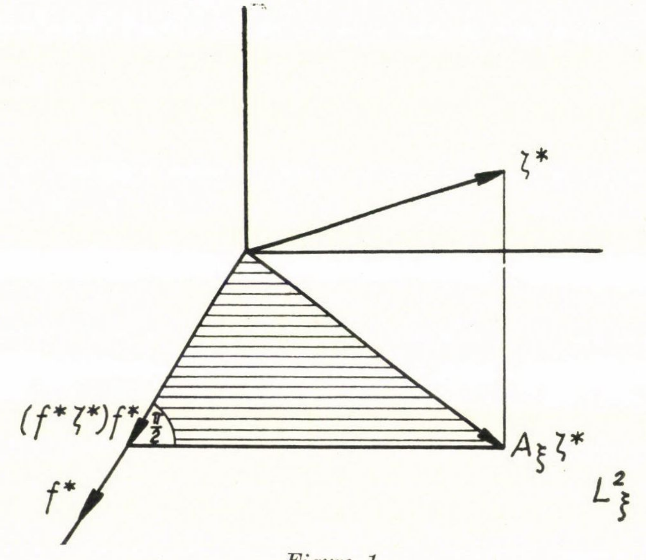

Figure 1.

The maximal correlation of ξ and η is defined as

$$\mathbf{S}(\xi, \eta) = \sup_{\substack{f \in L_\xi^2 \\ g \in L_\eta^2}} \mathbf{R}(f, g)\,. \tag{1.9}$$

This measure of stochastic connection was first defined by H. GEBELEIN [3]. Recently O. SARMANOV [6], [7] dealt with this problem and A. RÉNYI [5] generalized the notion.

Relating to the maximal correlation the following lemma is true:

Lemma 1.

$$\mathbf{S}(\xi, \eta) = \|\mathbf{A}_\eta\| = \|\mathbf{A}_\xi\|\,. \tag{1.10}$$

Proof. As

$$\sup_{\substack{f \in L_{\xi,0}^2 \\ \|f\|=1}} \|\mathbf{A}_\eta f\| = \sup_{\substack{f \in L_{\xi,0}^2 \\ \|f\|=1}} \left(\frac{\mathbf{A}_\eta f}{\|\mathbf{A}_\eta f\|}, f\right) \leq \mathbf{S}(\xi, \eta)$$

and in case $f \in L_{\xi,0}^2$, $\|f\| = 1$, $g \in L_{\eta,0}^2$, $\|g\| = 1$, according to equation (1.7)

$$|(f, g)| \leq \left(\frac{\mathbf{A}_\eta f}{\|\mathbf{A}_\eta f\|}, f\right),$$

thus

$$\mathbf{S}(\xi, \eta) \leq \sup_{\substack{f \in L^2_{\xi,0} \\ \|f\|=1}} \left(\frac{\mathbf{A}_\eta f}{\|\mathbf{A}_\eta f\|}, f \right),$$

consequently

$$\mathbf{S}(\xi, \eta) = \sup_{\substack{f \in L^2_{\xi,0} \\ \|f\|=1}} \|\mathbf{A}_\eta f\| = \|\mathbf{A}_\eta\|.$$

Similarly, $\mathbf{S}(\xi, \eta) = \|\mathbf{A}_\xi\|$. Thus our Lemma is proved.

A. RÉNYI [5] has shown that in the space $L^2_{\xi,0}$ the operator $\mathbf{A}_\xi \mathbf{A}_\eta$ is self-adjoint, positive definite and

$$\mathbf{S}^2(\xi, \eta) = \sup_{\substack{f \in L^2_{\xi,0} \\ \|f\|=1}} (\mathbf{A}_\xi \mathbf{A}_\eta f, f) = \|\mathbf{A}_\xi \mathbf{A}_\eta\|. \tag{1.11}$$

If

$$\begin{cases} \mathbf{A}_\eta f = \lambda g \\ \mathbf{A}_\xi g = \lambda f \end{cases} \quad f \in L^2_{\xi,0},\ g \in L^2_{\eta,0} \tag{1.12}$$

holds then we call λ an eigenvalue and f, g a pair of eigenfunctions of the pair of operators $\mathbf{A}_\eta$, $\mathbf{A}_\xi$. According to the results of the authors mentioned above, in (1.9) the supremum is the maximum if and only if $\mathbf{S}(\xi, \eta)$ is the highest of the values $|\lambda|$ in (1.12). In case of completely continuous $\mathbf{A}_\xi \mathbf{A}_\eta$, this condition is fulfilled.

