PLISKA
STUDIA MATHEMATICA BULGARICA

ПЛИСКА
БЪЛГАРСКИ МАТЕМАТИЧЕСКИ СТУДИИ

Pliska Stud. Math. Bulgar. **18** (2007), 379–386

PLISKA
STUDIA MATHEMATICA
BULGARICA

JOINT DENSITIES OF CORRELATION COEFFICIENTS FOR SAMPLES FROM MULTIVARIATE STANDARD NORMAL DISTRIBUTION

Evelina Veleva

We consider the joint distribution of the correlation coefficients for samples from multivariate standard normal distribution. Some marginal densities are obtained. Independence and conditional independence between sets of sample correlation coefficients are established.

1. Introduction

Let $\xi = (\xi_1, \dots, \xi_n)^T$ be a random vector with standard normal distribution $N_n(\vec{0}, \mathbf{I})$, where $\vec{0}$ is a zero $n \times 1$ vector, and $\mathbf{I}$ is the identity matrix of size n. Let $\xi^{(1)}, \dots, \xi^{(m)}$ be a sample from ξ with size m, $(m > n)$. Consider the matrix $\mathbf{R}$,

$$\mathbf{R} = \begin{pmatrix} 1 & \eta_{12} & \dots & \eta_{1n} \\ \eta_{12} & 1 & \dots & \eta_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ \eta_{1n} & \eta_{2n} & \dots & 1 \end{pmatrix},$$

where η_{ij} is the sample correlation coefficient of the random variables ξ_i and ξ_j, $1 \le i < j \le n$. The joint density of the elements η_{ij}, $1 \le i < j \le n$ of the matrix $\mathbf{R}$ is of the form (see [4] and [7]):

$$C_n \left(\det \mathbf{Y}_n\right)^{\frac{m-n-1}{2}} I_D \;, \tag{1}$$

2000 *Mathematics Subject Classification:* 62H10

Key words: Multivariate normal distribution, sample correlation coefficients, independence, conditional independence.

where:

- C_n is a suitable constant,

$$C_n = \frac{\left[\Gamma\left(\frac{m}{2}\right)\right]^{n-1}}{\Gamma\left(\frac{m-1}{2}\right)\Gamma\left(\frac{m-2}{2}\right)\dots\Gamma\left(\frac{m-n+1}{2}\right)\left[\Gamma\left(\frac{1}{2}\right)\right]^{\frac{n(n-1)}{2}}}, \tag{2}$$

  $\Gamma(\cdot)$ is the well known Gamma function;

- $\mathbf{Y}_n$ is a real symmetric matrix,

$$\mathbf{Y}_n = \begin{pmatrix} 1 & y_{12} & \dots & y_{1n} \\ y_{12} & 1 & \dots & y_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ y_{1n} & y_{2n} & \dots & 1 \end{pmatrix}; \tag{3}$$

- I_D is the indicator of the set D, consisting of all points $(y_{ij}, 1 \le i < j \le n)$ in the Euclidian real space $R_{n(n-1)/2}$ with dimension $n(n-1)/2$, for which the matrix $\mathbf{Y}_n$ is positive definite.

The multivariate normal (Gaussian) distribution is a basic distribution in many models of the multivariate statistical analysis (see [1, 2, 3, 5, 6]). The sample correlation matrix $\mathbf{R}$ arises naturally in the decision theory when we consider hypothesis, connected with the correlation between the factors in the experiment.

The author in [7] proves necessary and sufficient conditions for the positive definiteness of the matrix $\mathbf{Y}_n$. These conditions allow us to obtain some of the marginal densities of the density in (1) (see [7]).

2. Joint density of union of subsets

Let us denote by V_n the set $V_n = \{\eta_{ij}, 1 \le i < j \le n\}$. Let $S = \{\eta_{i_s j_s},\ s = 1, \dots, k\}$ be an arbitrary subset of the set of random variables V_n. To this subset S we can attach a graph $G(S)$ with nodes $\{1, 2, \dots, n\}$ and k undirected edges $\{\{i_s, j_s\}, s = 1, \dots k\}$. It is easy to find that the correspondence

$$S \to G(S)$$

is bilateral, i.e. there exists the inverse correspondence

$$G \to S(G).$$

Introduce the notation f_S for the joint density of the random variables belonging to the set S. If S is the empty set $\varnothing$ we define $f_S = 1$.

