МАТЕМАТИКА И МАТЕМАТИЧЕСКО ОБРАЗОВАНИЕ, 2006
MATHEMATICS AND EDUCATION IN MATHEMATICS, 2006
Proceedings of the Thirty Fifth Spring Conference of
the Union of Bulgarian Mathematicians
Borovets, April 5–8, 2006

CONDITIONAL INDEPENDENCE OF JOINT SAMPLE CORRELATION COEFFICIENTS OF INDEPENDENT NORMALLY DISTRIBUTED RANDOM VARIABLES*

Evelina I. Veleva

The aim of this paper is to obtain a formula for the densities of a class of joint sample correlation coefficients of independent normally distributed random variables. The established relation between the joint densities of certain sets of sample correlation coefficients shows their conditional independence.

Let $\boldsymbol{\xi} = (\xi_1, \dots \xi_p)'$ be a random vector with distribution $N_p(\vec{\mathbf{0}}, \mathbf{I})$, where $\vec{\mathbf{0}}$ is a zero $p \times 1$ vector, and $\mathbf{I}$ is the identity matrix of size p. Let $\boldsymbol{\xi}^{(1)}, \dots, \boldsymbol{\xi}^{(n)}$ be a sample from $\boldsymbol{\xi}$ of size n. Consider the sample correlation coefficient ν_{ij} of the random variables ξ_i and ξ_j, $1 \le i < j \le p$. The joint density of $\nu_{ij},\ 1 \le i < j \le p$ is of the form (see [2] and [3])

$$C_p \begin{vmatrix} 1 & x_{12} & \dots & x_{1p} \\ x_{12} & 1 & \dots & x_{2p} \\ \dots & \dots & \dots & \dots \\ x_{1p} & x_{2p} & \dots & 1 \end{vmatrix}^{\frac{(n-p-1)}{2}}, \tag{1}$$

for all points $\{x_{ij}, \quad 1 \le i < j \le p\}$ in $R_{p(p-1)/2}$ for which the symmetric matrix in (1) is positively definite, where

— C_p is a constant, given by

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

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

and $|\cdot|$ is a notation for the determinant of a matrix.

Denote by $\Delta^{+}_{k_1,\dots,k_s}$ the matrix, which can be obtained from the matrix in (1) after deleting the rows and columns with numbers $k_1, \dots, k_s$.

Let us introduce the notation f_M for the joint density of the random variables from a set M; if $M = \Phi$, i.e. the set M is empty we define $f_M = 1$.

*2000 Mathematics Subject Classification: 62H10

Lemma 1. *Let K be a set of natural numbers $K = \{k_1, \dots, k_s\}$, each one belonging to the interval $[1,p]$ and M be the set $M = \{\nu_{ij} \mid i \notin K, \quad j \notin K, \quad 1 \le i < j \le p\}$. Then*

$$f_M = C_{p-s} |\Delta^+_{k_1,\dots,k_s}|^{\frac{n-p+s-1}{2}}, \tag{3}$$

for all points in $R_{(p-s)(p-s-1)/2}$ for which the matrix $\Delta^+_{k_1,\dots,k_s}$ is positively definite.

Proof. We will use the following Theorem proved in [3]:

Theorem 1. *Let k be a fixed integer, $1 \le k \le p$. The joint density of the random variables $\{\nu_{ij} | i \ne k, \quad j \ne k, \quad 1 \le i < j \le p\}$ is equal to*

$$C_{p-1} |\Delta^+_k|^{\frac{n-p}{2}},$$

for all points in $R_{(p-1)(p-2)/2}$ for which the matrix Δ^+_k is positively definite.

If we substitute in this statement $k = k_1$ we get Lemma 1 for $s = 1$. Assume Lemma 1 is true for $s = l$ and let the set K have $l+1$ elements, $K = \{k_1, \dots, k_l, k_{l+1}\}$. Consider the sets $K' = \{k_1, \dots, k_l\}$ and $M' = \{\nu_{ij} \quad | \quad i \notin K', \quad j \notin K', \quad 1 \le i < j \le p\}$. According to the assumption we have

$$f'_M = C_{p-l} |\Delta^+_{k_1,\dots,k_l}|^{\frac{n-p+l-1}{2}}, \tag{4}$$

for all points in $R_{(p-l)(p-l-1)/2}$ for which the matrix $\Delta^+_{k_1,\dots,k_l}$ is positively definite.