1.5. Let $\{f_n\}$ and $\{g_n\}$ be complete orthonormal systems in $L^2_{\xi,0}$ and $L^2_{\eta,0}$, respectively. The positive square roots of the quantities

$$\begin{cases} |||\mathbf{A}_\eta|||^2 = \sum_i \sum_k (\mathbf{A}_\eta f_i, g_k)^2 \\ |||\mathbf{A}_\xi|||^2 = \sum_i \sum_k (f_i, \mathbf{A}_\xi g_k)^2 \end{cases} \tag{1.13}$$

are called the double norms of $\mathbf{A}_\eta$ resp. $\mathbf{A}_\xi$. Of course, these double norms may be infinite, too. As

$$\|\mathbf{A}_\eta f_i\|^2 = \sum_k (\mathbf{A}_\eta f_i, g_k)^2 \quad \text{and} \quad \|\mathbf{A}_\xi g_k\|^2 = \sum_i (f_i, \mathbf{A}_\xi g_k)^2$$

hence

$$|||\mathbf{A}_\eta|||^2 = \sum_i \|\mathbf{A}_\eta f_i\|^2 \quad \text{and} \quad |||\mathbf{A}_\xi|||^2 = \sum_k \|\mathbf{A}_\xi g_k\|^2. \tag{1.14}$$

From (1.1) and (1.13)

$$|||\mathbf{A}_\eta|||^2 = \sum_i \sum_k (f_i, g_k)^2 = |||\mathbf{A}_\xi|||^2. \tag{1.15}$$

By means of the double norm the notion of the mean square contingency introduced by K. PEARSON may be extended to arbitrary pairs of random variables ξ, η, by defining the contingency as

$$\mathbf{C}(\xi, \eta) = |||\mathbf{A}_\eta||| = |||\mathbf{A}_\xi|||. \tag{1.16}$$

According to the definition and (1.14)

$$\mathbf{C}^2(\xi, \eta) = \sum_i \theta_\eta^2(f_i) = \sum_k \theta_\xi^2(g_k)\,. \tag{1.17}$$

As shown later this definition is equivalent to that proposed by A. Rényi [4].

§ 2. An operator discussion of bivariate problems

2.1. Let the joint distribution function of ξ and η be $H(x, y)$ generating the measure P on the plane $[x, y]$; similarly, the marginal distribution functions $H_1(x)$ of ξ and $H_2(y)$ of η generating the measures P_1 and P_2, respectively, on the real line. The spaces L_ξ^2 and $L_{H_1}^2$, resp. L_η^2 and $L_{H_2}^2$ are isomorphic in the sense explained in 1.1. To the operator $\mathbf{A}_\eta$ defined in the space L_ξ^2 there corresponds an operator $\mathbf{A}_1$ transforming the elements of $L_{H_1}^2$ into $L_{H_2}^2$ in the following way: let be $f(x) \in L_{H_1}^2$ and $f = f(\xi) \in L_\xi^2$; then to the random variable $g = \mathbf{A}_\eta f$ there exists such a unique $g(y) \in L_{H_2}^2$ that $g = g(\eta)$. Then $\mathbf{A}_1$ is the operator which transforms $f(x)$ into this $g(y)$. Similarly, to $\mathbf{A}_\xi$ there corresponds an operator $\mathbf{A}_2$ transforming the elements of $L_{H_2}^2$ into $L_{H_1}^2$. It follows from the construction that $\mathbf{A}_1$ and $\mathbf{A}_2$, per analogiam, have the properties of $\mathbf{A}_\eta$ resp. $\mathbf{A}_\xi$. The question arises, under which conditions are $\mathbf{A}_1$ and $\mathbf{A}_2$ integral operators. Concerning this problem there holds the following

Theorem 1. *The pair of operators $\mathbf{A}_1$, $\mathbf{A}_2$ forms a pair of integral operators if and only if $P \ll P_1 \times P_2$. In this case there exists a function $K(x, y)$ such that $P(E) = \iint_E K(x, y)\, dP_1\, dP_2$ for any Borel set E in the plane $[x, y]$; furthermore,*

$$\mathbf{A}_1 f(x) = \int_{-\infty}^{\infty} K(x, y) f(x)\, dH_1(x) \quad \text{and} \quad \mathbf{A}_2\, g(y) = \int_{-\infty}^{\infty} K(x, y)\, g(y)\, dH_2(y)$$

hold.