Theorem 1. *Let S_1 and S_2 be two subsets of the set V_n, and $G_1 = G(S_1)$, $G_2 = G(S_2)$ be their corresponding graphs. Denote by K_i the set of the numbers of the nodes, which are ends of edges of the graph G_i, $i = 1, 2$. Let $K = K_1 \cap K_2$. If the set K contains at least two elements and for every two elements p and q of K, $p < q$ the random variable η_{pq} belongs simultaneously to S_1 and S_2 then*

$$f_{S_1 \cup S_2} = \frac{f_{S_1} f_{S_2}}{f_{S_1 \cap S_2}}, \tag{4}$$

i.e. the sets S_1 and S_2 are conditionally independent on the set $S_1 \cap S_2$.

Proof. Let us denote the number of the elements of the sets K_1, K_2 and K by k_1, k_2 and k respectively. Let $S_\cup$ and $S_\cap$ be the sets $S_\cup = S_1 \cup S_2$ and $S_\cap = S_1 \cap S_2$, and $G_\cup$, $G_\cap$ be their corresponding graphs, $G_\cup = G(S_\cup)$, $G_\cap = G(S_\cap)$. Re-number the vertices in the graphs G_1, G_2, $G_\cup$ and $G_\cap$ so that the nodes from the set $K_1 \setminus K$ to have the numbers from 1 to $k_1 - k$; the nodes from the set K to have the numbers from $k_1 - k + 1$ to k_1 and the nodes from the set $K_2 \setminus K$ to have the numbers from $k_1 + 1$ to $k_1 + k_2 - k$. With this re-numbering we get new graphs $G_1^\star$, $G_2^\star$, $G_\cup^\star$, $G_\cap^\star$ and corresponding new subsets $S_1^\star$, $S_2^\star$, $S_\cup^\star$ and $S_\cap^\star$ of V_n, $S_1^\star = S(G_1^\star)$, $S_2^\star = S(G_2^\star)$ $S_\cup^\star = S(G_\cup^\star)$ and $S_\cap^\star = S(G_\cap^\star)$. The next proposition can be found in [9].

Proposition 1. *Let $G(S)$ be the corresponding graph to a subset $S \subset V_n$. Let permute the numbers of the vertices of the graph $G(S)$ and let denote the new graph by $G^\star$. The subset $S^\star = S(G^\star)$ has the same joint distribution as the initial set S.*

According to this statement, the random variables from the set $S_1^\star$ have the same joint distribution as the random variables from the set S_1, i.e.

$$f_{S_1^\star} = f_{S_1} \ . \tag{5}$$

Analogically,

$$f_{S_2^\star} = f_{S_2} \quad , \quad f_{S_\cup^\star} = f_{S_\cup} \quad , \quad f_{S_\cap^\star} = f_{S_\cap} \ . \tag{6}$$

It is easy to see that

$$S_\cup^\star = S_1^\star \cup S_2^\star \quad , \quad S_\cap^\star = S_1^\star \cap S_2^\star \ . \tag{7}$$

Let us denote by U_1 and U_2 the sets $U_1 = \{\eta_{ij} \mid 1 \le i < j \le k_1\}$ and $U_2 = \{\eta_{ij} \mid k_1 - k + 1 \le i < j \le k_1 + k_2 - k\}$. It is obvious that $S_1^\star \subset U_1$, $S_2^\star \subset U_2$ and $U_1, U_2 \subset V_{k_1 + k_2 - k}$. The next proposition can be found in [7].

Proposition 2. *The joint density of the random variables from the set V_s, where s is an integer ($s < n$), has the form (1) with $n = s$.*

Consequently, the joint density of the random variables from the set $V_{k_1+k_2-k}$ has the form (1). We will use the next proposition which can be found in [8].

