MATHEMATICS

O. V. SARMANOV and V. K. ZAKHAROV

1960

SovietRxiv

Abstract

Full Text

MATHEMATICS

O. V. SARMANOV and V. K. ZAKHAROV

MAXIMAL COEFFICIENTS OF MULTIPLE CORRELATION

(Presented by Academician S. N. Bernstein on 8 X 1959)

In the present paper the concept of the maximal coefficient of correlation between two random variables, considered in works ([1],[2]), is extended to multiple correlation between n random variables having a density.

1. Let the random variables $x_1, x_2, \ldots, x_n$ take values in the domain $\Omega : \{a_i \leq x_i \leq b_i;\ i = 1, 2, \ldots, n\}$. In particular, Ω may coincide with the whole space R_n.

Consider a subspace R_m of the space R_n, $m \leq n$. Let R_m be divided into two nonintersecting subspaces R_k and R_{m-k}, of dimensions k and $m-k$, respectively. Denote by Ω_1 the intersection of Ω with R_k, and by Ω_2 the intersection of Ω with R_{m-k}, and by Q, Q_1, and Q_2 arbitrary vectors of the subspaces R_m, R_k, and R_{m-k}, respectively.

Let $p(Q)$ be the density of the distribution in the domain $\Omega_1 + \Omega_2$; then the distribution densities in Ω_1 and Ω_2 are determined, respectively, by the integrals

$$p_1(Q_1) = \int_{\Omega_2} p(Q)\, dQ_2, \qquad p_2(Q_2) = \int_{\Omega_1} p(Q)\, dQ_1. \tag{1}$$

It is always assumed that the restriction

$$\int_{\Omega_1+\Omega_2} \frac{p^2(Q)}{p_1(Q_1)p_2(Q_2)}\, dQ < \infty \tag{2}$$

is satisfied.

Definition 1. The **maximal coefficient of correlation** $\bar{r}(Q_1, Q_2)$ between the random vectors Q_1 and Q_2 will be called the greatest, in absolute value, value of the integral

$$I(\varphi, \psi) = \int_{\Omega_1+\Omega_2} p(Q)\, \varphi(Q_1)\psi(Q_2)\, dQ \tag{3}$$

in the class of functions φ, ψ satisfying the conditions

$$\int_{\Omega_1} p_1(Q_1)\varphi(Q_1)\,dQ_1 = \int_{\Omega_2} p_2(Q_2)\psi(Q_2)\,dQ_2 = 0, \tag{4}$$

$$\int_{\Omega_1} p_1(Q_1)\varphi^2(Q_1)\,dQ_1 = \int_{\Omega_2} p_2(Q_2)\psi^2(Q_2)\,dQ_2 = 1. \tag{5}$$

Remark. From the extremal meaning of the eigenvalues of the kernel

$$\frac{p(Q)}{\sqrt{p_1(Q_1)p_2(Q_2)}}$$

it follows that the maximal coefficient of correlation between Q_1 and Q_2 is equal to the first eigenvalue $\frac{1}{\lambda_1}$ of this kernel ([1],[2]).

2. **Main theorem.** *For the complete independence of the random variables $x_1, x_2, \dots, x_n$, it is necessary and sufficient that all possible maximal correlation coefficients vanish, each of which is computed for a pair of vectors taken from nonintersecting subspaces of the space R_n (the sum of the dimensions of these vectors ranges from 2 to n).*

Proof. The necessity is obvious, since in the case of independence of the random variables any two vectors Q_1 and Q_2 will be independent, $p(Q) = p_1(Q_1)p_2(Q_2)$, and the integral (3) is identically equal to zero, since the admissible functions satisfy condition (4).

On the other hand, it is easy to show that from the vanishing of $\bar{r}(Q_1, Q_2)$ there follows the independence of Q_1 and Q_2. Indeed, the bilinear expansion of the kernel

$$\frac{p(Q)}{\sqrt{p_1(Q_1)p_2(Q_2)}} \sim \sqrt{p_1(Q_1)p_2(Q_2)} + \sum_{i=1}^{\infty} \frac{\varphi_i(Q_1)\sqrt{p_1(Q_1)}\psi_i(Q_2)\sqrt{p_2(Q_2)}}{\lambda_i} \tag{6}$$

converges to it in the mean by virtue of condition (2).

The vanishing of $\bar{r}(Q_1, Q_2)$, equivalent to the vanishing of $\frac{1}{\lambda_1}$ and, consequently, of all the other eigenvalues $\frac{1}{\lambda_i}$, $i = 2, 3, \dots$, leads to the equality

$$\int_{\Omega_1+\Omega_2} \left[\frac{p(Q)}{\sqrt{p_1(Q_1)p_2(Q_2)}} - \sqrt{p_1(Q_1)p_2(Q_2)}\right]^2 dQ = 0, \tag{7}$$

whence the independence of Q_1 and Q_2 follows.