Proof. If $P \ll P_1 \times P_2$ then according to the Radon—Nikodym theorem there exists such a function $K(x, y)$ that for any Borel set E in the plane $[x, y]$

$$P(E) = \iint_E K(x, y)\, dH_1(x)\, dH_2(y) \tag{2.1}$$

holds. For $f(x) \in L_{H_1}^2$ and $g(y) \in L_{H_2}^2$

$$(f(\xi), g(\eta)) = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} f(x)\, g(y)\, dH(x, y) = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} K(x, y) f(x)\, g(y)\, dH_1(x)\, dH_2(y)\,,$$

consequently $K(x, y) f(x)\, g(y)$ is integrable according to the measure $P_1 \times P_2$ for any $f(x) \in L_{H_1}^2$, $g(y) \in L_{H_2}^2$. The Fubini theorem implies that

$$k(y) = \int_{-\infty}^{\infty} K(x, y) f(x)\, dH_1(x)$$

exists for any $f(x) \in L_{H_1}^2$ and

$$\int_{-\infty}^{\infty} k(y)\, g(y)\, dH_2(y)$$

exists for any $g(y) \in L_{H_2}^2$. Hence, (see e. g. A. C. Zaanen [8] p. 137), $k(y) \in L_H^2$

follows. Let be $k = k(\eta) \in L^2_\eta$. Then in case $f = f(\xi)$ and arbitrary $g = g(\eta) \in L^2_\eta$

$$(f, g) = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} K(x, y) f(x) g(y) \, dH_1(x) \, dH_2(y) = \int_{-\infty}^{\infty} k(y) g(y) \, dH_2(y) = (k, g)$$

wherefrom $(f - k, g) = 0$ for any $g \in L^2_\eta$; this results $k = \mathbf{A}_\eta f$, i. e.

$$\mathbf{A}_1 f(x) = \int_{-\infty}^{\infty} K(x, y) f(x) \, dH_1(x) \quad \text{for any } f(x) \in L^2_{H_1}; \tag{2.2}$$

analogously,

$$\mathbf{A}_2 g(y) = \int_{-\infty}^{\infty} K(x, y) g(y) \, dH_2(y) \quad \text{for any } g(y) \in L^2_{H_2}. \tag{2.3}$$

Conversely, if for operator $\mathbf{A}_1$ there exists a function $K(x, y)$ specified by (2.2), then for any pair of Borel sets A, B—using the notations $\mathcal{A} = \{\xi \in A\}$, $\mathcal{B} = \{\eta \in B\}$; $\chi_{\mathcal{A}} = \begin{cases} 1 & \xi \in A \\ 0 & \xi \notin A \end{cases}$ and $\chi_{\mathcal{B}} = \begin{cases} 1 & \eta \in B \\ 0 & \eta \notin B \end{cases}$ — the following are true:

$$P(A \times B) = P(\xi \in A, \eta \in B) = \mathbf{M}(\chi_{\mathcal{A}} \chi_{\mathcal{B}}) = (\chi_{\mathcal{A}}, \chi_{\mathcal{B}}) =$$
$$= (\mathbf{A}_\eta \chi_{\mathcal{A}}, \chi_{\mathcal{B}}) = \big(\mathbf{A}_1 \chi_A(x), \chi_B(y)\big) = \iint_{A \times B} K(x, y) \, dH_1(x) \, dH_2(y).$$

For any Borel set E in the plane the above equation can be uniquely extended to (2.1) from which $P \ll P_1 \times P_2$ follows. (Similarly from $\mathbf{A}_2$ being an integral operator.) Thus the Theorem is proved.

Remark. The proof shows that either both of $\mathbf{A}_1$ and $\mathbf{A}_2$ are integral operators or none of them.