The density (4) is again of the form (1). Consequently, applying Theorem 1 for $k = k_{l+1}$ to the density (4) we get that the joint density of the random variables from the set $\{\nu_{ij} \quad | \quad \nu_{ij} \in M', \quad i \ne k_{l+1}, \quad j \ne k_{l+1}\} = \{\nu_{ij} \quad | \quad i \notin K, \quad j \notin K, \quad 1 \le i < j \le p\}$ is equal to

$$C_{p-l-1} |\Delta^+_{k_1,\dots,k_l,k_{l+1}}|^{\frac{n-p+l}{2}},$$

for all points in $R_{(p-l-1)(p-l-2)/2}$ for which the matrix $\Delta^+_{k_1,\dots,k_l,k_{l+1}}$ is positively definite. Hence Lemma 1 is true for $s = l+1$. Therefore it is true by induction.

Lemma 2. *The joint density of the random variables from the set $\{\nu_{ij} \mid 1 \le i < j \le p\} \backslash \{\nu_{1p}, \dots, \nu_{kp}\}$, where k is an integer, $1 \le k \le p-2$ is equal to*

$$\frac{C_{p-1} C_{p-k} |\Delta^+_p|^{\frac{n-p}{2}} \; |\Delta^+_{1,\dots,k}|^{\frac{n-p+k-1}{2}}}{C_{p-k-1} |\Delta^+_{1,\dots,k,p}|^{\frac{n-p+k}{2}}}, \tag{5}$$

for all points in $R_{p(p-1)/2-k}$ for which the matrices Δ^+_p and $\Delta^+_{1,\dots,k}$ are both positively definite.

Proof. The proof uses induction on k. Let $k = 1$. We will use the formula for the joint density of the random variables from the set $M = \{\nu_{ij} \mid 1 \le i < j \le p\} \backslash \{\nu_{ls}\}$ (l and s are integers, $1 \le l < s \le p$):

$$f_M = C_p \frac{\Gamma\left(\dfrac{(n-p+1)}{2}\right) \Gamma\left(\dfrac{1}{2}\right)}{\Gamma\left(\dfrac{n-p+2}{2}\right)} \; \frac{(|\Delta^+_l| \; |\Delta^+_s|)^{\frac{(n-p)}{2}}}{|\Delta^+_{l,s}|^{\frac{(n-p+1)}{2}}} \tag{6}$$

that is valid for all points in $R_{p(p-1)/2-1}$ for which the matrices Δ_l^+ and Δ_s^+ are both positively definite. It is proved in [3]. Then the joint density of the random variables from the set $\{\nu_{ij} \quad | \quad 1 \le i < j \le p\}\backslash\{\nu_{1p}\}$ is equal to

$$C_p \frac{\Gamma\left(\dfrac{(n-p+1)}{2}\right)\Gamma\left(\dfrac{1}{2}\right)}{\Gamma\left(\dfrac{n-p+2}{2}\right)} \frac{(|\Delta_1^+|\ |\Delta_p^+|)^{\frac{(n-p)}{2}}}{|\Delta_{1,p}^+|^{\frac{(n-p+1)}{2}}}$$

for all points in $R_{p(p-1)/2-1}$ for which the matrices Δ_1^+ and Δ_p^+ are both positively definite. From equality (2) it can be easily seen that

$$\frac{C_{p-1}}{C_{p-2}} = \frac{C_p}{C_{p-1}} \; \frac{\Gamma\left(\dfrac{n-p+1}{2}\right)\Gamma\left(\dfrac{1}{2}\right)}{\Gamma\left(\dfrac{n-p+2}{2}\right)}. \tag{7}$$

Consequently for $k = 1$ the Lemma is true.

Assume the Lemma is true for k, $1 \le k < p-2$; we will prove it for $k+1$. We have that the joint density of random variables from the set $\{\nu_{ij} \quad | \quad 1 \le i < j \le p\}\backslash\{\nu_{1p}, \ldots, \nu_{kp}\}$ is equal to (5). In order to get the joint density of the variables from the set $\{\nu_{ij} \quad | \quad 1 \le i < j \le p\}\backslash\{\nu_{1p}, \ldots, \nu_{kp}, \nu_{k+1p}\}$ we have to integrate the density (5) with respect to x_{k+1p}. This variable presents only in the matrix $\Delta_{1,\ldots,k}^+$ and does not present in the matrices Δ_p^+ and $\Delta_{1,\ldots,k,p}^+$. According to Lemma 1 the expression

$$C_{p-k}|\Delta_{1,\ldots,k}^+|^{\frac{n-p+k-1}{2}},$$

gives the joint density of the variables from the set $\{\nu_{ij} \quad | \quad k+1 \le i < j \le p\}$ for all points in $R_{(p-k)(p-k-1)/2}$, for which the matrix $\Delta_{1,\ldots,k}^+$ is positively definite. This density is of the form (1). Consequently, if we integrate it with respect to x_{k+1p} and apply formula (6) with $p = p-k$, $l = k+1$ and $s = p$ we get the density

$$C_{p-k} \frac{\Gamma\left(\dfrac{(n-p+k+1)}{2}\right)\ \Gamma\left(\dfrac{1}{2}\right)}{\Gamma\left(\dfrac{n-p+k+2}{2}\right)} \; \frac{(|\Delta_{1,\ldots,k+1}^+|\ |\Delta_{1,\ldots,k,p}^+|)^{\frac{(n-p+k)}{2}}}{|\Delta_{1,\ldots,k+1,p}^+|^{\frac{(n-p+k+1)}{2}}}.$$

