

Distribution of the Sample Correlation Matrix and Applications

Thu Pham-Gia, Vartan Choulakian

Department of Mathematics and Statistics, Université de Moncton, Moncton, Canada
Email: Thu.pham-gia@umoncton.ca, Vartan.choulakian@umoncton.ca

Received 29 May 2014; revised 5 July 2014; accepted 15 July 2014

Copyright © 2014 by authors and Scientific Research Publishing Inc.
This work is licensed under the Creative Commons Attribution International License (CC BY).
<http://creativecommons.org/licenses/by/4.0/>



Open Access

Abstract

For the case where the multivariate normal population does not have null correlations, we give the exact expression of the distribution of the sample matrix of correlations R , with the sample variances acting as parameters. Also, the distribution of its determinant is established in terms of Meijer G-functions in the null-correlation case. Several numerical examples are given, and applications to the concept of system dependence in Reliability Theory are presented.

Keywords

Correlation, Normal, Determinant, Meijer G-Function, No-Correlation, Dependence, Component, Formatting, Style, Styling, Insert

1. Introduction

The correlation matrix plays an important role in multivariate analysis since by itself it captures the pairwise degrees of relationship between different components of a random vector. Its presence is very visible in *Principal Component Analysis* and *Factor Analysis*, where in general, it gives results different from those obtained with the covariance matrix. Also, as a test criterion, it is used to test the independence of variables, or subsets of variables ([1], p. 407).

In a normal distribution context, when the population correlation matrix $\Lambda = I$, the identity matrix, or equivalently, the population covariance matrix Σ is diagonal, *i.e.* $\Sigma = (\sigma_{11}, \dots, \sigma_{pp})$, the distribution of the sample correlation matrix R is relatively easy to compute, and its determinant has a distribution that can be expressed as a Meijer G-function distribution. But when $\Lambda \neq I$ no expression for the density of R is presently available in the literature, and the distribution of its determinant is still unknown, in spite of efforts made by several researchers. We will provide here the closed form expression of the distribution of R , with the sample variances as parameters, hence complementing a result presented by Fisher in [2].

As explained in [3], for a random matrix, there are at least three distributions of interest, its “entries distribution” which gives the joint distribution of its matrix entries, its “determinant distribution” and its “latent roots distribution”. We will consider the first two only and note that, quite often, the first distribution is also expressed in terms of the determinant, and can lead to some confusion. In Section 2 we recall some results related to the case where $\Lambda \neq I$, and establish the new exact expression of the density of \mathbf{R} , with the sample variances $\{s_{ii}\}$ as parameters, denoted $f_{\mathbf{R}}(\mathbf{R}, \{s_{ii}\})$. In Section 3, some simulation results are given. The distribution of $|\mathbf{R}|$, the determinant of \mathbf{R} , is given in Section 4 for $\Lambda = I$. Applications of the above results to the concept of dependence within a multi-component system are given in Section 5. Numerical examples are given throughout the latter part of the paper to illustrate the results.

2. Case of the Population Correlation Matrix Not Being Identity

2.1. Covariance and Correlation Matrices

Let us consider a random vector \mathbf{X} with mean $\boldsymbol{\mu}$ and covariance matrix $\boldsymbol{\Sigma}$, of the form of a $(p \times p)$ symmetric positive definite random matrix

$$\boldsymbol{\Sigma} = \begin{bmatrix} \sigma_{11} & \sigma_{12} & \cdots & \sigma_{1p} \\ \sigma_{21} & \sigma_{22} & \cdots & \sigma_{2p} \\ \vdots & \vdots & \ddots & \vdots \\ \sigma_{p1} & \sigma_{p2} & \cdots & \sigma_{pp} \end{bmatrix}$$

of pairwise covariances between components in the matrix.

We obtain the population correlation matrix Λ by dividing each σ_{ij} by $\sqrt{\sigma_{ii}\sigma_{jj}}$. Then

$\Lambda = [\rho_{ij}] = \mathbf{D}_{\Sigma}^{-\frac{1}{2}} \boldsymbol{\Sigma} \mathbf{D}_{\Sigma}^{-\frac{1}{2}}$, where $\mathbf{D}_{\Sigma} = \text{diag}(\sigma_{11}, \dots, \sigma_{pp})$, is symmetrical, with diagonals $\rho_{ii} = 1, i = 1, \dots, p$, i.e.

$$\Lambda = \begin{bmatrix} 1 & \rho_{12} & \cdots & \rho_{1p} \\ \rho_{21} & 1 & \cdots & \rho_{2p} \\ \vdots & \vdots & \ddots & \vdots \\ \rho_{p1} & \rho_{p2} & \cdots & 1 \end{bmatrix}.$$

For a sample of size n of observations from $N_p(\boldsymbol{\mu}, \boldsymbol{\Sigma})$, the sample mean $\bar{\mathbf{X}} = \sum_{\alpha=1}^n \mathbf{X}_{\alpha} / n$ and the “adjusted sample covariance” matrix

$$\mathbf{S} = \sum_{\alpha=1}^n (\mathbf{x}_{\alpha} - \bar{\mathbf{x}})(\mathbf{x}_{\alpha} - \bar{\mathbf{x}})^T, \tag{1}$$

(or matrix of sums of squares and products, $\mathbf{S} = \begin{bmatrix} s_{11} & s_{12} & \cdots & s_{1p} \\ s_{21} & s_{22} & \cdots & s_{2p} \\ \vdots & \vdots & \ddots & \vdots \\ s_{p1} & s_{p2} & \cdots & s_{pp} \end{bmatrix}$) are independent, with the latter having a

Wishart distribution $W_p(n-1, \boldsymbol{\Sigma})$. Similarly, the correlation matrix $\mathbf{R} = \begin{bmatrix} 1 & r_{12} & \cdots & r_{1p} \\ r_{21} & 1 & \cdots & r_{2p} \\ \vdots & \vdots & \ddots & \vdots \\ r_{p1} & r_{p2} & \cdots & 1 \end{bmatrix}$ is obtained from

\mathbf{S} by using the relations: $r_{ij} = s_{ij} / \sqrt{s_{ii}s_{jj}}$, or $\mathbf{R} = \mathbf{D}_s^{-\frac{1}{2}} \mathbf{S} \mathbf{D}_s^{-\frac{1}{2}}$, where $\mathbf{D}_s = \text{diag}(s_{11}, s_{22}, \dots, s_{pp})$. We also have

the relation between determinants: $|\boldsymbol{\Sigma}| = |\Lambda| \prod_{i=1}^p \sigma_{ii}$, and, similarly, $|\mathbf{S}| = |\mathbf{R}| \prod_{i=1}^p s_{ii}$.

It is noted that by considering the usual sample covariance matrix $S^* = S/(n-1)$, which is $W_p((n-1), \Sigma/(n-1))$, we have the relation $\left\{ (s_i^*)^2 = s_{ii}/(n-1) \right\}, i = 1, \dots, p$, between the two diagonal elements, but the sample correlation matrix is the same.

The $p(p-1)/2$ coefficients r_{ij} are (marginally) distributed independently of both the sample and population means [2]. It is to be noticed that while R can always be defined from S , the reverse is not true since S has $p(p+1)/2$ independent parameters. This fact explains the differences between results when either R or S is used.

In the bivariate case, Hotelling’s expression [4] clearly shows that the density of r depends only on the population correlation coefficient ρ and the sample size n (see Section 4.3) and we will see that, similarly, the density of the sample correlation matrix, with the sample variances as parameters, is dependent on the population variances and the sample size. However, r_{ij} are biased estimators of σ_{ij} , and Olkin and Pratt [5] have suggested using the modified estimator $r_{ij} \left\{ 1 + \frac{1-r_{ij}^2}{2(n-4)} \right\}$, with a table of corrective multipliers for r_{ij} for convenience.

2.2. Some Related Work

Several efforts have been carried out in the past to obtain the exact form of the density of R for the general case, where $\rho_{ij} \neq 0, i \neq j$ and $p \geq 3$. For example, Joarder and Ali [6] derived the distribution of R for a class of elliptical models. Ali, Fraser and Lee [7], starting from the identity correlation matrix case, derived the density for the general case when $\Lambda \neq I$, again by modulating the likelihood ratio to obtain a density of R containing

the function $H_n \left(\frac{\lambda_{ij} r_{ij}}{\sqrt{\lambda_{ii} \lambda_{jj}}} \right)$, already used by Fraser ([8], p. 196) in the bivariate case. But here, $H_n(\cdot)$, expressed as an infinite series, has a much more complicated expression.

