



---

Soviet-era science, translated into English

# MATHEMATICS

O. V. SARMANOV

1958

SovietRxiv

---

View the original and related papers at <https://sovietrxiv.org/items/ru-195801.62220>

Source: Math-Net.Ru and CyberLeninka. Machine translation. Verify with the original.

**Abstract**

**Full Text**

**MATHEMATICS**

**O. V. SARMANOV**

## **THE MAXIMAL CORRELATION COEFFICIENT**

**(THE SYMMETRIC CASE)**

*(Presented by Academician S. N. Bernstein on 23 I 1958)*

1. Let  $F(x, y) = F(y, x) \geq 0$  be a symmetric distribution density defining a correlation between the random variables  $x$  and  $y$  in the square domain  $[a \leq x; y \leq b]$ , which may also be infinite.

Denote by

$$p(x) = \int_a^b F(x, y) dy$$

the a priori density of  $x$ , and suppose that the square of the kernel

$$K(x, y) = \frac{F(x, y)}{\sqrt{p(x)p(y)}}$$

is integrable in both variables.

As was shown in <sup>(1)</sup>, the spectrum of the kernel  $K(x, y)$  has the form

$$\lambda_0 = 1, \lambda_1, \lambda_2, \dots, \quad \varphi_0(x) = 1 \cdot \sqrt{p(x)}, \varphi_1(x) \sqrt{p(x)}, \varphi_2(x) \sqrt{p(x)}, \dots \quad (1)$$

where  $|\lambda_1| > 1$ , since  $1/\lambda_1$  is the correlation coefficient between the eigenfunctions  $\varphi_1(x)$  and  $\varphi_1(y)$ .

2. **Definition.**  $R^* = 1/\lambda_1$  will be called the **maximal** (in absolute value) **correlation coefficient corresponding to the density  $F(x, y)$** . The name is justified by the extremal property of the eigenvalues of a symmetric kernel.

If one seeks the maximum of the modulus of the expression

$$I = \int_a^b \int_a^b \varphi(x) \varphi(y) F(x, y) dx dy \quad (2)$$

under the conditions

$$\int_a^b \varphi(x) p(x) dx = 0, \quad \int_a^b \varphi^2(x) p(x) dx = 1, \quad (3)$$

then this maximum is attained for  $\varphi(x) = \varphi_1(x)$  and is equal to  $1/|\lambda_1|$ .

3. The first eigenfunction  $\varphi_1(x)$ , together with the coefficient  $R^* = 1/\lambda_1$ , is computed by the method of successive approximations, whose convergence was proved, for example, in (1).

As the “zero approximation”  $r_0(x)$  one may take any function having a variance. If the first moment

$$c_0 = \int_a^b r_0(x)p(x) dx \neq 0,$$

then it is more convenient to replace  $r_0(x)$  by  $r_0(x) - c_0$ ; therefore, without loss of generality, we shall assume  $c_0 = 0$ .

Put

$$r_k(x) = \int_a^b r_{k-1}(y) \frac{F(x, y)}{p(x)} dy \quad (k = 1, 2, \dots); \quad (4)$$

then, as shown in (1),

$$c_1 \varphi_1(x) = \lim_{k \rightarrow \infty} r_k(x) \lambda_1^k; \quad (5)$$

$$c_1 = \int_a^b \varphi_1(x) r_0(x) p(x) dx$$

is determined by normalization.

If  $k$  is sufficiently large, then

$$R^* = \frac{1}{\lambda_1} \simeq \frac{r_k(x)}{r_{k-1}(x)}. \quad (6)$$

4. As is known, the ordinary correlation coefficient characterizes the dependence between  $x$  and  $y$  well only in the case of rectilinear correlation, and its vanishing (as well as the vanishing of the so-called correlation ratio, see (2), p. 385) does not imply independence of the random variables. The following theorem is therefore all the more important.

**Theorem 1.** *For independence of the random variables  $x$  and  $y$ , it is necessary and sufficient that the maximal correlation coefficient vanish.*

**Proof.** Necessity is obvious, since if  $x$  and  $y$  are independent, then  $\varphi_1(x)$  and  $\varphi_1(y)$  are also independent and the correlation coefficient between them is zero, i.e.  $R^* = 0$ .

Let now  $R^* = 1/\lambda_1 = 0$ ; then, according to (1), the kernel  $K(x, y)$  has no eigenfunctions except  $1 \cdot \sqrt{p(x)}$ .

Since the Fourier series

$$\sqrt{p(x)p(y)} + \sum_{i=1}^{\infty} \frac{\varphi_i(x)\varphi_i(y)\sqrt{p(x)p(y)}}{\lambda_i}$$

converges to  $K(x, y)$  in the mean, in the case  $1/\lambda_1 = 0$

$$\int_a^b \int_a^b \left[ \frac{F(x, y)}{\sqrt{p(x)p(y)}} - \sqrt{p(x)p(y)} \right]^2 dx dy = 0,$$

i.e.

$$F(x, y) = p(x)p(y) \tag{7}$$

for almost all  $x$  and  $y$ , as was required to prove.

**Theorem 2.** *If the correlation is rectilinear, then the ordinary correlation coefficient  $R$  between  $x$  and  $y$  coincides with the maximal correlation coefficient  $R^*$ .*

**Proof.** In this case the kernel  $K(x, y)$  has as its first eigenfunction a linear (and, consequently, monotone) function  $(x - c)/\sigma$ , where  $c$  is the mean and  $\sigma^2$  is the variance of  $x$ .