Therefore the required joint density of the variables from the set $\{\nu_{ij} \mid 1 \le i < j \le p\}\backslash\{\nu_{1p}, \ldots, \nu_{kp}, \nu_{k+1p}\}$ equals to

$$\frac{C_{p-1}C_{p-k}\Gamma\left(\dfrac{(n-p+k+1)}{2}\right)\Gamma\left(\dfrac{1}{2}\right)}{C_{p-k-1}\Gamma\left(\dfrac{n-p+k+2}{2}\right)} \; \frac{|\Delta_{1,\ldots,k+1}^+|^{\frac{(n-p+k)}{2}}\ |\Delta_p^+|^{\frac{(n-p)}{2}}}{|\Delta_{1,\ldots,k+1,p}^+|^{\frac{(n-p+k+1)}{2}}}$$

for all points in $R_{p(p-1)/2-k-1}$ for which the matrices Δ_p^+ and $\Delta_{1,\ldots,k+1}^+$ are both positively definite. To prove the Lemma for $k+1$ it remains to show that

$$\frac{C_{p-k-1}}{C_{p-k-2}} = \frac{C_{p-k}\Gamma\left(\dfrac{n-p+k+1}{2}\right)\Gamma\left(\dfrac{1}{2}\right)}{C_{p-k-1}\Gamma\left(\dfrac{n-p+k+2}{2}\right)}.$$

This can be easily seen if in (7) one changes p by $p-k$. This completes the proof.

Theorem 2. *Let s and k be arbitrary integers, such that $2 \le s \le p-1$ and $1 \le k \le p-2$. Let us denote by A and B the sets $A = \{\nu_{ij} \mid 1 \le i < j \le s\}$ and $B = \{\nu_{ij} \mid k+1 \le i < j \le p\}$. Then*

$$f_{A\cup B} = \frac{f_A\, f_B}{f_{A\cap B}}. \tag{8}$$

Proof. The proof is by induction on s. If $s = p-1$ then $A \cup B = \{\nu_{ij} \mid 1 \le i < j \le p\}\backslash\{\nu_{1p}, \ldots, \nu_{kp}\}$. According to Lemmas 1 and 2

$$f_{A\cup B} = \frac{C_{p-1}|\Delta_p^+|^{\frac{n-p}{2}}\ C_{p-k}|\Delta_{1,\ldots,k}^+|^{\frac{n-p+k-1}{2}}}{C_{p-k-1}|\Delta_{1,\ldots,k,p}^+|^{\frac{n-p+k}{2}}} = \frac{f_A f_B}{f_{A\cap B}}.$$

Hence, for $s = p-1$ Theorem 2 is true. Assume the Theorem is true for s, $2 \le s \le p-1$; we will prove it for $s-1$. Let us denote by A' the set $A' = \{\nu_{ij} \mid 1 \le i < j \le s-1\}$.

Let $k \le s-2$. Then$A' \cup B = A \cup B\backslash\{\nu_{1s}, \ldots, \nu_{ks}\}$. In order to get $f_{A'\cup B}$ we have to integrate $f_{A\cup B}$ with respect to $x_{1s}, \ldots, x_{ks}$. For $f_{A\cup B}$, according to the induction assumptions, representation (8) holds. The variables $x_{1s}, \ldots, x_{ks}$ are presented in the density f_A only and are not presented in f_B and $f_{A\cap B}$. From Lemma 1 we have that

$$f_A = C_s|\Delta_{s+1,\ldots,p}^+|^{\frac{n-s-1}{2}}, \tag{9}$$

for all points in $R_{s(s-1)/2}$ for which the matrix $\Delta_{s+1,\ldots,p}^+$ is positively definite. The density (9) is of the form (1). Integrating (9) with respect to $x_{1s}, \ldots, x_{ks}$ we get the expression