A. RÉNYI [4] defined the contingency as

$$\mathbf{C}^2(\xi, \eta) = \begin{cases} \displaystyle\int_{-\infty}^{\infty} \int_{-\infty}^{\infty} [K(x, y) - 1]^2 \, dH_1(x) \, dH_2(x) = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} K^2(x, y) \, dH_1(y) \, dH_2(y) - 1 & \text{if } P \ll P_1 \times P_2 \\ \infty & \text{in all other cases} \end{cases} \tag{2.4}$$

(where $K(x, y)$ is the function figuring in (2.1)).

A. RÉNYI proves his definition to involve the classical one for cases treated by K. PEARSON.

Theorem 2. *The definition of contingency given in* (1.16) *is equivalent to that of* (2.4).

Proof. Let

$$\{f_0(x), f_1(x), \ldots\} \quad \text{and} \quad \{g_0(y), g_1(y), \ldots\}$$

be complete orthonormal systems in $L^2_{H_1}$ resp. $L^2_{H_2}$, further $f_0(x) \equiv g_0(y) \equiv 1$. Then $\{f_1(x), f_2(x), \ldots\}$ and $\{g_1(y), g_2(y), \ldots\}$ will be complete orthonormal systems in the spaces $L^2_{H_1,0}$ resp. $L^2_{H_2,0}$.

Therefore

(2.5) $$\mathbf{C}^2(\xi,\eta)=|||\,\mathbf{A}_\eta\,|||^2=\sum_{i\geq 1}\sum_{k\geq 1}(\mathbf{A}_1 f_i(x),g_k(y))^2=\sum_{i\geq 0}\sum_{k\geq 0}(\mathbf{A}_1 f_i(x),g_k(y))^2-1\,,$$

viz. in case $i \geq 1$

$$\big(\mathbf{A}_1 f_i(x), g_0(y)\big)=\big(f_i(x), \mathbf{A}_2 g_0(y)\big)=(f_i(x), f_0(x))=0\,,$$

similarly in case $k \geq 1$

$$\big(\mathbf{A}_1 f_0(x), g_k(y)\big)=\big(g_0(y), g_k(y)\big)=0;$$

further

$$\big(\mathbf{A}_1 f_0(x), g_0(y)\big)=1\,.$$

If $P \ll P_1\times P_2$, i. e. by Theorem 1 if $\mathbf{A}_1$ is an integral operator then

(2.6) $$|||\,\mathbf{A}_1\,|||^2=\int_{-\infty}^{\infty}\int_{-\infty}^{\infty}K^2(x,y)\,dH_1(x)\,dH_2(y)=\sum_{i\geq 0}\sum_{k\geq 0}(\mathbf{A}_1 f_i(x),g_k(y))^2\,.$$

From (2.5) and (2.6) $\mathbf{C}^2(\xi,\eta)=|||\,\mathbf{A}_\eta\,|||^2=|||\,\mathbf{A}_1\,|||^2-1$, which is identical to (2.4).

On the other hand, if $|||\,\mathbf{A}_1\,|||<\infty$, then $\mathbf{A}_1$ is known to be a completely continuous integral operator, thus according to Theorem 1 $P \ll P_1\times P_2$ must hold; consequently if the latter does not hold, then $\mathbf{C}(\xi,\eta)=|||\,\mathbf{A}_1\,|||=\infty$. Thus our statement is proved.

2.2. Let the measure P be generated on the plane by the joint distribution of ξ and η, further let be

(2.7) $$\begin{cases} P_A(B)=P(A\times B)\\ P^B(A)=P(A\times B)\end{cases}$$

where A and B are arbitrary Borel sets. $P_A(B)$ for fixed A and $P^B(A)$ for fixed B are measures on the class of Borel sets on the real line. As

(2.8) $$P_A(B)\leq P_2(B) \quad\text{and}\quad P^B(A)\leq P_1(A)\,,$$

consequently $P_A \ll P_2$ and $P^B \ll P_1$. According to the Radon—Nikodym theorem there exist functions $P_1(A\mid y)$ and $P_2(B\mid x)$ such that

(2.9) $$\begin{cases} P_A(B)=\int_B P_1(A\,|\,y)\,dP_2\\ P^B(A)=\int_A P_2(B\,|\,x)\,dP_1\,,\end{cases}$$

respectively.