Schott ([9], p. 408), using $Vec(R)$, gives for a first-order approximation for R and the expression of $Var(Vec(R))$, and Kollo and Ruul [10], also using $Vec(R)$, presented a general method for approximating the density of R through another multivariate density, possibly one of higher dimension. Finally, Farrell ([11], p. 177) has approached the problem using exterior differential forms. However, no explicit expression for the density of R is given in any of these works, and the question rightly raised is whether that such an expression really exists.

2.3. Some New Results

We begin with the case of $\Sigma = \text{diag}(\sigma_{11}, \dots, \sigma_{pp})$. Then it can be easily established that the density of R is ([12], p. 107):

$$f(R) = \frac{\left[\Gamma\left(\frac{n-1}{2}\right) \right]^p |R|^{\frac{n-p-2}{2}}}{\Gamma_p\left(\frac{n-1}{2}\right)}, \quad -1 \leq r_{ij} = r_{ji} \leq 1, r_{ii} = 1, 1 \leq i, j \leq p \tag{2}$$

By showing that the joint density of the diagonal elements $\{s_{11}, \dots, s_{pp}\}$ with R , denoted $f(R, s_{11}, \dots, s_{pp})$, can be factorized into the product of two densities, $f_1(s_{11}, \dots, s_{pp})$ and $f_2(R)$, which has expression (2) above. $\{s_{11}, \dots, s_{pp}\}$ and R are hence independent.

We can also show that (2) is a density, *i.e.* it integrates to 1 within its definition domain, by using the approach given in Mathai and Haubold ([13], p. 421), based on matrix decomposition.

In what follows, following Kshirsagar’s approach [14], which itself is a variation of Fisher’s [2] original method, we establish first the expression of a similar joint distribution, when Σ is not diagonal.

THEOREM 1:

Let $X \sim N_p(\mu, \Sigma)$, where we suppose that the population correlation $\Lambda \neq I$ and its inverse Λ^{-1} has λ_{ii} as its diagonal elements. Then the correlation matrix R of a random sample of n observations has its distribution given by:

$$f_{\mathbf{R}}(\mathbf{R}) = \frac{\left[\Gamma\left(\frac{n-1}{2}\right) \right]^p \exp\left\{ -\sum_{i < j} \frac{\lambda_{ij} s_{ij}}{\sqrt{\sigma_{ii} \sigma_{jj}}} \right\}}{\Gamma_p\left(\frac{n-1}{2}\right) \left[|\mathbf{\Lambda}| \prod_{i=1}^p (\lambda_{ii}) \right]^{\frac{n-1}{2}}} |\mathbf{R}|^{\frac{n-p-2}{2}}, \quad (3)$$

where $\Gamma_p\left(\frac{n-1}{2}\right) = \pi^{\frac{p(p-1)}{4}} \prod_{i=1}^p \Gamma\left(\frac{n-i}{2}\right)$, with the sample covariance $\{s_{ij}\}$, $1 \leq i < j \leq p$, serving as parameters.

PROOF: Let us consider the ‘‘adjusted sample covariance matrix’’ \mathbf{S} , given by (1). We know that $\mathbf{S} \sim W_p(n-1, \mathbf{\Sigma})$, with density:

$$f(\mathbf{S}) = C |\mathbf{S}|^{\frac{n-p-2}{2}} \exp\left(-\frac{\mathbf{\Sigma}^{-1} \mathbf{S}}{2}\right),$$

with C being an appropriate constant.

Our objective is to find the joint density of \mathbf{R} with a set of variables $\{u_i\}_{i=1}^p$ so that the density of \mathbf{R} can be obtained by integrating out the variables u_i .

Let $\mathbf{\Lambda}^{-1}$ be the inverse of $\mathbf{\Lambda}$, the (population) matrix of correlations, with diagonals λ_{ii} , $1 \leq i \leq p$ and non-diagonals λ_{ij} , $1 \leq i, j \leq p$. We first transform the $p(p-1)/2$ variables s_{ij} , $i > j$, into r_{ij} . Since the diagonal s_{ii} remain unchanged, the Jacobian of the transformation from \mathbf{S} to $\{\mathbf{R}, \{s_{ii}\}\}$ is then $\prod_{i=1}^p s_{ii}^{(p-1)/2}$ and the new joint density is:

$$f(\mathbf{R}, \{s_{ii}\}) = B |\mathbf{\Sigma}|^{\frac{n-1}{2}} |\mathbf{R}|^{\frac{n-p-2}{2}} \prod_{j=1}^p s_{jj}^{\frac{n-3}{2}} \exp\left(-\frac{\text{tr}[\mathbf{D} \mathbf{\Lambda}^{-1} \mathbf{D}] \mathbf{R}}{2}\right) d\mathbf{R} \prod_{k=1}^p ds_{kk},$$

where $B = \left[2^{(n-1)p/2} \pi^{p(p-1)/4} \prod_{j=1}^p \Gamma\left(\frac{n-j}{2}\right) \right]$ and $\mathbf{D} = \text{diag}\left(\sqrt{\frac{s_{11}}{\sigma_{11}}}, \sqrt{\frac{s_{22}}{\sigma_{11}}}, \dots, \sqrt{\frac{s_{pp}}{\sigma_{pp}}}\right)$.

We set

$$u_i = b_i \sqrt{s_{ii}}, \quad 1 \leq i \leq p, \quad (4)$$

with $\sqrt{\frac{\lambda_{ii}}{\sigma_{ii}}} = b_i$, and form the symmetric matrix

$$\mathbf{\Phi} = \begin{bmatrix} 1 & \gamma_{12} & \gamma_{13} & \cdots & \gamma_{1p} \\ \gamma_{21} & 1 & \gamma_{23} & \cdots & \gamma_{2p} \\ \vdots & \vdots & \vdots & & \vdots \\ \gamma_{p1} & \gamma_{p2} & \gamma_{p3} & \cdots & 1 \end{bmatrix}, \quad (5)$$

where $\gamma_{ij} = r_{ij} \left[\frac{\lambda_{ij}}{\sqrt{\lambda_{ii} \lambda_{jj}}} \right] = s_{ij} / \sqrt{s_{ii} s_{jj}} \left[\frac{\lambda_{ij}}{\sqrt{\lambda_{ii} \lambda_{jj}}} \right]$, which depends on \mathbf{S} .

The joint density of $\{u_1, \dots, u_p\}$ and \mathbf{R} is now:

$$\frac{1}{2^{\frac{p(n-3)}{2}} \pi^{p(p-1)/4} \prod_{j=1}^p \Gamma\left(\frac{n-j}{2}\right)} \frac{1}{|\mathbf{\Lambda}|^{\frac{n-1}{2}} \prod_{j=1}^p \lambda_{jj}^{\frac{n-1}{2}}} |\mathbf{R}|^{\frac{n-p-2}{2}} \exp\left(-\frac{\mathbf{u}^T \mathbf{\Phi} \mathbf{u}}{2}\right) \prod_{j=1}^p u_j^{n-2} d\mathbf{R} du_1 \cdots du_p.$$

\mathbf{R} and $\{u_j\}_{j=1}^p$ are independent only if the above expression can be expressed as the product

$f_1(u_1, \dots, u_p) \cdot f_2(\mathbf{R})$, which is not the case since Φ contains r_{ij} .

To integrate out the vector \mathbf{u} , we set

$$G(\Phi) = \int_0^\infty \dots \int_0^\infty \exp\left\{-\frac{1}{2} \mathbf{u}^T \Phi \mathbf{u}\right\} \prod_{i=1}^p u_i^{n-2} du_1 \dots du_p,$$

(This integral is denoted $F_{n-2}(\gamma_{ij})$ by Fisher [2], and by $F_{n-2}(\Phi)$ by Kshirsagar [14] who used the notation $F(\Gamma)$ and, in both instances, was left non-computed).