$$\frac{C_{s-1}C_{s-k}|\Delta_{s,\ldots,p}^+|^{\frac{n-s}{2}}\ |\Delta_{1,\ldots,k,s+1,\ldots,p}^+|^{\frac{n-s+k-1}{2}}}{C_{s-k-1}|\Delta_{1,\ldots,k,s,\ldots,p}^+|^{\frac{n-s+k}{2}}}, \tag{10}$$

which follows from Lemma 2 with $p = s$. Using Lemma 1 it is easy to see that (10) equals to

$$\frac{f'_A f_{A\cap B}}{f_{A'\cap B}}.$$

Consequently,

$$f_{A'\cup B} = \frac{f'_A f_{A\cap B}}{f_{A'\cap B}}\frac{f_B}{f_{A\cap B}} = \frac{f'_A f_B}{f_{A'\cap B}},$$

which is the desired representation.

We now turn to the case $k > s-2$. It is easy to see that here $A \cap B = A' \cap B = \Phi$ and, by definition, $f_{A\cap B} = f_{A'\cap B} = 1$. In this case we have $A' \cup B = A \cup B \backslash \{\nu_{1s}, \dots, \nu_{s-1,s}\}$ and to get $f_{A'\cup B}$ we have to integrate $f_{A\cup B}$ with respect to the variables $x_{1s}, \dots, x_{s-1,s}$. By the induction assumption, representation (8) holds. The variables $x_{1s}, \dots, x_{s-1,s}$ are present only in the density f_A and are not present in f_B and $f_{A\cap B}$. Integrating f_A with respect to $x_{1s}, \dots, x_{s-1,s}$ we get exactly density f'_A. Therefore

$$f_{A'\cup B} = \frac{f'_A f_B}{f_{A\cap B}} = \frac{f'_A f_B}{f_{A'\cap B}},$$

which is our claim. This completes the proof.

The next Corollary follows immediately from Lemma 1 and Theorem 2:

Corollary. *Let s and k be integers, such that $2 \le s \le p-1$ and $1 \le k \le p-2$. Let us denote by A and B the sets $A = \{\nu_{ij} \mid 1 \le i < j \le s\}$ and $B = \{\nu_{ij} \mid k+1 \le i < j \le p\}$. Then*

$$f_{A\cup B} = \frac{C_s C_{p-k} |\Delta^+_{s+1,\dots,p}|^{\frac{n-s-1}{2}} \ |\Delta^+_{1,\dots,k}|^{\frac{n-p+k-1}{2}}}{f_{A\cap B}},$$

where

$$f_{A\cap B} = \begin{cases} C_{s-k} |\Delta^+_{1,\dots,k,s+1,\dots,p}|^{\frac{n-s+k-1}{2}}, & \textit{if} \quad k \le s-2 \\ 1, & \textit{if} \quad k > s-2 \end{cases}.$$

Let us denote by $f_{A\cup B/A\cap B}$, $f_{A/A\cap B}$ and $f_{B/A\cap B}$ the densities of $A \cup B$, A and B, conditioned on the random variables from the set $A \cap B$. If we divide the two sides of equality (8) by $f_{A\cap B}$ we will get the relation

$$f_{A\cup B/A\cap B} = f_{A/A\cap B} f_{B/A\cap B}. \tag{11}$$

According to the definition of conditional independence, given in [1], equality (11) shows that the random variables from the sets A and B are conditionally independent relative to the random variables from the set $A \cap B$.

REFERENCES

[1] A. P. David. Conditional independence in statistical theory (with discussion). *J. Roy. Stat. Soc.*, Ser. B, **41** (1979), No 1, 1–31.

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

[3] E. I. Veleva. Some marginal densities of the sample correlation coefficients of independent normal distributed random variables. *Annuaire de l'Universite de Sofia "St.Kliment Ohridski", Faculte des Sciences Economiques et de Gestion*, **5** (2005), to appear.

Evelina Ilieva Veleva
Department of Numerical Methods and Statistics
Rouse University
8, Studentska Str.
7017 Rouse, Bulgaria
e-mail: eveleva@abv.bg

УСЛОВНА НЕЗАВИСИМОСТ НА МНОЖЕСТВА ОТ ЕМПИРИЧНИ КОРЕЛАЦИОННИ КОЕФИЦИЕНТИ НА НЕЗАВИСИМИ НОРМАЛНО РАЗПРЕДЕЛЕНИ СЛУЧАЙНИ ВЕЛИЧИНИ

Евелина И. Велева

Получена е формула за пресмятане на един клас от маргинални плътности на съвместната плътност на емпиричните корелационни коефициенти при наблюдение над независими нормално разпределени случайни величини. Изведеното съотношение между съвместните плътности на разглежданите съвкупности от емпирични корелационни коефициенти показва тяхната условна независимост.