Further let be $\chi_A(x)=\begin{cases}1 & x\in A\\ 0 & x\notin A\end{cases}$ where A is any Borel set on the real line; then

$$\int_B \mathbf{A}_1\,\chi_A(x)\,dP_2=\int_{-\infty}^{\infty}\chi_B(y)\,\mathbf{A}_1\,\chi_A(x)\,dH_2(y)=\big(\mathbf{A}_1\,\chi_A(x),\chi_B(y)\big)=P_A(B)\,,$$

therefore

$$\mathbf{A}_1\, \chi_A(x) = P_1(A \mid y) \tag{2.10}$$

for almost every y according to the measure P_2. Similarly,

$$\mathbf{A}_2\, \chi_B(y) = P_2(B \mid x) \tag{2.11}$$

for almost every x according to the measure P_1. Thus $P_1(A \mid y)$ resp. $P_2(B \mid x)$ are the conditional distributions, which generate probability measures (see J. L. Doob [2] p. 29).

If $P_1(A) = 0$ then by (2.8)

$$P_A(B) = \int_B P_1(A \mid y)\, dP_2 = 0\,,$$

whence $P_1(A \mid y) = 0$ follows for almost every y according to the measure P_2. This does not imply however, $P_1(A \mid y) \ll P_1(A)$. As to the validity of the latter there holds the following

Lemma 2. $P_1(A \mid y) \ll P_1(A)$, $P_2(B \mid x) \ll P_2(B)$ *and* $P \ll P_1 \times P_2$ *do or do not hold simultaneously*

Proof. If $P \ll P_1 \times P_2$ then $P_1(A \mid y) = \int_A K(x, y)\, dP_1$ and $P_2(B \mid x) = \int_B K(x, y)\, dP_2$ because of

$$\int_B \left[\int_A K(x, y)\, dP_1\right] dP_2 = P(A \times B) = P_A(B)$$

and

$$\int_A \left[\int_B K(x, y)\, dP_2\right] dP_1 = P(A \times B) = P^B(A)\,,$$

wherefrom $P_1(A \mid y) \ll P_1(A)$ resp. $P_2(B \mid x) \ll P_2(B)$ follow.

Further if $P_1(A \mid y) \ll P_1(A)$ then according to the Radon—Nikodym theorem there exists a function $K(x, y)$ for which

$$P_1(A \mid y) = \int_A K(x, y)\, dP_1\,.$$

For this function $K(x, y)$

$$\iint_{A \times B} K(x, y)\, dP_1\, dP_2 = \int_B \left[\int_A K(x, y)\, dP_1\right] dP_2 = \int_B P_1(A \mid y)\, dP_2 = P(A \times B)\,,$$

which means that $P \ll P_1 \times P_2$. (Similarly from $P_2(B \mid x) \ll P_2(B)$.)

Thus our lemma is true.

Corollary. *In consequence of Theorem* 1 *and Lemma* 2, *the following three statements are equivalent :*

1° $P_1(A \mid y) \ll P_1(A)$ $\big($*and* $P_2(B \mid x) \ll P_2(B)\big)$
2° $P \ll P_1 \times P_2$
3° $\mathbf{A}_1$ *is an integral operator (and* $\mathbf{A}_2$ *too).*

2.3. In many cases by the symmetry of the bivariate distribution in its variables $\big(H(x, y) = H(y, x)\big)$ the solution of the problems discussed before may be facilitated. Namely, if $H(x, y) = H(y, x)$, the problem reduces to the spectral decomposition of a single operator. Viz. if we have a symmetric distribution $P(A \times B) = P(B \times A)$ then $P_1 = P_2$, thus $L^2_{H_1} = L^2_{H_2}$ and by this $\mathbf{A}_1 = \mathbf{A}_2$. The pair of operator equations becomes

$$\begin{cases} \mathbf{A}_1 f(x) = \lambda g(y) \\ \mathbf{A}_1 g(x) = \lambda f(y) . \end{cases} \tag{2.12}$$

If $f(x) = g(x)$, these equations are identical and $f(x)$ is an eigenfunction of $\mathbf{A}_1$. If $f(x) \neq g(x)$, then by subtracting the second equation of (2.12) from the first one we obtain

$$\mathbf{A}_1\big(f(x) - g(x)\big) = -\lambda\big(f(y) - g(y)\big)$$

whence $f(x) - g(x)$ is an eigenfunction of $\mathbf{A}_1$. Above considerations show the possibility of reducing the spectral decomposition of a pair of operators to that of a single operator.