Proposition 3. *Let p and q be arbitrary integers, such that $2 \le p \le n-1$ and $1 \le q \le n-2$. Let A and B be subsets of the set V_n, $A = \{\eta_{ij} \mid 1 \le i < j \le p\}$ and $B = \{\eta_{ij} \mid q+1 \le i < j \le n\}$. Then*

$$f_{A\cup B} = \frac{f_A f_B}{f_{A\cap B}} \ .$$

From the last equality it follows that:

$$f_{U_1\cup U_2} = \frac{f_{U_1} f_{U_2}}{f_{U_1\cap U_2}} \ . \tag{8}$$

It can be easily seen that

$$S_1^\star \cap S_2^\star = \{\eta_{ij} \mid k_1 - k + 1 \le i < j \le k_1\} \ .$$

Consequently,

$$S_1^\star \cap S_2^\star \equiv U_1 \cap U_2 \ .$$

Let us integrate the two sides of the representation (8) with respect to the variables, corresponding to the random variables from the set $(U_1 \setminus S_1^\star) \cup (U_2 \setminus S_2^\star)$. On the left we get the density $f_{S_1^\star \cup S_2^\star}$. On the right, the variables, corresponding to the random variables from the set $U_1 \setminus S_1^\star$ appear only in the density f_{U_1}, and those related to the random variables from the set $U_2 \setminus S_2^\star$ appear only in the density f_{U_2}. Therefore we get that

$$f_{S_1^\star \cup S_2^\star} = \frac{f_{S_1^\star} f_{S_2^\star}}{f_{U_1\cap U_2}} = \frac{f_{S_1^\star} f_{S_2^\star}}{f_{S_1^\star \cap S_2^\star}} \ ,$$

whence by the equalities (5) – (7) the representation (4) follows. □

Theorem 2. *Let S_1 and S_2 be two subsets of the set V_n, and $G_1 = G(S_1)$, $G_2 = G(S_2)$ be their corresponding graphs. Let us denote by K_i the set of the numbers of the nodes, which are ends of edges of the graph G_i, $i = 1, 2$. Let $K = K_1 \cap K_2$. If the set K contains at most one element then*

$$f_{S_1\cup S_2} = f_{S_1} f_{S_2} \ ,$$

i.e. the sets S_1 and S_2 are independent.

P r o o f. Let us denote the number of the elements of the sets K_1, K_2 and K by k_1, k_2 and k respectively. Let $S_\cup$ be the union $S_\cup = S_1 \cup S_2$, and $G_\cup$ be its corresponding graph, $G_\cup = G(S_\cup)$. We consider two cases:

Case I. Let $k = 0$. Let us re-number the vertices in the graphs G_1, G_2 and $G_\cup$ so that the nodes with numbers from the set K_1 can take values from 1 to k_1; the nodes from the set K_2 can take values from $k_1 + 1$ to $k_1 + k_2$. With this re-numbering we get new graphs $G_1^\star$, $G_2^\star$, $G_\cup^\star$ and corresponding new subsets $S_1^\star$, $S_2^\star$ and $S_\cup^\star$ of V_n, $S_1^\star = S(G_1^\star)$, $S_2^\star = S(G_2^\star)$ and $S_\cup^\star = S(G_\cup^\star)$. According to Proposition 1, the random variables from the set $S_1^\star$ have the same joint distribution as the random variables from the set S_1, i.e. $f_{S_1^\star} = f_{S_1}$. Analogically, $f_{S_2^\star} = f_{S_2}$ and $f_{S_\cup^\star} = f_{S_\cup}$. It is easy to see that $S_\cup^\star = S_1^\star \cup S_2^\star$.

Let us denote by U_1 and U_2 the sets $U_1 = \{\eta_{ij} \mid 1 \le i < j \le k_1\}$ and $U_2 = \{\eta_{ij} \mid k_1 + 1 \le i < j \le k_1 + k_2\}$. According to Proposition 2, the joint density of the random variables from the set $V_{k_1+k_2}$ is of the form (1). In accordance with Proposition 3, for U_1 and U_2 the representation (8) holds. Since in this case $U_1 \cap U_2 = \varnothing$ then $f_{U_1 \cap U_2} = 1$ and hence

