

ON SOME PROPERTIES OF MULTIDIMENSIONAL DISTRIBUTIONS

By J. LUKOMSKI

If, in a system of random variables x_1, x_2, \dots, x_n , some variables are connected by a functional (exact) dependence, the n -dimensional distribution law has a degenerated character. In other words, in this case the probability is not distributed over the whole n -dimensional space, but is concentrated on a manifold of a smaller number of dimensions which may be called the *skeleton of the distribution*.

The character and the dimensionality of this manifold are determined by the character and the number of functional connections between the variables x_1, x_2, \dots, x_n . If all these connections are linear, the skeleton will be a linear manifold (hyperplane). The investigation of the skeleton of distribution represents obviously an interest from the theoretical as well as from the practical point of view.

In the present paper we establish some criteria which enable us to determine, for any distribution possessing finite moments of the first and second order, the linear skeleton and to find the variations of the dimensionality of this manifold when the variables are subjected to a linear transformation.¹

We also apply the obtained results to the case of a multidimensional normal distribution (generalized by H. Cramér to the case of linear dependence between variables).

§1

Let

$$(1) \quad x_1, x_2, \dots, x_n$$

be a system of random variables defined in the n -dimensional euclidean space R_n by the multidimensional distribution function $F(x_1, x_2, \dots, x_n)$. The function F is defined on all Borel sets in R_n . We assume the existence of the following moments:

$$E(x_i) = \int \int \dots \int_{R_n} x_i dd \dots dF(x_1, x_2, \dots, x_n) = 0$$
$$E(x_i x_j) = \int \int \dots \int_{R_n} x_i x_j dd \dots dF(x_1, x_2, \dots, x_n) = \mu_{ij}$$

where the integrals are to be understood in the sense of Lebesgue-Radon.

¹The questions of degeneracy of a statistical distribution were for the first time considered—from a somewhat different point of view—by R. Frisch [1].

If the variables x_1, x_2, \dots, x_n are connected by a relation of the form $C_1x_1 + C_2x_2 + \dots + C_nx_n = 0$ ($\Sigma C^2 \neq 0$) (are linearly dependent), we call this relation a *linear bond of the distribution F*.

We shall call a system of linear bond of the distribution *F complete*, if all bonds of the system are linearly independent and every linear bond of the distribution depends linearly on the bonds of the system.

By the (linear) *decrement* of the distribution *F* (we denote it by $k(F)$ or simply k) we understand the number of bonds in a complete system. We may, correspondingly, call the difference between the number of variables and the decrement of the distribution the (linear) *rank* of the distribution, or the dimensionality of the linear skeleton.

The decrement (rank) is given by the following

THEOREM 1.² *The decrement (rank) of the distribution F is equal to the decrement³ (rank) of the matrix*

$$|| \mu_{ij} || \quad i, j = 1, 2, \dots, n$$

of the moments of the second order of this distribution; that is

$$(2) \quad k(F) = k(|| \mu_{ij} ||), \quad i, j = 1, 2, \dots, n$$

PROOF. Consider the form

$$(3) \quad v = t_1x_1 + t_2x_2 + \dots + t_nx_n$$

where t_1, t_2, \dots, t_n are arbitrary real numbers, not all equal to zero. Let

$$(4) \quad \begin{aligned} Q^2 = E(v^2) &= \int \int \dots \int_{R_n} (t_1x_1 + t_2x_2 + \dots + t_nx_n)^2 dd \dots \\ &\dots dF(x_1, x_2, \dots, x_n) \\ &= \sum_{i,j=1}^n t_i t_j \int \int \dots \int_{R_n} x_i x_j dd \dots dF(x_1, x_2, \dots, x_n) = \sum_{i,j=1}^n t_i t_j \mu_{ij}. \end{aligned}$$

Q^2 is a non-negative quadratic form in the variables t_1, t_2, \dots, t_n . The system of values t_1, t_2, \dots, t_n , for which the expression (3) becomes zero is a double point of the form Q^2 .

The coordinates of the double point can be found from the system of homogeneous equations:

$$(5) \quad \begin{aligned} \mu_{11}t_1 + \mu_{12}t_2 + \dots + \mu_{1n}t_n &= 0 \\ \mu_{21}t_1 + \mu_{22}t_2 + \dots + \mu_{2n}t_n &= 0 \\ \dots \dots \dots \dots \dots \dots \dots & \\ \mu_{n1}t_1 + \mu_{n2}t_2 + \dots + \mu_{nn}t_n &= 0. \end{aligned}$$

² This theorem was proved by a different method by R. Frisch [1].

³ By the decrement of a (rectangular) matrix we call, after B. Kagan, the difference between the number of its rows and its rank.