Hence, if $G(\Phi)$ is constant, as in the case when $\Phi = \mathbf{I}$, \mathbf{R} would have density

$$f_{\mathbf{R}}(\mathbf{R}) = \left\{ \frac{G(\Phi)}{2^{p \frac{n-3}{2}} \Gamma_p\left(\frac{n-1}{2}\right) \left[|\Lambda| \prod_{i=1}^p (\lambda_{ii})\right]^{\frac{n-1}{2}}} \right\} |\mathbf{R}|^{\frac{n-p-2}{2}}, \tag{6}$$

with $\Gamma_p\left(\frac{n-1}{2}\right) = \pi^{\frac{p(p-1)}{4}} \prod_{i=1}^p \Gamma\left(\frac{n-i}{2}\right)$, being the p -gamma function. However, in the general case this is not true.

Consider the quadratic form:

$$\mathbf{u}^T \Phi \mathbf{u} = \sum_{i=1}^p \gamma_{ii} u_i^2 + 2 \sum_{i < j} \gamma_{ij} u_i u_j = \sum_{i=1}^p \frac{\lambda_{ii}}{\sigma_{ii}} s_{ii} + 2 \sum_{i < j} \frac{\lambda_{ij} s_{ij}}{\sqrt{\sigma_{ii} \sigma_{jj}}},$$

and

$$G(\Phi) = \int_0^\infty \dots \int_0^\infty \exp\left\{-\frac{1}{2} \sum_{i=1}^p \frac{\lambda_{ii}}{\sigma_{ii}} s_{ii} - \sum_{i < j} \frac{\lambda_{ij} s_{ij}}{\sqrt{\sigma_{ii} \sigma_{jj}}}\right\} \left(\prod_{i=1}^p u_i^{n-2}\right) du_1 \dots du_p.$$

Changing to s_{ij} , we have

$$\begin{aligned} G(\Phi) &= \int_0^\infty \dots \int_0^\infty \exp\left\{-\frac{1}{2} \sum_{i=1}^p \frac{\lambda_{ii}}{\sigma_{ii}} s_{ii} - \sum_{i < j} \frac{\lambda_{ij} s_{ij}}{\sqrt{\sigma_{ii} \sigma_{jj}}}\right\} \left(\prod_{i=1}^p (b_i \sqrt{s_{ii}})^{n-2}\right) \left(\prod_{i=1}^p \left(\frac{1}{2} b_i \frac{ds_{ii}}{\sqrt{s_{ii}}}\right)\right) \\ &= \left(\frac{1}{2}\right)^p \left(\prod_{i=1}^p b_i\right)^{n-1} \exp\left\{-\sum_{i < j} \frac{\lambda_{ij} s_{ij}}{\sqrt{\sigma_{ii} \sigma_{jj}}}\right\} \prod_{i=1}^p \left(\int_0^\infty \exp\left(-\frac{1}{2} \frac{\lambda_{ii}}{\sigma_{ii}} s_{ii}\right) s_{ii}^{\frac{n-3}{2}} ds_{ii}\right). \end{aligned}$$

Since each integral is a gamma density in s_{ii} , with value $\Gamma\left(\frac{n-1}{2}\right) \left(\frac{2\sigma_{ii}}{\lambda_{ii}}\right)^{\frac{n-1}{2}}$, we obtain

$$\begin{aligned} G(\Phi) &= \left(\frac{\Gamma\left(\frac{n-1}{2}\right)}{2}\right)^p \left(\prod_{i=1}^p \frac{2\sigma_{ii}}{\lambda_{ii}}\right)^{\frac{n-1}{2}} \left(\prod_{i=1}^p \sqrt{\frac{\lambda_{ii}}{\sigma_{ii}}}\right)^{n-1} \exp\left\{-\sum_{i < j} \frac{\lambda_{ij} s_{ij}}{\sqrt{\sigma_{ii} \sigma_{jj}}}\right\} \\ &= 2^{\frac{(n-3)p}{2}} \left[\Gamma\left(\frac{n-1}{2}\right)\right]^p \exp\left\{-\sum_{i < j} \frac{\lambda_{ij} s_{ij}}{\sqrt{\sigma_{ii} \sigma_{jj}}}\right\}. \end{aligned}$$

$G(\Phi)$ contains all non-diagonal entries of the sample covariance matrix, and depends on \mathbf{S} . We now obtain from (6) expression (3) of **Theorem 1**. Here, the off-diagonal sample covariance, $\{s_{ij}\}$, serve as parameters of this density.

QED.

Alternately, using the corresponding correlation coefficients, we have:

$$f_{\mathbf{R}}(\mathbf{R}) = \frac{\left[\Gamma\left(\frac{n-1}{2}\right)\right]^p}{\Gamma_p\left(\frac{n-1}{2}\right)} \frac{1}{\left[|\mathbf{\Lambda}|\prod_{i=1}^p(\lambda_{ii})\right]^{\frac{n-1}{2}}} \cdot \exp\left\{-\sum_{i<j} \lambda_{ij} r_{ij} \sqrt{\frac{s_{ii}s_{jj}}{\sigma_{ii}\sigma_{jj}}}\right\} |\mathbf{R}|^{\frac{n-p-2}{2}}, \tag{7}$$

REMARKS:

1) For $\mathbf{\Lambda} = \mathbf{I}$ our results given above should reduce to known results, and they do. Indeed, since we now have $\mathbf{\Phi} = \mathbf{I}$

$$G(\mathbf{I}) = \int_0^\infty \exp\left(-\frac{1}{2}u^T \mathbf{\Phi} u\right) \prod_{i=1}^p u_i^{n-2} du_1 \cdots du_p = 2^{\frac{p(n-3)}{2}} \left[\Gamma\left(\frac{n-1}{2}\right)\right]^p.$$

Hence,

$$f_{\mathbf{R}}(\mathbf{R}) = \frac{\left[\Gamma\left(\frac{n-1}{2}\right)\right]^p |\mathbf{R}|^{\frac{n-p-2}{2}}}{\pi^{\frac{p(p-1)}{4}} \prod_{i=1}^p \Gamma\left(\frac{n-i}{2}\right)} = \frac{\left[\Gamma\left(\frac{n-1}{2}\right)\right]^p |\mathbf{R}|^{\frac{n-p-2}{2}}}{\Gamma_p\left(\frac{n-1}{2}\right)}$$

as in (2) since we now have $|\mathbf{\Lambda}| = \prod_{i=1}^p (\lambda_{ii})^{\frac{n-1}{2}} = 1$.

Here, only the value of p needs to be known and this explains why expression (2) depends only on n and p. Also, as pointed out by Muirhead ([15], p. 148), if we do not suppose normality, the same results can be obtained under some hypothesis.

2) Expression (3) can be interpreted as the density of \mathbf{R} when $\{\sigma_{ii}\}_{i=1}^p$ are known, and $\{s_{ij}\}, 1 \leq i < j \leq p$, is a set of constant sample covariances. But when this set is considered as a random vector, with a certain distribution, $f_{\mathbf{R}}(\mathbf{R})$ is called mixture distribution, defined in two steps:

- a) $\{s_{ij}\}$ has an (known or unknown) distribution φ_0 , i.e. $\{s_{ij}\} \sim \varphi_0$.
- b) $(\mathbf{R}; \{s_{ij}\}) \sim \varphi_0$ has the density given by (3), denoted \mathfrak{Z}_0 .

The distribution of \mathbf{R} is then $\mathfrak{Z}_0 * \varphi_0$, where $*$ denotes the mixture operation.

However, a closed form for this mixture is often difficult to obtain.

Alternately, $f_{\mathbf{R}}(\mathbf{R})$ could be $\mathfrak{Z}_1 * \varphi_1$, with φ_1 being the density of the diagonal sample variances $\{s_{ii}\}$ and off-diagonal sample correlations, $\{r_{ij}\}_{1 \leq i < j \leq p}$ while \mathfrak{Z}_1 is given by (7).

