

Trace of the Wishart Matrix and Applications

T. Pham-Gia^{1*}, Dinh N. Thanh², Duong T. Phong³

¹Université de Moncton, Moncton, Canada

²University of Science, Hochiminh City, Vietnam

³Ton Duc Thang University, Hochiminh City, Vietnam

Email: *thu.pham-gia@umoncton.ca, dnthanh@hcmus.edu.vn, dtphong@itam.tdt.edu.vn

Received 19 January 2015; accepted 22 April 2015; published 27 April 2015

Copyright © 2015 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

The trace of a Wishart matrix, either central or non-central, has important roles in various multivariate statistical questions. We review several expressions of its distribution given in the literature, establish some new results and provide a discussion on computing methods on the distribution of the ratio: the largest eigenvalue to trace.

Keywords

Trace, Wishart Matrix, Sphericity, Latent Roots, A-Hypergeometric Functions, Humbert Function

1. Introduction

Let $\mathbf{X} \sim N_p(\boldsymbol{\mu}, \boldsymbol{\Sigma})$ be a normal vector and $\{\mathbf{X}_1, \dots, \mathbf{X}_N\}$ be a sample taken from this multivariate normal population. Classical results show that the sample mean vector $\bar{\mathbf{X}} = \frac{1}{N} \sum_{i=1}^N \mathbf{X}_i$ is independent of the sample covariance matrix \mathbf{S} , where

$$n\mathbf{S} = \mathbf{V} = \sum_{j=1}^N (\mathbf{X}_j - \bar{\mathbf{X}})(\mathbf{X}_j - \bar{\mathbf{X}})^T, \quad n = N - 1 \geq p,$$

and they are distributed respectively, as a normal vector $\bar{\mathbf{X}} \sim N_p(\boldsymbol{\mu}, \boldsymbol{\Sigma}/N)$ and a central Wishart matrix $\mathbf{V} \sim W_p(n, \boldsymbol{\Sigma})$, with n degrees of freedom and covariance matrix $\boldsymbol{\Sigma}$, with density

*Corresponding author.

$$f(\mathbf{V}) = \frac{|\mathbf{V}|^{\frac{n-p-1}{2}} \operatorname{etr}\left(-\frac{1}{2}\boldsymbol{\Sigma}^{-1}\mathbf{V}\right)}{(2\pi)^{np/2} \Gamma_p(n/2) |\boldsymbol{\Sigma}|^{n/2}}, \quad \mathbf{V} > \mathbf{0}, \quad (1)$$

where $\operatorname{etr}(\mathbf{X}) = e^{\operatorname{tr}(\mathbf{X})}$, $\mathbf{V} > \mathbf{0}$ means that \mathbf{V} is positive definite matrix, and

$$\Gamma_p(\alpha) = \pi^{\frac{p(p-1)}{4}} \Gamma(\alpha) \Gamma\left(\alpha - \frac{1}{2}\right) \cdots \Gamma\left(\alpha - \frac{p-1}{2}\right),$$

$$\alpha > \frac{p-1}{2}.$$

If $n < p$, the distribution is singular and no density exists. The case of pseudo-Wishart matrices will not be considered in detail here.

Several important results in Multivariate analysis are associated with either the determinant, trace or the eigenvalues of this matrix.

For $\mathbf{U} = \sum_{j=1}^N \mathbf{X}_j \mathbf{X}_j^T$, we have the non-central Wishart $\mathbf{U} \sim W_p(N, \boldsymbol{\Sigma}, \boldsymbol{\Theta})$, with non-centrality parameter

$\boldsymbol{\Theta} = N\boldsymbol{\Sigma}^{-1}\boldsymbol{\mu}\boldsymbol{\mu}^T$, which has the more complicated density expression:

$$f(\mathbf{U}) = \frac{|\mathbf{U}|^{\frac{N-p-1}{2}} \operatorname{etr}\left(-\frac{1}{2}\boldsymbol{\Sigma}^{-1}\mathbf{U}\right)}{(2\pi)^{Np/2} \Gamma_p(N/2) |\boldsymbol{\Sigma}|^{N/2}} \cdot \operatorname{etr}\left(-\frac{1}{2}\boldsymbol{\Theta}\right) {}_0F_1\left(\frac{N}{2}; \frac{1}{4}\boldsymbol{\Theta}\boldsymbol{\Sigma}^{-1}\mathbf{U}\right), \quad (2)$$

where $\mathbf{U} > \mathbf{0}$, and ${}_0F_1(\cdot)$ is the hypergeometric function with one matrix argument, and reduces to (1) when $\boldsymbol{\Theta} = \mathbf{0}$.