It will be shown that any bivariate distribution may be usefully replaced by a symmetric one. Let us consider instead of the original bivariate distribution P the symmetric distribution $\overline{P}$ defined by

$$\overline{P}(A \times B) = \big(\mathbf{A}_2\, \mathbf{A}_1\, \chi_A(x), \chi_B(x)\big) = \big(\mathbf{A}_1\, \chi_A(x), \mathbf{A}_1\, \chi_B(x)\big), \tag{2.13}$$

where A and B are arbitrary Borel sets on the real line and the first scalar product is calculated in $L^2_{H_1}$ and the second one in $L^2_{H_2}$, respectively. It is easy to see that $\overline{P}$ is a probability measure on the plane. Furthermore, trivially $\overline{P}(A \times B) = \overline{P}(B \times A)$ and for the corresponding distribution function $\overline{H}(x, y) = \overline{H}(y, x)$. It is also obvious that both marginal distributions of $\overline{P}$ are equal to the distribution of ξ, i. e.

$$\overline{P}_1 = \overline{P}_2 = P_1$$

and thus

$$L^2_{\overline{H}_1} = L^2_{\overline{H}_2} = L^2_{H_1} .$$

The last term of (2.13) — equaling $\int\limits_{-\infty}^{\infty} P_1(A \mid y)\, P_1(B \mid y)\, dH_2(y)$ — suggests the following interpretation of the distribution $\overline{P}$: Let us fix any arbitrary value of η and choose an independent pair of values of ξ; then $\overline{P}$ is the mixture of the distributions of such pairs with respect to η.

Such transformations were proposed in special cases by H. GEBELEIN [3] and O. SARMANOV [7], too. The role of both operators $\mathbf{A}_1$ and $\mathbf{A}_2$ will be taken over by $\mathbf{A}_2\, \mathbf{A}_1$ according to the distribution $\overline{P}$, as shown in

Theorem 3.

$$\int\limits_{-\infty}^{\infty} \int\limits_{-\infty}^{\infty} f(x)\, g(y)\, d\overline{H}(x, y) = \big(\mathbf{A}_2\, \mathbf{A}_1\, f(x), g(x)\big) \tag{2.14}$$

where $f(x) \in L^2_{H_1}$ *and* $g(x) \in L^2_{H_1}$.

Proof. According to the definition of $\overline{P}$ in (2.13), for characteristic functions (2.14) holds. Therefore our statement is valid for any pair of step functions. Furthermore any function can be approximated to any desired exactness by step functions. Thus in consequence of the continuity of its terms, (2.14) holds in general.

Remark. It is evident from (2.14) that

$$\int_{-\infty}^{\infty}\int_{-\infty}^{\infty} f(x)\,f(y)\,d\overline{H}(x,y) = \theta_\eta^2(f(\xi)) \qquad f(x)\in L^2_{H_1,0},\ \|f(x)\| = 1\,.$$

Corollary 1. $\overline{\mathbf{A}} = \mathbf{A}_2\,\mathbf{A}_1$ *is the operator generating the conditional expectation in the sense*

$$\overline{\mathbf{A}}\,f(x) = \int_{-\infty}^{\infty} f(x)\,d\overline{H}_1(x\,|\,y) \qquad f(x)\in L^2_{H_1} \tag{2.15}$$

where $\overline{H}_1\,(x\,|\,y)$ *denotes the conditional distribution function corresponding to* $\overline{H}(x,y)$.