3) We have, using (3), and the relation $|\mathbf{S}| = |\mathbf{D}_S| \cdot |\mathbf{R}|$ the following equivalent expression, where $\mathbf{D}_S = \text{diag}(s_{11}, \dots, s_{pp})$:

$$f_{\mathbf{R}}(\mathbf{R}; \mathbf{S} = \mathbf{S}_0) = \frac{\left[\Gamma\left(\frac{n-1}{2}\right)\right]^p}{\Gamma_p\left(\frac{n-1}{2}\right)} \frac{\exp\left\{-\sum_{i<j} \lambda_{ij} (s_0)_{ij}\right\}}{|\mathbf{D}_S|^{\frac{n-p-2}{2}} \cdot \left[|\mathbf{\Lambda}|\left(\prod_{i=1}^p \lambda_{ii}\right)\right]^{\frac{n-1}{2}}} \cdot |\mathbf{S}_0|^{\frac{n-p-2}{2}}. \tag{8}$$

Expression (8) gives the positive numerical value of $f_{\mathbf{R}}(\mathbf{R}; \mathbf{S} = \mathbf{S}_0)$ upon knowledge of the value of \mathbf{S} . It will serve to set up simulation computations in Section 3. However, it also shows that $f_{\mathbf{R}}(\mathbf{R})$ can be defined with \mathbf{S} , and when \mathbf{S} has a certain distribution the values of $f_{\mathbf{R}}(\mathbf{R})$ are completely determined with that distribution. This highlights again the fact that, in statistics, using \mathbf{R} or \mathbf{S} can lead to different results, as mentioned previously.

2.4. Other Known Results

1) The distribution of the sample coefficient of correlation in the bivariate normal case can be determined

fairly directly when integrating out s_1 and s_2 , and this fact is mentioned explicitly by Fisher ([2], p. 4), who stated: “This, however, is not a feasible path for more than two variables.” In the bivariate case,

$$r = \frac{\sum_{i=1}^n (x - \bar{x})(y - \bar{y})}{\sqrt{\sum_{i=1}^n (x - \bar{x})^2 \sum_{i=1}^n (y - \bar{y})^2}},$$

and has been well-studied by several researchers, using different approaches and,

as early as 1915, Fisher [16] gave its density as:

$$f(r) = \frac{(1 - \rho^2)^{\frac{n-1}{2}}}{\pi \Gamma(n-2)} (1 - r^2)^{\frac{n-4}{2}} \frac{d^{n-2}}{d(\rho r)^{n-2}} \left\{ \frac{\arccos(-\rho r)}{\sqrt{1 - (\rho r)^2}} \right\}, \quad -1 \leq r \leq 1, \tag{9}$$

Using geometric arguments, with ρ being the population coefficient of correlation. We refer to ([17], p. 524-534) for more details on the derivation of the above expression. Other equivalent expressions, reportedly as numerous as 52, were obtained by other researchers, such as Hotelling [4], Sawkins [18], Ali *et al.* [7].

2) Using Equation (3) on the sample correlation matrix $\begin{bmatrix} 1 & r \\ r & 1 \end{bmatrix}$, obtained from the sample covariance matrix

$$\begin{bmatrix} s_{11} & s_{12} = r\sqrt{s_{11}s_{22}} \\ s_{21} = r\sqrt{s_{11}s_{22}} & s_{22} \end{bmatrix},$$

Together with $p = 2$, we can also arrive at one of these forms (see also Section 4.3, where the determinant of \mathbf{R} provides a more direct approach).

3) Although Fisher [2] did not give the explicit form of the integral $G(\Phi)$ above, he included several interesting results on $G(\Phi) = F_{n-2}(\Phi)$, as a function of the sample size n and coefficients $\gamma_{ij} = r_{ij} \left[\frac{\lambda_{ij}}{\sqrt{\lambda_{ii}\lambda_{jj}}} \right]$. For

example, in the case all $\rho_{ij} = 0, i \neq j$, the generalized volume in the $p(p - 1)/2$ -dimension space of the region of integration for r_{ij} is found to be a function of p , having a maximum at $p = 6$. In the case all $\gamma_{ij} = 0$, the expressions of the partial derivatives of $\text{Log } G(\Phi)$ w.r.t. γ_{ij} can be obtained, and so are the mixed derivatives.

For the case $n = 2$, and $p = 3$, $G(\Phi)$ has an interesting geometrical interpretation as $2^{(p-2)/2} \Gamma\left(\frac{n}{2}\right) \frac{V}{D}$, where

V = generalized volume defined by $-\gamma_{ij}$ and D is the volume defined by the p unit vectors in a transformation where $-\gamma_{ij}$ are the cosines of the angles between pairs of edges.

3. Some Computation and Simulation Results

3.1. Simulations Related to R

A matrix equation such as (3) can be difficult to visualize numerically, especially when the dimensions are high, *i.e.* $p \geq 3$. Ideally, to illustrate (7), a figure giving $f_{\mathbf{R}}(\mathbf{R})$ in function of the matrix \mathbf{R} itself is most informative, but, naturally, impossible to obtain. One question we can investigate is how the values of $W = f_{\mathbf{R}}(\mathbf{R})$ distributed, for a normal model $\mathbf{X} \sim N_p(\boldsymbol{\mu}, \boldsymbol{\Sigma})$? Simulation using (8) can provide some information on this distribution in some specific cases. For example, we can start from the (4×4) population covariance matrix

$$\boldsymbol{\Sigma} = \begin{bmatrix} 0.122 & 0.097 & 0.016 & 0.010 \\ & 0.140 & 0.011 & 0.009 \\ & & 0.030 & 0.006 \\ & & & 0.011 \end{bmatrix},$$

taken from our analysis of Fisher’s iris data [19]. It concerns the Setosa iris variety, with x_1 = sepal length, x_2 = sepal with, x_3 = petal length and x_4 = petal width. It gives the population correlation matrix

$$\Lambda = \begin{bmatrix} 1 & 0.7425 & 0.2672 & 0.2781 \\ & 1 & 0.1777 & 0.2328 \\ & & 1 & 0.3316 \\ & & & 1 \end{bmatrix},$$

where all $\rho_{ij} \neq 0$. We generate 10,000 samples of 100 observations each from $N_4(\boldsymbol{\mu}, \boldsymbol{\Sigma})$, which give 10,000 values of the covariance matrix \mathbf{S} , which, in turn, give matrix values for \mathbf{R} , scalar values for $|\mathbf{R}|$ and, finally, positive scalar values for $W = f_{\mathbf{R}}(\mathbf{R}; \mathbf{S} = \mathbf{S}_0)$, as given by (8). Recall that for \mathbf{X} normal \mathbf{S} has a Wishart distribution $W_p(n-1, \boldsymbol{\Sigma})$. **Figure 1** gives the corresponding histogram, which shows that values of W are distributed along a unimodal density, denoted by $h(W)$, with a very small variation interval, *i.e.* most of its values are concentrated around the mode.

Note: A special approach to graphing distributions of covariance matrices, using the principle of decomposing a matrix into scale parameters and correlations, is presented in: Tokuda, T., Goodrich, B., Van Mechelen, I., Gelman, A. and Tuerlinckx, F., *Visualizing Distributions of Covariance Matrices* (Document on the Internet).

It is also mentioned there that for the Inverted Wishart case, with ν degrees of freedom, then

$$f(\mathbf{R}) = |\mathbf{R}|^{\frac{(\nu-1)(p-1)}{2}-1} \left(\prod_{i=1}^p |\mathbf{R}_{ii}| \right)^{-\nu/2},$$

where \mathbf{R}_{ii} is the i -th principal sub-matrix of \mathbf{R} , obtained by removing row and column i (p. 12).

3.2. Simulations Related to $|\mathbf{R}|$

Similarly, the same application above gives the approximate simulated distribution of $|\mathbf{R}|$ presented in **Figure 2**. We can see that it is a unimodal density which depends on the correlation coefficients r_{ij} .

Using expression (7), which exhibits explicitly r_{ij} , and replacing r_{ij} by the corrected value $r_{ij} \left\{ 1 + \frac{1-r_{ij}^2}{2(n-4)} \right\}$,

the unbiased estimator of ρ_{ij} , we obtain $|\hat{\mathbf{R}}|$. However, since an unbiased estimator of $|\Lambda|$ is still to be found we cannot use neither $|\mathbf{R}|$ nor $|\hat{\mathbf{R}}|$ as point-estimate of $|\Lambda|$. **Figure 2** gives the simulated distributions of $|\mathbf{R}|$ and of $|\hat{\mathbf{R}}|$. We can see that the two approximate densities are different, and the density of $|\hat{\mathbf{R}}|$ has higher mean and median, resulting in a shift to the right. But, again, the two variation intervals are very small.