It is, however, known that the number of the independent double points of the form, Q^2 , i.e. the number of linearly independent untrivial solutions of the system (5) is equal to the decrement of the matrix $|| \mu_{ij} ||, i, j = 1, 2, \dots n$.

Consequently, there exist only $k(|| \mu_{ij} ||)$ independent linear connections between the variables x_1, x_2, \dots, x_n , which proves the theorem.

Hence it follows that the variables x_1, x_2, \dots, x_n are linearly independent ($k(F) = 0$) if and only if the form Q^2 is positive definite and, consequently, the discriminant $| \mu_{ij} |$ of the form is positive.

The following two theorems may be used for determination of a complete system of linear bonds. The first of them is a special case of the second, but is stated separately in order to simplify the proof.

THEOREM 2. *If $k(F) = 1$, we obtain the linear bond of the distribution by replacing in the determinant on the left hand side of the equation*

$$(6) \quad \begin{vmatrix} \mu_{11} & \mu_{12} & \dots & \mu_{1n} \\ \mu_{21} & \mu_{22} & \dots & \mu_{2n} \\ \dots & \dots & \dots & \dots \\ \mu_{n1} & \mu_{n2} & \dots & \mu_{nn} \end{vmatrix} = 0$$

the elements of one (arbitrary) row by x_1, x_2, \dots, x_n respectively.

For instance, replacing the first row, we have

$$(7) \quad \begin{vmatrix} x_1 & x_2 & \dots & x_n \\ \mu_{21} & \mu_{22} & \dots & \mu_{2n} \\ \dots & \dots & \dots & \dots \\ \mu_{n1} & \mu_{n2} & \dots & \mu_{nn} \end{vmatrix} = 0.$$

PROOF. Since the decrement of the matrix $|| \mu_{ij} ||, i, j = 1, 2, \dots n$ is equal to 1, for the unique nontrivial independent solution of the system (5) (t_1, t_2, \dots, t_n) may be taken, as we know, the system of algebraical supplements of the elements of any row of the determinant $| \mu_{ij} |, i, j = 1, 2, \dots n$. (Among the algebraical supplements of elements of each row there is at least one different from zero, since the algebraical supplements of corresponding elements of any pair of rows are proportional to each other.)

Hence, since $t_1x_1 + t_2x_2 + \dots + t_nx_n = 0$, the theorem follows.

THEOREM 3. *If $k(F) > 0$, we obtain a complete system of linear bonds of the distribution F replacing in each of the k equations*

$$(8) \quad \begin{vmatrix} \mu_{ki} & \mu_{k,k+1} & \dots & \mu_{kn} \\ \mu_{k+1,i} & \mu_{k+1,k+1} & \dots & \mu_{k+1,n} \\ \dots & \dots & \dots & \dots \\ \mu_{ni} & \mu_{n,k+1} & \dots & \mu_{nn} \end{vmatrix} = 0, \quad i = 1, 2, \dots k$$

one (arbitrary) row of the determinant respectively by x_i, x_{k+1}, \dots, x_n , where x_{k+1}, \dots, x_n are chosen in such a way that

$$\begin{vmatrix} \mu_{k+1,k+1} & \dots & \mu_{k+1,n} \\ \dots & \dots & \dots \\ \mu_{n,k+1} & \dots & \mu_{nn} \end{vmatrix} > 0.$$

PROOF. Consider a system of forms in arbitrary linearly independent parameters $\xi_1, \xi_2, \dots, \xi_n$:

$$\begin{aligned}
 v_1 &= a_{11}\xi_1 + a_{12}\xi_2 + \dots + a_{1n}\xi_n \\
 v_2 &= a_{21}\xi_1 + a_{22}\xi_2 + \dots + a_{2n}\xi_n \\
 &\dots\dots\dots \\
 v_m &= a_{m1}\xi_1 + a_{m2}\xi_2 + \dots + a_{mn}\xi_n \\
 &\dots\dots\dots \\
 v_{m+1} &= a_{m+1,1}\xi_1 + a_{m+1,2}\xi_2 + \dots + a_{m+1,n}\xi_n \\
 &\dots\dots\dots \\
 v_{m+k} &= a_{m+k,1}\xi_1 + a_{m+k,2}\xi_2 + \dots + a_{m+k,n}\xi_n
 \end{aligned}
 \tag{14}$$

such that the matrix of the system (14) coincides with the elongated matrix of the transformation.

For

$$v_{m+1} = 0, \quad v_{m+2} = 0, \quad \dots, \quad v_{m+k} = 0
 \tag{15}$$

the system (14) reduces to the system (12).