$$f_{U_1 \cup U_2} = f_{U_1} f_{U_2} \ . \tag{9}$$

It is obvious that $S_1^\star \subset U_1$ and $S_2^\star \subset U_2$. Let us integrate the two sides of the representation (9) with respect to the variables, corresponding to the random variables from the set $(U_1 \setminus S_1^\star) \cup (U_2 \setminus S_2^\star)$. On the left we get the density $f_{S_1^\star \cup S_2^\star}$. On the right, the variables, corresponding to the random variables from the set $U_1 \setminus S_1^\star$ appear only in the density f_{U_1}, and those related to the random variables from the set $U_2 \setminus S_2^\star$ appear only in the density f_{U_2}. Therefore we get the equality

$$f_{S_1^\star \cup S_2^\star} = f_{S_1^\star} f_{S_2^\star} \ ,$$

consequently it follows that

$$f_{S_1 \cup S_2} = f_{S_1} f_{S_2} \ ,$$

i.e. the two subsets of random variables S_1 and S_2 are independent.

Case II. Let $k = 1$. The proof is by analogy with Case I, but here we re-number the vertices in the graphs G_1, G_2 and $G_\cup$ so that the nodes with numbers from the set K_1 can take values from 1 to k_1; the nodes from the set K_2 can take values from k_1 to $k_1 + k_2 - 1$. After that we consider the sets U_1 and U_2, $U_1 = \{\eta_{ij} \mid 1 \le i < j \le k_1\}$, $U_2 = \{\eta_{ij} \mid k_1 \le i < j \le k_1 + k_2 - 1\}$. For U_1 and U_2, according to Propositions 2 and 3, the representation (9) holds. The rest of the proof is analogically to the Case I. Thus the proof is complete. □

Theorem 3. *Let r be an integer, $r \geq 2$, and $S_1, \ldots, S_r$ be a sequence of subsets of the set V_n, such that*

$$S_t = \{\eta_{ij} \mid i, j \in M_t, i < j\} ,$$

where M_t, $t = 1, \ldots, r$ are subsets of the set $\{1, \ldots, n\}$. Let for all t, $t = 2, \ldots, r$ the intersection

$$(M_1 \cup \ldots \cup M_{t-1}) \cap M_t , \tag{10}$$

satisfies one of the next two conditions:

1. *The set in (10) contains at most one element;*
2. *The set in (10) contains at least two elements. In this case for every two elements p and q from (10), the random variable η_{pq}, $(p < q)$, belongs to the set*
$$(S_1 \cup \ldots \cup S_{t-1}) \cap S_t .$$

Then

$$f_{S_1 \cup S_2 \cup \ldots \cup S_r} = \frac{f_{S_1} f_{S_2} \ldots f_{S_r}}{f_{S_1 \cap S_2} f_{(S_1 \cup S_2) \cap S_3} \cdots f_{(S_1 \cup \ldots \cup S_{r-1}) \cap S_r}} . \tag{11}$$

P r o o f. The proof is by induction on r. Let $r = 2$, and S_1, S_2 be subsets of the set V_n of the form

$$S_1 = \{\eta_{ij} \mid i, j \in M_1, i < j\} \quad , \quad S_2 = \{\eta_{ij} \mid i, j \in M_2, i < j\} ,$$

where M_1 and M_2 are subsets of the set $\{1, \ldots, n\}$. It is easy to see that the intersection $S_1 \cap S_2$ will have the form

$$S_1 \cap S_2 = \{\eta_{ij} \mid i, j \in M_1 \cap M_2, i < j\} . \tag{12}$$

Let us denote by G_1 and G_2 the corresponding graphs to the subsets S_1 and S_2, i.e. $G_1 = G(S_1)$ and $G_2 = G(S_2)$. Let K_i be the set of the numbers of the nodes, which are ends of edges of the graph G_i, $i = 1, 2$. It is easy to see that $K_1 = M_1$ and $K_2 = M_2$. Denote by K the set

$$K = K_1 \cap K_2 = M_1 \cap M_2 .$$

Let the set K contain at most one element. According to Theorem 2, we have

$$f_{S_1 \cup S_2} = f_{S_1} f_{S_2} .$$

From (12) it follows that the set $S_1 \cap S_2$ is empty, whence $f_{S_1 \cap S_2} = 1$. Consequently, we get the representation

$$f_{S_1 \cup S_2} = \frac{f_{S_1} f_{S_2}}{f_{S_1 \cap S_2}} \ . \tag{13}$$

Let the set K contains at least two elements and for every two elements p and q of K, the random variable η_{pq}, $(p < q)$, belongs to the intersection $S_1 \cap S_2$. Then, according to Theorem 1 the equality (13) holds. Therefore, this Theorem 3 is true for $r = 2$.