If, however, any two vectors with different components from R_n are independent, then all n random variables are jointly independent, as was required to prove.

Remark 1. It is clear that the number of conditions ensuring independence can be greatly reduced; for example, it is sufficient to ensure the independence of $n-1$ pairs of vectors:

$$x_1 \text{ and } x_2; \quad (x_1, x_2) \text{ and } x_3; \dots; \quad (x_1, x_2, \dots, x_{n-1}) \text{ and } x_n,$$

i.e., it is sufficient to consider only $n-1$ coefficients

$$\bar{r}_{(1,2,\dots,i-1);\,i} = \bar{r}\{(x_1, x_2, \dots, x_{i-1}), x_i\}, \quad i = 2, 3, \dots, n. \tag{8}$$

Remark 2. If the density is symmetric with respect to $x_1, x_2, \dots, x_n$ (up to now symmetry has nowhere been assumed), then the number of conditions ensuring independence can be reduced still further; thus, for $n = 2^k$, instead of the $n-1$ coefficients (8), it is sufficient to consider only k coefficients

$$\bar{r}\{(x_1, x_2, \dots, x_i), (x_{i+1}, x_{i+2}, \dots, x_{2i})\}, \quad i = 1, 2, 2^2, \dots, 2^{k-1}. \tag{9}$$

From the vanishing of all coefficients (9) there follows the complete independence of $n = 2^k$ random variables in the symmetric case.

3. The dependence between n random variables can also be characterized by a single number

$$\bar{r}(x_1, x_2, \dots, x_n) = \frac{\sum\limits_{i=2}^{n} \bar{r}^2_{(1,2,\dots,i-1);\,i}}{\sum\limits_{i=2}^{n} \left|\bar{r}_{(1,2,\dots,i-1);\,i}\right|}, \tag{10}$$

which we shall call the **maximal summary correlation coefficient** between n random variables.

From the main theorem and formula (10) there follows the

Corollary. *For complete independence of n random variables it is necessary and sufficient that the reduced maximal correlation coefficient (10) vanish.*

4. In the case of normal correlation the eigenfunctions $\varphi_i(Q_1)$, $\psi_i(Q_2)$ in the expansion (6) will be multidimensional Hermite polynomials, and the first eigenfunctions will be polynomials of the first degree; therefore the maximum of the modulus of the functional (3) is attained in this case on the linear functions $\varphi_1(Q_1)$ and $\psi_1(Q_2)$. A. N. Kolmogorov drew our attention to the latter circumstance.

5. Let us consider, as an example, the normal correlation among three variables x, y, z with means 0 and variances 1; it is known that in this case the density of the joint distribution is completely determined by specifying the three ordinary correlation coefficients r_{12}, r_{13}, r_{23}.

From the expansion (6) in this case we find that

$$\bar{r}_{(1,2),3} = \sqrt{\frac{r_{13}^2 + r_{23}^2 - 2r_{12}r_{13}r_{23}}{1 - r_{12}^2}} = r_{(1,2),3}, \tag{11}$$

i.e. it coincides with the ordinary partial correlation coefficient between z and the pair (x, y); the expression for the latter coefficient is given, for example, on p. 392 of the book (^{3}) (formula 114).

Moreover, in this case

$$\varphi_1(x, y) = \frac{\alpha x + \beta y}{\sqrt{\alpha^2 + \beta^2 + 2\alpha\beta r_{12}}}, \qquad \psi_1(z) = z, \tag{12}$$

where

$$\alpha = \frac{r_{13} - r_{12}r_{23}}{1 - r_{12}^2}, \qquad \beta = \frac{r_{23} - r_{12}r_{13}}{1 - r_{12}^2}, \tag{13}$$

with $\bar{r}_{12} = r_{12}$, and

$$\bar{r}(x, y, z) = \frac{r_{12}^2 + r_{(1,2);3}^2}{|r_{12}| + |r_{(1,2);3}|}. \tag{14}$$

In the symmetric case, when $r_{12} = r_{13} = r_{23} = r$, formula (14) takes the form

$$\bar{r}(x, y, z) = |r| \frac{3 + |r|}{1 + |r| + \sqrt{2(1 + |r|)}}; \tag{15}$$

thus, for normal variables the fact that the necessary and sufficient conditions for independence are the conditions $r_{12} = r_{13} = r_{23} = 0$ follows from formulas (11) and (14).

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

Received
6 X 1959

References

1. O. V. Sarmanov, DAN, **120**, No. 4 (1958).

2. O. V. Sarmanov, DAN, **121**, No. 1 (1958).

3. S. N. Bernstein, *Probability Theory*, 4th ed., 1946.

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