If the spectrum of a stochastic kernel contains a monotone function, then it always belongs to the first eigenvalue (1, 3); therefore the correlation coefficient between  $x$  and  $y$ , equal to the correlation coefficient between the functions  $(x - c)/\sigma$  and  $(y - c)/\sigma$ , coincides with the maximal correlation coefficient, as was required to prove.

The last theorem shows why the ordinary correlation coefficient characterizes only rectilinear correlation well—because only in this case is it the maximal correlation coefficient.

**5.** For discrete random variables the maximal correlation coefficient is defined analogously.

Let the correlation dependence between the discrete random variables  $x$  and  $y$  be defined by the square symmetric matrix

$$\{p_{ij}\}, \quad i; j = 1, 2, \dots, n, \tag{8}$$

where

$$0 \leq p_{ij} = p_{ji} = \mathbf{P}\{x = x_i, y = x_j\}; \quad \sum_{i,j} p_{ij} = 1; \tag{9}$$

$$p_i = \sum_{j=1}^n p_{ij} = \mathbf{P}\{x = x_i\} = \mathbf{P}\{y = x_i\}; \quad i = 1, 2, \dots, n.$$

Then  $R^* = 1/\lambda_1$ , where  $\lambda_1^{-1}$  is the first eigenvalue of the matrix

$$\left\{ \frac{p_{ij}}{\sqrt{p_i p_j}} \right\},$$

is called the **maximum coefficient of correlation between random variables with correlation table** (8).

The process of successive approximations for finding  $R^*$  does not differ in any essential way from that described in Sec. 3.

Let  $r_0(x)$  be an arbitrary vector with coordinates  $\{x_1^{(0)}, x_2^{(0)}, \dots, x_n^{(0)}\}$ , chosen so that

$$\sum_{i=1}^n x_i^{(0)} p_i = 0.$$

The  $k$ -th iteration is the vector whose coordinates are determined by the formulas

$$x_i^{(k)} = \sum_{j=1}^n \frac{p_{ij}}{p_i} x_j^{(k-1)}, \quad i = 1, 2, \dots, n; \quad k = 1, 2, \dots \quad (10)$$

If  $k$  is sufficiently large, then

$$R^* = \frac{1}{\lambda_1} \approx \frac{x_i^{(k)}}{x_i^{(k-1)}}. \quad (11)$$

**Remark.** If there is a rectilinear correlation between the discrete  $x$  and  $y$ , then as the initial vector  $r_0(x)$  one should take the vector  $r(x)$ , whose coordinates are the values of  $x$ ; then the very first iteration will lead to the goal, i.e. the ratio

$$\frac{x_i^{(1)} - \bar{x}}{x_i - \bar{x}} \approx \frac{1}{\lambda_1} = R^* = R$$

will be practically constant for  $i = 1, 2, \dots, n$ .

6. One can give a direct proof of Theorem 1 in the discrete case. Necessity is still obvious, and therefore we shall confine ourselves to proving sufficiency.

Let  $R^* = 0$ ; this means that the quadratic form

$$H = \sum_{i,j} \frac{p_{ij}}{\sqrt{p_i p_j}} \xi_i \xi_j \equiv 0 \quad (12)$$

for any set of  $n$  numbers  $\xi_i$  satisfying the conditions

$$\sum_{i=1}^n \xi_i^2 = 1, \quad (A)$$

$$\sum_{i=1}^n \xi_i \sqrt{p_i} = 0. \quad (B)$$

Let  $x_1, x_2, \dots, x_n$  be any  $n$  real numbers, among which there are at least two distinct ones. With their help form  $n$  numbers

$$\xi_i = \frac{x_i - c}{\sigma} \sqrt{p_i}, \quad i = 1, 2, \dots, n, \quad (13)$$

where

$$c = \sum_{i=1}^n x_i p_i; \quad \sigma^2 = \sum_{i=1}^n x_i^2 p_i - c^2.$$

The numbers (13) satisfy conditions (A) and (B); consequently, for them identity (12) holds, which is equivalent to the identity

$$H_1 = \sum_{i,j} (p_{ij} - p_i p_j) x_i x_j \equiv 0, \quad (14)$$

whence, in view of the arbitrariness of  $x_i$ , the conditions follow

$$p_{ij} = p_i p_j \quad (15)$$

for all  $i$  and  $j$ , which also means the independence of  $x$  and  $y$ .

7. If  $n = 2$ , then conditions (A) and (B) determine  $\xi_1$  and  $\xi_2$ , and the form  $H$  takes the constant value

$$H = \frac{p_{11} - p^2}{p - p^2}, \quad (16)$$

where  $p$  is the probability of one of the outcomes, for example “success,” and  $p_{11}$  is the probability of the joint occurrence of two successes. In this case  $R^*$

coincides with the value (16) and is equal to the so-called correlation coefficient between two events (see <sup>(2)</sup>, p. 33).

Mathematical Institute named after V. A. Steklov  
Academy of Sciences of the USSR

Received  
23 I 1958

## REFERENCES

- <sup>1</sup> O. V. Sarmanov, DAN, **53**, No. 9 (1946).
- <sup>2</sup> S. N. Bernstein, *Theory of Probability*, 2nd ed., 1946.
- <sup>3</sup> M. K. Nomokonov, DAN, **72**, No. 6 (1950).

*Note: Figure translations are in progress. See original paper for figures.*

*Source: Math-Net.Ru and CyberLeninka. Machine translation. Verify with the original.*