Proof. The definition of the conditional expectation — see (1.2) — may be written in function terminology in the form

$$\int_A \mathbf{A}_1 f(x)\,dP_2 = \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} f(x)\,\chi_A(y)\,dP \qquad f(x)\in L^2_{H_1} \tag{2.16}$$

for any Borel set A on the real line. We shall see that $\overline{\mathbf{A}}f(x)$ satisfies (2.16) and therefore (2.15), too. Namely, because of Theorem 3

$$\int_A \overline{\mathbf{A}}\,f(x)\,dP_1 = (\overline{\mathbf{A}}\,f(x),\chi_A(x)) = \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} f(x)\,\chi_A(y)\,d\overline{P}.$$

Hence also $\overline{P_1}\,(A\,|\,y) = \overline{\mathbf{A}}\chi_A\,(x)$ holds and the correlation ratios concerning $\overline{P}$ equal $\|\,\overline{\mathbf{A}}f(x)\,\|$ $\;(f(x)\in L^2_{H_1,0}\,,\ \|\,f(x)\,\| = 1)$.

Corollary 2. *The maximal correlation of* $\overline{H}(x,y)$ *is equal to the square of that of* $H(x,y)$.

Proof. In consequence of (1.11) and (2.14)

$$\mathbf{S}^2(\xi,\eta) = \sup_{\substack{f(x)\in L^2_{H_1,0}\\ \|f(x)\|=1}} \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} f(x)\,f(y)\,d\overline{P} \leq \sup_{\substack{f(x)\in L^2_{H_1,0},\|f(x)\|=1\\ g(x)\in L^2_{H_1,0},\|g(x)\|=1}} \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} f(x)\,g(y)\,d\overline{P}.$$

On the other hand from (2.14)

$$\sup_{\substack{f(x)\in L^2_{H_1,0},\|f(x)\|=1\\ g(x)\in L^2_{H_1,0},\|g(x)\|=1}} \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} f(x)\,g(y)\,d\overline{P} = \sup_{\substack{f(x)\in L^2_{H_1,0},\|f(x)\|=1\\ g(x)\in L^2_{H_1,0},\|g(x)\|=1}} \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} (\mathbf{A}_2\,\mathbf{A}_1\,f(x),g(x)) =$$

$$= \sup_{\substack{f(x)\in L^2_{H_1,0},\|f(x)\|=1\\ g(x)\in L^2_{H_1,0},\|g(x)\|=1}} (\mathbf{A}_1\,f(x),\mathbf{A}_1\,g(x)) \leq \sup_{\substack{f(x)\in L^2_{H_1,0},\|f(x)\|=1\\ g(x)\in L^2_{H_1,0},\|g(x)\|=1}} \|\mathbf{A}_1 f(x)\|\,\|\mathbf{A}_1 g(x)\| = \mathbf{S}^2(\xi,\eta).$$

Hence our statement follows.

The mean square contingency belonging to the measure $\overline{P}$ equals

$$\sum_i \sum_k (\mathbf{A}_1 f_i(x), \mathbf{A}_1 f_k(x))^2 = \sum_i \| \overline{\mathbf{A}} f_i(x) \|^2 \tag{2.17}$$

where $\{f_i(x)\}$ is a complete orthonormal system in $L^2_{H_1,0}$. If $\mathbf{A}_1$ is an integral operator with kernel $K(x, y)$ then $\overline{\mathbf{A}}$ is an integral operator with kernel

$$\overline{K}(x, y) = \int_{-\infty}^{\infty} K(x, z)\, K(y, z)\, dH_2(z) \tag{2.18}$$

and (2.17) can be calculated as

$$\int_{-\infty}^{\infty} \int_{-\infty}^{\infty} \overline{K}^2(x, y)\, dH_1(x)\, dH_1(y) - 1 .$$

If the pair of operators $\mathbf{A}_1$, $\mathbf{A}_2$ has a pair of eigenfunctions $f(x)$, $g(y)$ belonging to the eigenvalue λ then $f(x)$ is an eigenfunction of $\overline{\mathbf{A}}$ and belongs to λ^2. (This can be seen by substituting one of the operator equations into the other.)

It is easy to see that (2.17) — if it is finite — is not less than $\sum_i \theta_\eta^4 (f_i)$ and the equality holds if and only if $\{f_i(x)\}$ is a system of eigenfunctions (for $(\mathbf{A}_2 \mathbf{A}_1 f_i(x), f_i(x)) = \| \mathbf{A}_2 \mathbf{A}_1 f_i(x) \|$ holds in the case $\mathbf{A}_2 \mathbf{A}_1 f_i(x) = \lambda_i^2 f_i(x)$). If $\overline{\mathbf{A}}$ has a complete eigenfunction system, (2.17) may be written as $\sum_i \lambda_i^4$ where the λ_i-s are the eigenvalues of the pair of operators $\mathbf{A}_1$, $\mathbf{A}_2$.