We can also have the matrix \mathbf{X} formed by the n column vectors $\mathbf{X}_j \sim N_p(\boldsymbol{\mu}, \boldsymbol{\Sigma})$, and consider the product matrix $\mathbf{X}\mathbf{X}^T$. We have then $\mathbf{X}\mathbf{X}^T \sim W_p(n, \boldsymbol{\Sigma}, \boldsymbol{\Theta})$, where $\boldsymbol{\Theta} = n\boldsymbol{\Sigma}^{-1}\boldsymbol{\mu}\boldsymbol{\mu}^T$. More general is the case where \mathbf{X}_j are independent observations from $N_p(\boldsymbol{\mu}_j, \boldsymbol{\Sigma})$, $1 \leq j \leq n$, with different values for $\boldsymbol{\mu}_j$. We then form the $p \times n$ matrix $\boldsymbol{\mu} = [\boldsymbol{\mu}_1 \boldsymbol{\mu}_2 \cdots \boldsymbol{\mu}_n]$ and we again have:

$$\mathbf{X}\mathbf{X}^T \sim W_p(n, \boldsymbol{\Sigma}, \boldsymbol{\Theta}), \quad \text{where } \boldsymbol{\Theta} = \boldsymbol{\Sigma}^{-1}\boldsymbol{\mu}\boldsymbol{\mu}^T.$$

If we consider at the start the $p \times n$ rectangular matrix as variate, *i.e.* $\mathbf{X} \sim N_{p \times n}(\mathbf{M}, \boldsymbol{\Sigma} \otimes \mathbf{I}_n)$, $n \geq p$, the product matrix $\mathbf{X}\mathbf{X}^T$ is Wishart, *i.e.* $\mathbf{X}\mathbf{X}^T \sim W_p(n, \boldsymbol{\Sigma}, \boldsymbol{\Theta})$, where $\boldsymbol{\Theta} = \boldsymbol{\Sigma}^{-1}\mathbf{M}\mathbf{M}^T$, central if $\mathbf{M} = \mathbf{0}$ and $\boldsymbol{\Theta} = \mathbf{0}$, or non-central with density (2) otherwise.

We wish to avoid too technical results in this article, that could digress us from the real purpose of this survey-type article, which is to gather results on the distribution of the trace of the Wishart matrix, that are still scattered in the literature. But several new research results related to this trace, are also presented. In most cases, we will present both the central and the non-central cases, or the null and non-null distributions of a test criterion. It is also natural that we will encounter zonal polynomials, the values of which are not completely known. Finally, due to the extremely complicated mathematical expressions of certain results we will refer the reader to the original publications when this approach appears to be more convenient.

The non-central Wishart distribution has an important role in theoretical Multivariate analysis, but recently has also found some applications, for example in Image Processing [1].

The Wishart distribution has been generalized in several directions and the most general extension of the Wishart is made by Díaz-García and Gutiérrez-Jáimez [2] to which we refer the reader for additional details. Concerning the product of several positive common random univariate random variables, or the ratio of two positive random variates, H-function, or G-function distributions [3] will be used but we will not discuss the best technique to compute the values of these functions by the residue theorem, since this challenging mathematical problem is already an important topic in itself. Maple and Mathematica can deal with fairly complex cases.

In Section 2 we will first recall several special functions that will be used later. In Section 3 we consider the central Wishart distribution and its trace. Similar results are established for the non-central Wishart and its trace in Section 4. Section 5 studies the moments of the trace while Section 6 considers the Wishartness of some quadratic forms. Section 7 considers the sphericity problem where the trace of the Wishart matrix has an important

role. Finally, Section 8 considers the latent roots and their ratios to the trace and shows the need of further research in this area. It also proposes the simulation approach that has proven to be very effective in some of our previous works.