Suppose that the Theorem 3 is true for some r, $r \geq 2$. Let $S_1, \dots, S_{r+1}$ be a sequence of subsets of the set V_n, such that

$$S_t = \{\eta_{ij} \mid i, j \in M_t, i < j\} \ ,$$

where M_t, $t = 1, \dots, r+1$ are subsets of the set $\{1, \dots, n\}$. Let for all t, $t = 2, \dots, r+1$ the intersection (10) satisfies one of the conditions 1 and 2. Let us denote by A and B the sets $A = S_1 \cup \ldots \cup S_r$, $B = S_{r+1}$. Let G_1 and G_2 be the corresponding graphs of the subsets A and B, i.e. $G_1 = G(A)$ and $G_2 = G(B)$. Denote by K_i the set of the numbers of the nodes, which are ends of edges of the graph G_i, $i = 1, 2$. It is easy to see that $K_1 = M_1 \cup \ldots \cup M_r$ and $K_2 = M_{r+1}$. Let K be the intersection $K = K_1 \cap K_2 = (M_1 \cup \ldots \cup M_r) \cap M_{r+1}$.

Suppose that the set K contains at most one element. According to Theorem 2, we have

$$f_{A \cup B} = f_A f_B \ .$$

The set $A \cap B$ is empty, therefore $f_{A \cap B} = 1$. Consequently, we get the representation

$$f_{A \cup B} = \frac{f_A f_B}{f_{A \cap B}} \ . \tag{14}$$

Let the set K contains at least two elements and for every two elements p and q of K, the random variable η_{pq}, $(p < q)$, belongs to the set

$$A \cap B = (S_1 \cup \ldots \cup S_r) \cap S_{r+1} \ .$$

Then, according to Theorem 1 the equality (14) holds. From the induction assumption we have the representation (11) for the density f_A. Hence, by the equality (14) the Theorem 3 follows for $r + 1$. Thus the proof of this Theorem 3 is complete. □

REFERENCES

[1] Anderson, T. W. An Introduction to Multivariate Statistical Analysis. New York, John Wiley & Sons, 1958.

[2] Dempster, A. P. Elements of Continuous Multivariate Statistical Analysis. Mass., Academic Press, Reading, 1969.

[3] Eaton, M. L. Multivariate Statistics. New York, John Wiley & Sons, 1983.

[4] Ignatov, T. G., A. D. Nikolova About Wishart's Distribution. *Annuaire de l'Univerite de Sofia "St.Kliment Ohridski", Faculte des Sciences Economiques et de Gestion* **3** (2004), 79–94.

[5] Muirhead, R. J. Aspects of Multivariate Statistical Theory. New York, John Wiley & Sons, 1982.

[6] Rao, C. R. Linear Statistical Inference and Its Application. New York, John Wiley & Sons, 1973.

[7] Veleva E. I. Some marginal densities of the sample correlation coefficients of independent normally distributed random variables (in Bulgarian). *Annuaire de l'Universite de Sofia "St.Kliment Ohridski", Faculte des Sciences Economiques et de Gestion* **5** (2006), 353–369.

[8] Veleva E. I. Conditional independence of joint sample correlation coefficients of independent normally distributed random variables. *Proceedings of 35-th Spring conference of the Union of Bulgarian Mathematicians* 5 - 8 April, 2006, Borovec, 315–320.

[9] Veleva E. I., T. G. Ignatov Distributions of joint sample correlation coefficients of independent normally distributed random variables. *Advanced Studies in Contemporary Mathematics* **12(2)** (2006), 247–254.

Evelina Veleva
Department of Numerical Methods and Statistics,
University of Rousse "Angel Kanchev",
Studentska str. 8,
Rouse, 7017, Bulgaria,
e-mail: `eveleva@abv.bg`