3.3. Expression of $G(\Phi)$

In the proof of **Theorem 1**, we have established that

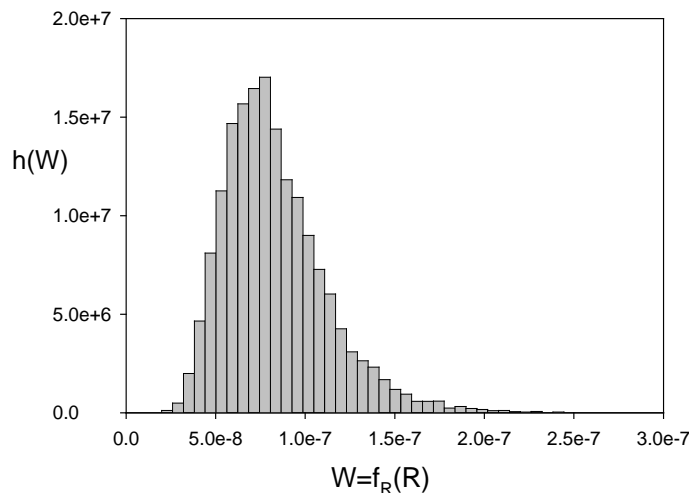


Figure 1. Simulated density of $W = f_{\mathbf{R}}(\mathbf{R}; \mathbf{S} = \mathbf{S}_0)$.

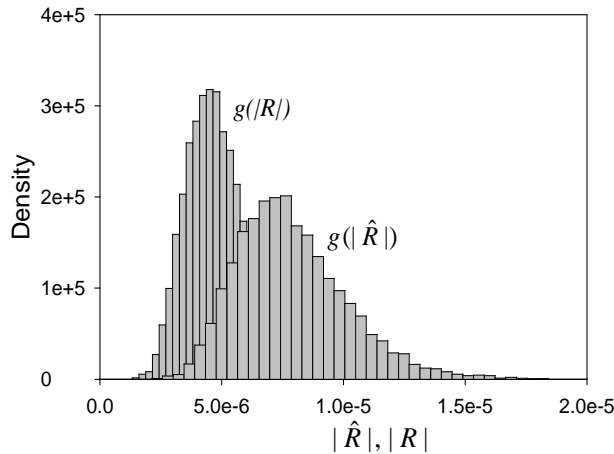


Figure 2. Simulated densities of $|\mathbf{R}|$ and $|\hat{\mathbf{R}}|$.

$$G(\Phi) = \int_0^\infty \dots \int_0^\infty \exp\left\{-\frac{1}{2} \mathbf{u}^T \Phi \mathbf{u}\right\} \prod_{i=1}^p u_i^{n-2} du_1 \dots du_p = 2^{\frac{(n-3)p}{2}} \left[\Gamma\left(\frac{n-1}{2}\right)\right]^p \exp\left\{-\sum_{i < j} \frac{\lambda_{ij} s_{ij}}{\sqrt{\sigma_{ii} \sigma_{jj}}}\right\}.$$

Using the above matrix Σ , for a simulated sample, say

$$S = \begin{bmatrix} 0.1106 & -0.019 & 0.015 & 0.017 \\ & 0.0384 & 0.0004 & 0.0092 \\ & & 0.0098 & 0.0034 \\ & & & 0.172 \end{bmatrix},$$

we compute directly the left side by numerical integration, and the right side by using the algebraic expression. The results are extremely close to each other, with both around the numerical value 1.238523012×10^5 .

4. Distribution of $|\mathbf{R}|$, the Determinant of \mathbf{R}

First, let $\det(\mathbf{R})$ be denoted by $|\mathbf{R}|$. In this case $\Lambda \neq \mathbf{I}$, this distribution is very complex and no related result is known when $p \geq 3$. Nagar and Castaneda [20], for example, established some results in the general case, for $p = 2$.

Theoretically, we can obtain the density of $|\mathbf{R}|$ from (3) by applying the transformation $\mathbf{R} \rightarrow |\mathbf{R}|$, with differential $d|\mathbf{R}| = |\mathbf{R}| \text{tr}(\mathbf{R}^{-1}(d\mathbf{R}))$, but the expression obtained quickly becomes intractable. Only in the case of $\Lambda = \mathbf{I}$ that we can derive some analytical results on $|\mathbf{R}|$, as presented in the next section. Gupta and Nagar [21] established some results for the case of a mixture of normal models, but again under the hypothesis of $\Lambda = \mathbf{I}$.

4.1. Density of the Determinant $|\mathbf{R}|$

When considering Meijer \mathbf{G} -functions and their extensions, Fox's \mathbf{H} -functions [22], for $\Lambda = \mathbf{I}$ the density of $|\mathbf{R}|$ can be expressed in closed forms, as are those of other related multivariate statistics [23]. Let us recall that

the Meijer function $\mathbf{G}(x)$, and the Fox function $\mathbf{H}(x)$, are defined as follows: $\mathbf{G}(x) = \mathbf{G}_{p,q}^{m,r} \left[x \left| \begin{matrix} a_1, \dots, a_p \\ b_1, \dots, b_q \end{matrix} \right. \right]$ is

the integral along the complex contour L of a rational expression of Gamma functions

$$\mathbf{G}_{p,q}^{m,r} \left[x \left| \begin{matrix} a_1, \dots, a_p \\ b_1, \dots, b_q \end{matrix} \right. \right] = \frac{1}{2\pi i} \int_L \frac{\prod_{j=1}^m \Gamma(b_j - s) \prod_{j=1}^r \Gamma(1 - a_j + s)}{\prod_{j=m+1}^q \Gamma(1 - b_j + s) \prod_{j=r+1}^p \Gamma(a_j - s)} x^{-s} ds \tag{10}$$

It is a special case, when $\alpha_i = \beta_j = 1, \forall i, j$, of Fox's H -function, defined as:

$$H_{p,q}^{m,r} \left[x \left| \begin{matrix} (a_1, \alpha_1), \dots, (a_p, \alpha_p) \\ (b_1, \beta_1), \dots, (b_q, \beta_q) \end{matrix} \right. \right] = \frac{1}{2\pi i} \int_L \frac{\prod_{j=1}^m \Gamma(b_j - \beta_j s) \prod_{j=1}^r \Gamma(1 - a_j + \alpha_j s)}{\prod_{j=m+1}^q \Gamma(1 - b_j + \beta_j s) \prod_{j=r+1}^p \Gamma(a_j - \alpha_j s)} x^{-s} ds.$$

Under some fairly general conditions on the poles of the gamma functions in the numerator, the above integrals exist.

THEOREM 2: When $\Lambda = I$, for a random sample of size n from $N_p(\mu, \Sigma)$, the density of $|\mathbf{R}|$ depends only on n and p :

$$g(u; n, p) = \frac{\left[\Gamma\left(\frac{n-1}{2}\right) \right]^{p-1}}{\Gamma\left(\frac{n-2}{2}\right) \dots \Gamma\left(\frac{n-p}{2}\right)} \mathbf{G}_{p-1}^{p-1, 0} \left[u \left| \begin{matrix} \frac{n-3}{2}, \dots, \frac{n-3}{2} \\ \frac{n-4}{2}, \dots, \frac{n-(p+2)}{2} \end{matrix} \right. \right], 0 \leq u \leq 1. \tag{11}$$

PROOF:

From (2) the moments of order t of $|\mathbf{R}|$ are:

$$E(|\mathbf{R}|^t) = \frac{\left[\Gamma\left(\frac{n-1}{2}\right) \right]^p \prod_{j=1}^p \Gamma\left(\frac{n-j}{2} + t\right)}{\left[\Gamma\left(\frac{n-1}{2} + t\right) \right]^p \prod_{j=1}^p \Gamma\left(\frac{n-j}{2}\right)} = \frac{\left[\Gamma\left(\frac{n-1}{2}\right) \right]^{p-1} \prod_{j=2}^p \Gamma\left(\frac{n-j}{2} + t\right)}{\left[\Gamma\left(\frac{n-1}{2} + t\right) \right]^{p-1} \prod_{j=2}^p \Gamma\left(\frac{n-j}{2}\right)}, \tag{12}$$