2. Some Special Functions

2.1. Special Functions

Table 1 gives all the probability densities treated here.

Advanced statistics make use of several special functions and integral transforms: the Humbert function of the second kind and the Lauricella D-function. They are both defined as infinite series, and extended by analytic continuation and are related to each other. We define:

Table 1. Table of densities.

| RANDOM VARIABLE X | RANDOM MATRIX Y |
|---|---|
| <p>Central chi-square $X \sim \chi^2_\nu$</p> $f(\nu; x) = \frac{x^{\nu/2-1} e^{-x/2}}{2^{\nu/2} \Gamma(\nu/2)}, \quad \nu > 0, \quad x \geq 0$ <p>or as a Bessel function.</p> | <p>Central Wishart $Y \sim W_m(n, \Sigma)$</p> $f(n, \Sigma; Y) = \frac{ Y ^{n-m-1} \text{etr}\left(-\frac{\Sigma^{-1}Y}{2}\right)}{2^{np/2} \Gamma_m\left(\frac{n}{2}\right) \Sigma ^{p/2}}, \quad Y > 0$ |
| <p>Non-central Chi-square $X \sim \chi^2_p(\tau^2)$, $p, \tau^2 > 0$</p> $f(p, \tau^2; x) = \frac{e^{-\frac{\tau^2+x}{2}} \left(\frac{\tau^2+x}{2}\right)^{\frac{p}{2}-1} \sum_{k=0}^{\infty} \left(\frac{\tau^2}{4}\right)^k \frac{x^k}{k! \Gamma\left(\frac{p}{2} + k\right)}, \quad x \geq 0.$ <p>τ^2 : Non-centrality parameter, or as a Bessel function.</p> <p>It has MGF $\varphi(t) = \frac{\exp\left(\frac{\tau^2 t}{1-2t}\right)}{(1-2t)^{p/2}}$.</p> | <p>Non-central Wishart $Y \sim W_m(n, \Sigma, \Theta)$</p> $f(n, \Sigma, \Theta; Y) = \frac{ Y ^{n-m-1} \text{etr}\left(-\frac{\Theta - \Sigma^{-1}Y}{2}\right) {}_0F_1\left(\frac{n}{2}; \frac{1}{4} \Theta \Sigma^{-1} Y\right)}{2^{np/2} \Gamma_m\left(\frac{n}{2}\right) \Sigma ^{p/2}}$ <p>$Y > 0$</p> |
| <p>Central Gamma $X \sim \Gamma(a, \lambda)$</p> $f(a, \lambda; x) = \frac{x^{a-1} e^{-x/\lambda}}{\lambda^a \Gamma(a)}, \quad a, \lambda > 0, x \geq 0.$ | <p>Central Gamma $Y \sim \Gamma_m(a, \Lambda)$</p> $f(a, \Lambda; Y) = \frac{\text{etr}(-\Lambda^{-1}Y) Y ^{a-\frac{m+1}{2}}}{\Gamma_m(a) \Lambda ^a}, \quad Y > 0, \quad a > \frac{m+1}{2}$ |
| <p>Randomized Gamma $X \sim RG(\rho, \mu)$</p> $f(\rho, \mu; t) = e^{-(\mu+x)} \sqrt{\left(\frac{x}{\mu}\right)^\rho} I_\rho(2\sqrt{\mu x}), \quad x \geq 0$ <p>where $I_\nu(z) = \sum_{m=0}^{\infty} \frac{1}{m! \Gamma(m+\nu+1)} \left(\frac{z}{2}\right)^{2m+\nu}$</p> | <p>Non-central Gamma $Y \sim \Gamma_m(a, \Lambda, K)$</p> <p>$K$: Non-centrality parameter matrix.</p> $f(a, \Lambda, K; Y) = \frac{\text{etr}(-K - \Lambda^{-1}Y) Y ^{a-\frac{m+1}{2}} {}_0F_1(a, K\Lambda^{-1}Y)}{\Gamma_m(a) \Lambda ^a}, \quad Y > 0$ |
| <p>Non-central Gamma $X \sim \Gamma(a, \lambda, \gamma)$</p> <p>$\gamma$ = non-centrality parameter</p> $f(a, \lambda, \gamma; x) = e^{-\gamma/\lambda} \sum_{m=0}^{\infty} \frac{(\gamma/\lambda)^m}{m!} \frac{e^{-x/\lambda} x^{a+m-1}}{\lambda^{a+m} \Gamma(a+m)}.$ | |