If the decrement of the matrix of the system is equal to s , there exist only $m + k - s$ linearly independent forms v_i , and each of the remaining s forms is a linear combination of the first.

By Steinitz's theorem we can always include in a subsystem of independent forms the forms v_{m+1}, \dots, v_{m+k} (since these forms are independent).

Denoting all forms of the subsystem by $v_{s+1}, \dots, v_m, v_{m+1}, \dots, v_{m+k}$, let us write the s relations connecting each of the remaining forms with the forms of our subsystem in the form:

$$\begin{aligned}
 g_{11}v_1 + g_{1,s+1}v_{s+1} + \dots + g_{1m}v_m + g_{1,m+1}v_{m+1} + \dots + g_{1,m+k}v_{m+k} &= 0 \\
 g_{22}v_2 + g_{2,s+1}v_{s+1} + \dots + g_{2m}v_m + g_{2,m+1}v_{m+1} + \dots + g_{2,m+k}v_{m+k} &= 0 \\
 \dots\dots\dots \\
 g_{ss}v_s + g_{s,s+1}v_{s+1} + \dots + g_{sm}v_m + g_{s,m+1}v_{m+1} + \dots + g_{s,m+k}v_{m+k} &= 0
 \end{aligned}
 \tag{16}$$

where $g_{11}, g_{22}, \dots, g_{ss} \neq 0$.

Assigning to the variables in these equations the values (15) we clearly obtain s linear relations between the variables u_1, u_2, \dots, u_m

$$\begin{aligned}
 g_{11}u_1 + g_{1,s+1}u_{s+1} + \dots + g_{1m}u_m &= 0 \\
 g_{22}u_2 + g_{2,s+1}u_{s+1} + \dots + g_{2m}u_m &= 0 \\
 \dots\dots\dots \\
 g_{ss}u_s + g_{s,s+1}u_{s+1} + \dots + g_{sm}u_m &= 0.
 \end{aligned}
 \tag{17}$$

The equations (17) are linearly independent, since the matrix of the system (17)

The decrement and the linear bonds of the distribution may be determined from the matrix of the coefficients $\| | c_{rs} \|$ $r, s = 1, 2, \dots, n$ on ground of the general theorems of §1.

Let, as before,

$$(11) \quad \begin{aligned} u_1 &= a_{11}x_1 + a_{12}x_2 + \dots + a_{1n}x_n \\ u_2 &= a_{21}x_1 + a_{22}x_2 + \dots + a_{2n}x_n \\ &\dots\dots\dots\dots\dots\dots\dots\dots\dots\dots\dots\dots \\ u_m &= a_{m1}x_1 + a_{m2}x_2 + \dots + a_{mn}x_n \end{aligned}$$

be a system of linear forms in the variables x_1, x_2, \dots, x_n . We shall prove the following

THEOREM 5. *The variables u_1, u_2, \dots, u_m are subject to the generalized normal distribution law the decrement of which is equal to the decrement of the elongated matrix of the transformation*

$$\left\| \begin{array}{cccc} a_{11} & a_{12} & \dots & a_{1n} \\ \dots\dots\dots\dots\dots\dots\dots\dots\dots\dots\dots\dots \\ a_{m1} & a_{m2} & \dots & a_{mn} \\ a_{m+1,1} & a_{m+1,2} & \dots & a_{m+1,n} \\ \dots\dots\dots\dots\dots\dots\dots\dots\dots\dots\dots\dots \\ a_{m+k,1} & a_{m+k,2} & \dots & a_{m+k,n} \end{array} \right\|.$$

PROOF. Consider the characteristic function of the distribution $G_1(u_1, u_2, \dots, u_m)$,

$$(23) \quad f_1(t_1, t_2, \dots, t_m) = \int \int \dots \int_{R_m} e^{i(t_1u_1+t_2u_2+\dots+t_mu_m)} d d \dots dG_1(u_1, u_2, \dots, u_m).$$

Performing in this expression the substitution (11), we obtain

$$\begin{aligned} &f_1(t_1, t_2, \dots, t_m) \\ &= \int \int \dots \int_{R_n} e^{i\left(t_1 \sum_{j=1}^n a_{1j}x_j + t_2 \sum_{j=1}^n a_{2j}x_j + \dots + t_m \sum_{j=1}^n a_{mj}x_j\right)} d d \dots \\ (24) \quad &\dots dG_1\left\{ \sum_{j=1}^n a_{1j}x_j, \sum_{j=1}^n a_{2j}x_j, \dots, \sum_{j=1}^n a_{mj}x_j \right\} \\ &= \int \int \dots \int_{R_n} e^{i\left(x_1 \sum_{p=1}^m a_{p1}t_p + x_2 \sum_{p=1}^m a_{p2}t_p + \dots + x_n \sum_{p=1}^m a_{pn}t_p\right)} d d \\ &\dots dG(x_1, x_2, \dots, x_n). \end{aligned}$$

($d d \dots dG(x_1, x_2, \dots, x_n)$ in the expression (24) does not, in general, coincide with $d d \dots dG(x_1, x_2, \dots, x_n)$ in the expression (22)).