Evidently, all what has been stated in this part for $\mathbf{A}_2 \mathbf{A}_1$ may be shown analogously for $\mathbf{A}_1 \mathbf{A}_2$ as well.

The fact that both marginals of the distribution $\overline{P}$ equal one of the original marginal distributions is evidently advantageous in cases we have one discrete or differentiable marginal.

(Received April 28, 1960.)

REFERENCES

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

[2] Doob, J. L.: *Stochastic processes.* Wiley, New York, 1953.

[3] 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.

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

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

[6] Сарманов, О. В.: „Максимальный коэффициент корреляции (симметричный случай)". *Доклады Академии Наук* **120** (1958) 715—718.

[7] Сарманов, О. В.: „Максимальный коэффициент корреляции (несимметричный случай)". *Доклады Академии Наук* **121** (1958) 52—55.

[8] Zaanen, A. C.: *Linear analysis.* North-Holland Publishing Co., Amsterdam, 1956.

О СТОХАСТИЧЕСКИХ СВЯЗЯХ С ДВУМЯ ПЕРЕМЕННЫМИ

P. CSÁKI и J. FISCHER

Резюме

Работа исследует ряд свойств стохастических связей с двумя переменными с помощью гильбертовых пространств. Этот метод исследования возможен потому, что некоторые характеристики стохастической связи между случайными величинами ξ и η (условное распределение, условное математическое ожидание, ряд мер силы стохастической связи) могут быть описаны определенными в соответствующих гильбертовых пространствах парами операторов $\mathbf{A}_\xi$, $\mathbf{A}_\eta$ или $\mathbf{A}_1$, $\mathbf{A}_2$. Так условное распределение ξ относительно η

$$P_1(A\,|\,y) = \mathbf{A}_1\,\chi_A(x)\,,$$

условное математическое ожидание

$$\mathbf{M}(\xi\,|\,\eta) = \mathbf{A}_\eta\,\xi\,,$$

корреляционное отношение

$$\theta_\eta(\xi) = \|\,\mathbf{A}_\eta\,\xi\,\|$$

(относительно стандартного ξ). Далее максимальная корреляция ξ и η

$$\mathbf{S}(\xi,\eta) = \|\,\mathbf{A}_\eta\,\| = \|\,\mathbf{A}_\xi\,\|\,;$$

а понятие контингенции обобщается для любой пары случайных величин ξ, η:

$$\mathbf{C}(\xi,\eta) = |||\,\mathbf{A}_\eta\,||| = |||\,\mathbf{A}_\xi\,|||\,.$$

Доказывается, что $\mathbf{A}_1$, $\mathbf{A}_2$ тогда и только тогда пара интегральных операторов, если $P \ll P_1 \times P_2$ (теорема 1). На основании этого доказывается, что вышеприведенное определение контингенции совпадает с определением A. Rényi (теорема 2).

С помощью некоторой леммы, обобщая теорему 1, доказывается эквивалентность следующих трех определений:

1. $P_1(A|y) \ll P_1(A)$ и $P_2(B|x) \ll P_2(B)$,
2. $P \ll P_1 \times P_2$,
3. $\mathbf{A}_1$ и $\mathbf{A}_2$ интегральные операторы.

С помощью оператора $\overline{\mathbf{A}} = \mathbf{A}_1\,\mathbf{A}_2$ дается метод замены любой плоскостной вероятностной меры P симметричной мерой $\overline{P}$, эквивалентной с точки зрения рассматриваемых вопросов. В связи с этим доказывается, что

$$\int_{-\infty}^{+\infty}\int_{-\infty}^{+\infty} f(x)\,g(y)\,d\overline{P} = \big(\overline{\mathbf{A}}f(x),\,g(x)\big)$$

(теорема 3). На основании этого доказывается, что относящаяся к $\overline{P}$ максимальная корреляция $\mathbf{S}^2(\xi,\eta)$.