$t \geq 1$, which is a product of moments of order t of independent beta variables. Upon identification of (12) with these products, we can see that $|\mathbf{R}| \sim X_1 \dots X_{p-1}$, with $X_j \sim \beta\left(\frac{n-j}{2}, \frac{j-1}{2}\right)$, with $j = 2, 3, \dots, p$. Using [23], the product of k independent betas, $\beta(\alpha^{(i)}, \beta^{(i)})$, has as density

$$g(y; \alpha^{(1)}, \beta^{(1)}, \dots, \alpha^{(k)}, \beta^{(k)}) = \prod_{i=1}^k \frac{\Gamma(\alpha^{(i)} + \beta^{(i)})}{\Gamma(\alpha^{(i)})} \mathbf{G}_{k-1}^{k-1, 0} \left[y \left| \begin{matrix} \alpha^{(1)} + \beta^{(1)} - 1, \dots, \alpha^{(k)} + \beta^{(k)} - 1 \\ \alpha^{(1)} - 1, \dots, \alpha^{(k)} - 1 \end{matrix} \right. \right], 0 \leq y \leq 1$$

Hence, we have here, the density of $|\mathbf{R}|$ as:

$$g(r; n, p) = \prod_{i=1}^{p-1} \frac{\Gamma(\alpha^{(i)} + \beta^{(i)})}{\Gamma(\alpha^{(i)})} \mathbf{G}_{p-1}^{p-1, 0} \left[r \left| \begin{matrix} \alpha^{(1)} + \beta^{(1)} - 1, \dots, \alpha^{(p-1)} + \beta^{(p-1)} - 1 \\ \alpha^{(1)} - 1, \dots, \alpha^{(p-1)} - 1 \end{matrix} \right. \right], 0 \leq r \leq 1,$$

where $\alpha^{(1)} = \frac{n-2}{2}, \beta^{(1)} = \frac{1}{2}, \alpha^{(2)} = \frac{n-3}{2}, \beta^{(2)} = \frac{2}{2}$ and $\alpha^{(p-1)} = \frac{n-p}{2}, \beta^{(p-1)} = \frac{p-1}{2}$.

And, hence, we obtain:

$$g(u; n, p) = \frac{\left[\Gamma\left(\frac{n-1}{2}\right) \right]^{p-1}}{\Gamma\left(\frac{n-2}{2}\right) \dots \Gamma\left(\frac{n-p}{2}\right)} \mathbf{G}_{p-1}^{p-1, 0} \left[u \left| \begin{matrix} \frac{n-3}{2}, \dots, \frac{n-3}{2} \\ \frac{n-4}{2}, \dots, \frac{n-(p+2)}{2} \end{matrix} \right. \right], 0 \leq u \leq 1 \tag{13}$$

QED.

The density of $|\mathbf{R}|$ can easily be computed and graphed, and percentiles of $|\mathbf{R}|$ can be determined numerically. For example, for $p = 4, n = 8$. The 2.5th and 97.5th percentiles can be found to be 0.04697 and 0.7719 respectively.

4.2. Product and Ratio

Let \mathbf{R}_1 and \mathbf{R}_2 be two independent correlation matrices, obtained from 2 populations, each with zero popula-

tion correlation coefficients. The determinant of their product is also a G -function distribution, and its density can be obtained. This result is among those which extend relations obtained in the univariate case by Pham-Gia and Turkkan [24], and also has potential applications in several domains.

THEOREM 3: Let $\{X_1^{(1)}, X_2^{(1)}, \dots, X_{n_1}^{(1)}\}$ and $\{X_1^{(2)}, X_2^{(2)}, \dots, X_{n_2}^{(2)}\}$ be two independent random samples from $N_{p_1}(\mu_1, \Sigma_1)$ and $N_{p_2}(\mu_2, \Sigma_2)$ respectively, both Σ_i being diagonal. Then the determinant $|\mathbf{R}|$ of the product $\mathbf{R} = \mathbf{R}_1 \mathbf{R}_2$ of the two correlation matrices \mathbf{R}_1 and \mathbf{R}_2 has density:

$$g(v; n_1, n_2, p) = A G_{p-2, p-2}^{p-2, 0} \left[v \left| \begin{array}{c} \frac{n_1-3}{2}, \dots, \frac{n_1-3}{2}, \frac{n_2-3}{2}, \dots, \frac{n_2-3}{2} \\ \frac{n_1-4}{2}, \dots, \frac{n_1-(p_1+2)}{2}, \frac{n_2-4}{2}, \dots, \frac{n_2-(p_2+2)}{2} \end{array} \right. \right], 0 \leq v \leq 1, \quad (14)$$

where $A = \prod_{i=1}^2 \frac{\left[\Gamma\left(\frac{n_i-1}{2}\right) \right]^{p_i-1}}{\Gamma\left(\frac{n_i-2}{2}\right) \dots \Gamma\left(\frac{n_i-p_i}{2}\right)}$ and $p = p_1 + p_2$.

PROOF: Immediate by using multiplication of G -densities presented in [23].

QED.

Figure 3 shows the density of $|\mathbf{R}| = |\mathbf{R}_1| |\mathbf{R}_2|$, for $n_1 = 8, p_1 = 4$ and $n_2 = 10, p_2 = 5$.

Using again results presented in [23] we can similarly derive the density of the ratio $|\mathbf{R}_1|/|\mathbf{R}_2|$ in terms of G -functions. Its expression is not given here to save space but is available upon request.

4.3. Particular Cases

1) Bivariate normal case: a) for the bivariate case we have $|\mathbf{R}| = 1 - r^2$, and when ρ is zero, we have from (11) the density of $|\mathbf{R}|$ as

$$\frac{\Gamma\left(\frac{n-1}{2}\right)}{\Gamma\left(\frac{n-2}{2}\right)} G_{1, 1}^{1, 0} \left[u \left| \begin{array}{c} \frac{n-3}{2} \\ \frac{n-4}{2} \end{array} \right. \right], 0 \leq u \leq 1,$$

which is the G -function form of the beta $\beta\left(\frac{n-2}{2}, \frac{1}{2}\right)$. Hence, the distribution of r^2 is: $\beta(1/2, (n-2)/2)$,

$0 \leq r^2 \leq 1$, and the density of r is:

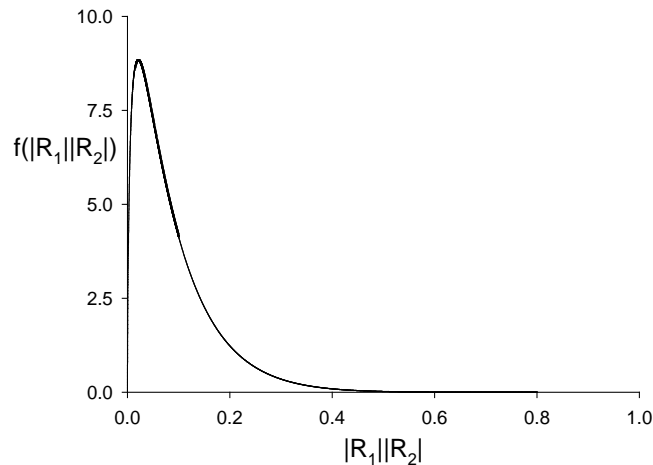


Figure 3. Density of product of independent correlation determinants.

$$f(r) = (1-r^2)^{(n-4)/2} / B(1/2, (n-2)/2), \tag{15}$$

for $-1 \leq r \leq 1$. Testing $H_0 : \rho = 0$ is much simpler when using Student's t -distribution, $t = \frac{R\sqrt{n-2}}{\sqrt{1-R^2}}$, and is

covered in most textbooks.

Pitman [25] has given an interesting distribution-free test when $\rho = 0$.

2) When $\rho \neq 0$ Hotelling ([4]) gave the density of r as:

$$f(r) = \frac{(n-2)\Gamma(n-1)}{\sqrt{2\pi}\Gamma\left(n-\frac{1}{2}\right)} (1-\rho^2)^{\frac{n-1}{2}} (1-r^2)^{\frac{n-4}{2}} {}_2F_1\left(\frac{1}{2}, \frac{1}{2}, \frac{2n-1}{2}; \frac{\rho r+1}{2}\right), -1 \leq r \leq 1, \tag{16}$$

where ${}_2F_1(a, b; c; x)$ is Gauss hypergeometric function with parameters a, b and c .