Taking into account (22), we obtain

$$f_1 = e^{-iQ}$$

where

$$\begin{aligned} Q_1^2 &= \sum_{r,s=1}^n \left\{ C_{rs} \left(\sum_{p=1}^m a_{pr} t_p \right) \left(\sum_{q=1}^m a_{qs} t_q \right) \right\} \\ (26) \quad &= \sum_{r,s=1}^n \left\{ C_{rs} \sum_{p,q=1}^m a_{pr} a_{qs} t_p t_q \right\} \\ &= \sum_{p,q=1}^m \left\{ t_p t_q \sum_{r,s=1}^n a_{pr} a_{qs} C_{rs} \right\} = \sum_{p,q=1}^m t_p t_q v_{pq}. \end{aligned}$$

Q_1^2 is a non-negative quadratic form in t_1, t_2, \dots, t_m , the coefficients of which coincide with the moments of the second order of the distribution $G_1(u_1, u_2, \dots, u_m)$.

Consequently, the distribution G_1 is a generalized normal distribution.

By Theorem 4 the decrement of the distribution G_1 is equal to the decrement of the matrix $\| a_{pr} \|$ $p = 1, 2, \dots, m + k$; $r = 1, 2, \dots, n$, the last k rows of which consist of the coefficients of the complete system of linear bonds of the distribution G .

Let now x_1, x_2, \dots, x_n be a system of random variables subjected to a proper Gaussian law. The density function of the distribution of the system is

$$y = Ce^{-x^2} = \frac{1}{(2\pi)^{\frac{n}{2}} \prod_{i=1}^n \sqrt{\mu_{ii}} \sqrt{R}} \exp \left[-\frac{1}{2} R \left\{ \sum_{i=1}^n R_{ii} \frac{x_i^2}{\mu_{ii}} + 2 \sum_i \sum_{i>j} R_{ij} \frac{x_i x_j}{\sqrt{\mu_{ii} \mu_{jj}}} \right\} \right]$$

where

$$R = \begin{vmatrix} 1 & r_{12} & \dots & r_{1n} \\ r_{21} & 1 & \dots & r_{2n} \\ \dots & \dots & \dots & \dots \\ r_{n1} & r_{n2} & \dots & 1 \end{vmatrix}.$$

R_{ij} are the algebraical supplements in R , $r_{ij} = \frac{\mu_{ij}}{\sqrt{\mu_{ii} \mu_{jj}}}$, and x^2 is a positive definite quadratic form in the variables x_1, x_2, \dots, x_n .

Again let u_1, u_2, \dots, u_m be a system of linear forms in the variables x_1, x_2, \dots, x_n

$$\begin{aligned} u_1 &= a_{11}x_1 + a_{12}x_2 + \dots + a_{1n}x_n \\ u_2 &= a_{21}x_1 + a_{22}x_2 + \dots + a_{2n}x_n \\ &\dots \dots \dots \dots \dots \dots \dots \dots \dots \dots \dots \dots \dots \dots \dots \dots \dots \\ u_m &= a_{m1}x_1 + a_{m2}x_2 + \dots + a_{mn}x_n. \end{aligned} \tag{11}$$

Then from Theorem 5 follows the

COROLLARY. *The random variables u_1, u_2, \dots, u_m are subject to the m -dimensional properly normal distribution law of Gauss if and only if the matrix*

$$\begin{vmatrix} a_{11} & a_{12} & \cdots & a_{1n} \\ a_{21} & a_{22} & \cdots & a_{2n} \\ \dots & \dots & \dots & \dots \\ a_{m1} & a_{m2} & \cdots & a_{mn} \end{vmatrix}$$

of the system of forms (11) has the rank m .

REFERENCES

- [1] R. Frisch, "Correlation and Scatter in Statistical Variables," *Nordisk Statistisk Tidskrift*, vol. 1 (1929), 36-102.
- [2] H. Cramér, "Über eine Eigenschaft der normalen Verteilungsfunktion," *Math. Zeitschr.*, vol. 41 (1936),
- [3] H. Cramér, *Random Variables and Probability Distributions*, Cambridge 1937.

Moscow, U. S. S. R.