3) **Mixture of Normal Distributions:** With X coming now from the mixture: $\varepsilon N_p(\boldsymbol{\mu}_1, \boldsymbol{\Sigma}) + (1-\varepsilon) N_p(\boldsymbol{\mu}_2, \boldsymbol{\Sigma})$, $0 < \varepsilon < 1$, Gupta and Nagar [21] consider $\mathbf{W} = \det(\mathbf{S}^*) / \prod_{i=1}^p s_{ii}$, and give the density of \mathbf{W} in terms of Meijer G -functions, but for the case $\boldsymbol{\Sigma} = \mathbf{I}$ only. The complicated form of this density contains the hypergeometric function ${}_2F_1(\cdot)$, as expected.

For the bivariate case, Nagar and Castaneda [20] established the density of r and gave its expression for both cases, $\rho \neq 0$ and $\rho = 0$. In the first case the density of r , when only one population is considered, reduces to the expression obtained by Hotelling [4] above.

5. Dependence between Components of a Random Vector

5.1. Dependence and $|\mathbf{R}|$

Correlation is useful in multiple regression analysis, where it is strongly related to collinearity. As an example of how individual correlation coefficients are used in regression, the variance inflation factor (VIF), well adopted now in several statistical softwares, measures how much the variance of a coefficient is increased by collinearity, or in other words, how much of the variation in one independent variable is explained by the others. For the j -th variable, VIF_j is the j -th diagonal element of \mathbf{R}^{-1} . We know that it equals $(1 - \mathbf{R}_{j,12\dots j-1,j+1,\dots,p}^2)^{-1}$, with $\mathbf{R}_{j,12\dots j-1,j+1,\dots,p}$ being the multiple correlation of the j -th variable regressed on the remaining $p-1$ others.

When all correlation measures are considered together, measuring intercorrelation by a single number has been approached in different ways by various authors. Either the value of $|\mathbf{R}|$ or those of its latent roots can be used. Rencher ([26], p. 21) mentions six of these measures, among them $1 - \sqrt{|\mathbf{R}|}$, where $|\mathbf{R}|$ is an observed value of \mathbf{R} , and takes the value 1 if the variables are independent, and 0 if there is an exact linear dependence. But since the exact distribution of \mathbf{R} is not available this sample measure is rather of descriptive type and no formal inferential process has really been developed.

Although the notion of independence between different components of a system is of widespread use in the study of the system structure, reliability and performance, its complement, the notion of dependence has been a difficult one to deal with. There are several dependence concepts, as explained by Jogdev [27], but using the covariance matrix between different components in a joint distribution remains probably the most direct approach. Other more theoretical approaches, are related to the relations between marginal and joint distributions, and Joe [28] can be consulted on these approaches. Still, other aspects of dependence are explored in Bertail *et al.* [29]. But two random variables can have zero correlation while being dependent. Hence, no-correlation and independence are two different concepts, as pointed out in Drouet Mari and Kotz [30]. Furthermore, for two independent events, the product of their probabilities gives the probability of the intersection event, which is not necessarily the case for two non-correlated events. Fortunately, these two concepts are equivalent, when the underlying population is supposed normal, a hypothesis that we will suppose in this section.

5.2. Inner Dependence of a System

When considering only two variables, several measures of dependence have also been suggested in the literature

(Lancaster [31]), and especially in system reliability (Hoyland and Rausand [32]), but a joint measure of the degree of dependence between several components of a random vector, or within a system \mathcal{G} , or *inner dependence* of \mathcal{G} , denoted by $\delta(\mathcal{G})$, is still missing. We approach this dependence concept here by way of the correlation matrix, where a single measure attached to it would reflect the overall degree of dependence. This concept has been presented first in Bekker, Roux and Pham-Gia [33], to which we refer for more details. It is defined as $\delta(\mathcal{G}) = (1 - |\Lambda|)^{1/p}$, with $0 \leq \delta(\mathcal{G}) \leq 1$. The measure of independence within the system is then $1 - (1 - |\Lambda|)^{1/p}$, estimated by $1 - (1 - |\hat{\Lambda}|)^{1/p}$, where $\hat{\Lambda}$ is a point estimation of Λ based on \mathbf{R} , the correlation matrix associated with a sample of n observations of the p -component system. In the general case this estimation question is still unresolved, except for the binormal case, $\mathcal{G} \sim N_2(\boldsymbol{\mu}, \boldsymbol{\Sigma})$. We then have $|\Lambda| = 1 - \rho^2$, and $\delta(\mathcal{G}) = |\rho|$, where ρ is the coefficient of correlation with its estimation well known, depending on either ρ is supposed to be zero or not. The associated sample measure being $d(\mathcal{G}) = |\rho|$, it is of interest to study the distribution of the sample *inner dependence* $d(\mathcal{G})$, based on a sample of n observations of the system.

In the language of Reliability Theory, a p -component normal system is *fully statistically independent* when the $p(p-1)/2$ correlation coefficients ρ_{ij} of its components are all zero. We have:

THEOREM 4: 1) Let the *fully statistically independent* system \mathcal{G} have p components with a joint normal distribution with $\boldsymbol{\Sigma} = \mathbf{I}$, where $p \geq 2$.

a) Then the distribution of the sample coefficient of *inner dependence* $d(\mathcal{G})$ is:

$$f(d; n, p) = \frac{\left[\Gamma\left(\frac{n-1}{2}\right) \right]^{p-1}}{\Gamma\left(\frac{n-2}{2}\right) \cdots \Gamma\left(\frac{n-p}{2}\right)} p(1-d)^{p-1} \mathbf{G}_{p-1}^{p-1, 0} \left[1 - (1-d)^p \left| \begin{matrix} \frac{n-3}{2}, \dots, \frac{n-3}{2} \\ \frac{n-4}{2}, \dots, \frac{n-(p+2)}{2} \end{matrix} \right. \right], \quad (17)$$

$0 \leq d \leq 1$

b) For the two-component case ($p = 2$), we have:

$$f(d; n) = 2(1-d^2)^{(n-4)/2} / B(1/2, (n-2)/2), \quad 0 \leq d \leq 1 \quad (18)$$

2) For a *non-fully independent* two-component binormal system ($\rho \neq 0$), for $0 \leq d \leq 1$:

$$f(d; n) = A \cdot (1-d^2)^{(n-4)/2} \left\{ (1-\rho d)^{-\binom{n-3}{2}} {}_2F_1\left(1/2, 1/2; (2n-1)/2, (1+\rho d)/2\right) + (1+\rho d)^{-\binom{n-3}{2}} {}_2F_1\left(1/2, 1/2; (2n-1)/2, (1-\rho d)/2\right) \right\} \quad (19)$$

where $A = \frac{(n-2)\Gamma(n-1)(1-\rho^2)^{(n-1)/2}}{\Gamma(n-(1/2))(2\pi)^{1/2}}$ and ${}_2F_1(\alpha, \beta; \lambda; x)$ is Gauss hypergeometric function.

PROOF: a) For $\rho_{i,j} = 0$, the density of $d(\mathcal{G})$, as given by (17), is obtained from (11) by a change of variable. **Figure 4** gives the density of the sample coefficient of *inner dependence* $d(\mathcal{G})$, for $n = 10$ and $p = 4$.

Expression (18) is obtained from (15) by the change of variable $d = |\rho|$. Again, the density of d , as given by (19), can be derived from (16) by considering the same change of variable. **QED.**

Numerical computations give $E(d) = 0.17665$. Estimation of $\delta(\mathcal{G})$ from $d(\mathcal{G})$ now follows the same principles as ρ from r .

Figure 5 gives the density $f(d)$, as given by (19), for $n = 8, p = 2, \rho = 0.25$.

6. Conclusion

In this article we have established an original expression for the density of the correlation matrix, with the sample variances as parameters, in the case of the multivariate normal population with non-identity population correlation matrix. We have, furthermore, established the expression of the distribution of the determinant of that random matrix in the case of identity population correlation matrix, and computed its value. Applications are

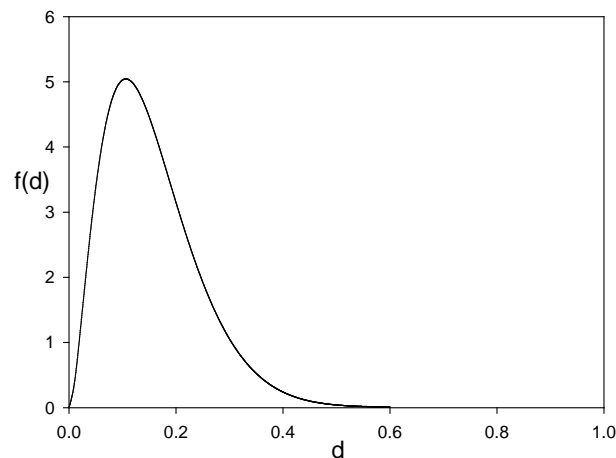


Figure 4. Density (17) for sample coefficient of *inner dependence* $d(\vartheta)$ (normal system with four components, $n = 10$, $p = 4$).

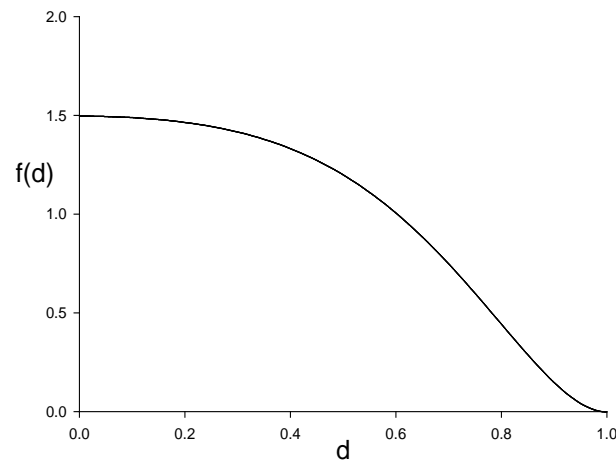


Figure 5. Density (19) of sample measure of a binormal system dependence, $d(\vartheta)$.

made to the dependence among p components of a system. Also, expressions for the densities of a sample measure of a system *inner dependence* are established.

References

- [1] Johnson, D. (1998) Applied Multivariate Methods for Data Analysis. Duxbury Press, Pacific Grove.
- [2] Fisher, R.A. (1962) The Simultaneous Distribution of Correlation Coefficients. *Sankhya, Series A*, **24**, 1-8.
- [3] Pham-Gia, T. and Turkkan, N. (2011) Distributions of the Ratio: From Random Variables to Random Matrices. *Open Journal of Statistics*, **1**, 93-104. <http://dx.doi.org/10.4236/ojs.2011.12011>
- [4] Hotelling, H. (1953) New Light on the Correlation Coefficient and Its Transform. *Journal of the Royal Statistical Society: Series B*, **15**, 193.
- [5] Olkin, I. and Pratt, J.W. (1958) Unbiased Estimation of Certain Correlation Coefficients. *Annals of Mathematical Statistics*, **29**, 201-211. <http://dx.doi.org/10.1214/aoms/1177706717>
- [6] Joarder, A.H. and Ali, M.M. (1992) Distribution of the Correlation Matrix for a Class of Elliptical Models. *Communications in Statistics—Theory and Methods*, **21**, 1953-1964. <http://dx.doi.org/10.1080/03610929208830890>
- [7] Ali, M.M., Fraser, D.A.S. and Lee, Y.S. (1970) Distribution of the Correlation Matrix. *Journal of Statistical Research*,

- 4, 1-15.
- [8] Fraser, D.A.S. (1968) *The Structure of Inference*. Wiley, New York.
- [9] Schott, J. (1997) *Matrix Analysis for Statisticians*. Wiley, New York.
- [10] Kollo, T. and Ruul, K. (2003) Approximations to the Distribution of the Sample Correlation Matrix. *Journal of Multivariate Analysis*, **85**, 318-334. [http://dx.doi.org/10.1016/S0047-259X\(02\)00037-4](http://dx.doi.org/10.1016/S0047-259X(02)00037-4)
- [11] Farrell, R. (1985) *Multivariate Calculation*. Springer, New York. <http://dx.doi.org/10.1007/978-1-4613-8528-8>
- [12] Gupta, A.K. and Nagar, D.K. (2000) *Matrix Variate Distribution*. Hall/CRC, Boca Raton.
- [13] Mathai, A.M. and Haubold, P. (2008) *Special Functions for Applied Scientists*. Springer, New York. <http://dx.doi.org/10.1007/978-0-387-75894-7>
- [14] Kshirsagar, A. (1972) *Multivariate Analysis*. Marcel Dekker, New York.
- [15] Muirhead, R.J. (1982) *Aspects of Multivariate Statistical Theory*. Wiley, New York. <http://dx.doi.org/10.1002/9780470316559>
- [16] Fisher, R.A. (1915) The Frequency Distribution of the Correlation Coefficient in Samples from an Indefinitely Large Population. *Biometrika*, **10**, 507-521.
- [17] Stuart, A. and Ord, K. (1987) *Kendall's Advanced Theory of Statistics*, Vol. 1. 5th Edition, Oxford University Press, New York.
- [18] Sawkins, D.T. (1944) Simple Regression and Correlation. *Journal and Proceedings of the Royal Society of New South Wales*, **77**, 85-95.
- [19] Pham-Gia, T., Turkkan, N. and Vovan, T. (2008) Statistical Discriminant Analysis Using the Maximum Function. *Communications in Statistics-Simulation and Computation*, **37**, 320-336. <http://dx.doi.org/10.1080/03610910701790475>
- [20] Nagar, D.K. and Castaneda, M.E. (2002) Distribution of Correlation Coefficient under Mixture Normal Model. *Metrika*, **55**, 183-190. <http://dx.doi.org/10.1007/s001840100139>
- [21] Gupta, A.K. and Nagar, D.K. (2004) Distribution of the Determinant of the Sample Correlation Matrix from a Mixture Normal Model. *Random Operators and Stochastic Equations*, **12**, 193-199.
- [22] Springer, M. (1984) *The Algebra of Random Variables*. Wiley, New York.
- [23] Pham-Gia, T. (2008) Exact Distribution of the Generalized Wilks's Statistic and Applications. *Journal of Multivariate Analysis*, **99**, 1698-1716. <http://dx.doi.org/10.1016/j.jmva.2008.01.021>
- [24] Pham-Gia, T. and Turkkan, N. (2002) Operations on the Generalized F-Variables, and Applications. *Statistics*, **36**, 195-209. <http://dx.doi.org/10.1080/02331880212855>
- [25] Pitman, E.J.G. (1937) Significance Tests Which May Be Applied to Samples from Any Population, II, the Correlation Coefficient Test. *Supplement to the Journal of the Royal Statistical Society*, **4**, 225-232. <http://dx.doi.org/10.2307/2983647>
- [26] Rencher, A.C. (1998) *Multivariate Statistical Inference and Applications*. Wiley, New York.
- [27] Jogdev, K. (1982) Concepts of Dependence. In: Johnson, N. and Kotz, S., Eds., *Encyclopedia of Statistics*, Vol. 2, Wiley, New York, 324-334.
- [28] Joe, H. (1997) *Multivariate Models and Dependence Concepts*. Chapman and Hall, London. <http://dx.doi.org/10.1201/b13150>
- [29] Bertail, P., Doukhan, P. and Soulier, P. (2006) Dependence in Probability and Statistics. *Springer Lecture Notes on Statistics* 187, Springer, New York. <http://dx.doi.org/10.1007/0-387-36062-X>
- [30] Drouet Mari, D. and Kotz, S. (2004) *Correlation and Dependence*. Imperial College Press, London.
- [31] Lancaster, H.O. (1982) Measures and Indices of Dependence. In: Johnson, N. and Kotz, S., Eds., *Encyclopedia of Statistics*, Vol. 2, Wiley, New York, 334-339.
- [32] Hoyland, A. and Rausand, M. (1994) *System Reliability Theory*. Wiley, New York.
- [33] Bekker, A., Roux, J.J.J. and Pham-Gia, T. (2005) Operations on the Matrix Beta Type I and Applications. Unpublished Manuscript, University of Pretoria, Pretoria.

Scientific Research Publishing (SCIRP) is one of the largest Open Access journal publishers. It is currently publishing more than 200 open access, online, peer-reviewed journals covering a wide range of academic disciplines. SCIRP serves the worldwide academic communities and contributes to the progress and application of science with its publication.

Other selected journals from SCIRP are listed as below. Submit your manuscript to us via either submit@scirp.org or [Online Submission Portal](#).