1) The Lauricella D-function, in $n + 2$ parameters and n scalar variables, by:

$$F_D^{(n)}(c; \mathbf{a}; \mathbf{b}; \mathbf{t}) = \sum_{m_n=0}^{\infty} \cdots \sum_{m_1=0}^{\infty} \frac{(c, m_1 + \cdots + m_n)(a_1, m_1) \cdots (a_n, m_n)}{(b, m_1 + \cdots + m_n)} \cdot \frac{t_1^{m_1}}{m_1!} \cdots \frac{t_n^{m_n}}{m_n!},$$

where

$$\mathbf{a} = (a_1, \dots, a_n), \quad \mathbf{t} = (t_1, \dots, t_n), \quad (\alpha, k) = \frac{\Gamma(\alpha + k)}{\Gamma(\alpha)},$$

and $(\alpha, 0) = 1$, which converges for all values of $|t_i| < 1, i = 1, \dots, n$.

2) Similarly, we define the Humbert function, in $n + 1$ parameters and n scalar variables, by:

$$\Phi_2^{(n)}(\mathbf{a}; \mathbf{b}; \mathbf{t}) = \sum_{m_n=0}^{\infty} \cdots \sum_{m_1=0}^{\infty} \frac{(a_1, m_1) \cdots (a_n, m_n)}{(b, m_1 + \cdots + m_n)} \frac{t_1^{m_1}}{m_1!} \cdots \frac{t_n^{m_n}}{m_n!},$$

which converges for all values of \mathbf{t} .

We have the Dirichlet distribution (in $n + 1$ parameters and n variables), $\mathbf{X} \sim \text{Dir}_n(\mathbf{a}; b)$, with $\mathbf{a} = (a_1, \dots, a_n)$, which also has a key role in multivariate analysis. It has density:

$$f(x_1, \dots, x_n) = \frac{\Gamma\left(b + \sum_{i=1}^n a_i\right)}{\Gamma(b) \prod_{i=1}^n \Gamma(a_i)} \left(1 - \sum_{i=1}^n x_i\right)^{b-1} \prod_{i=1}^n x_i^{a_i-1},$$

where $x_i > 0, \sum_{i=1}^n x_i < 1$.

The $\Phi_2^{(n)}(\cdot)$ function in n variables is, in fact, the Laplace transform of the Dirichlet distribution. We have:

$$\Phi_2^{(n)}(\mathbf{a}; \mathbf{b}; \mathbf{t}) = E(\exp(\langle \mathbf{t}, \mathbf{X} \rangle)),$$

where $\mathbf{t} = (t_1, \dots, t_n), \mathbf{a} = (a_1, \dots, a_n) \in \mathbb{R}^n$ and $\mathbf{X} = (X_1, \dots, X_n) \sim \text{Dir}_n(\mathbf{a}; b)$.

The relation between $\Phi_2^{(n)}(\cdot)$ and $F_D^{(n)}(\cdot)$ is [4]:

$$F_D^{(n)}(c; \mathbf{a}; \mathbf{b}; \mathbf{t}) = \frac{1}{\Gamma(c)} \int_0^{\infty} z^{c-1} e^{-z} \Phi_2^{(n)}(\mathbf{a}; \mathbf{b}; z\mathbf{t}) dz. \tag{3}$$

An extension of $\Phi_2^{(n)}(\cdot)$ to the matrix variates is given by [5] and an application of $\Phi_2^{(n)}(\cdot)$ in renewal processes can be found in [6]. On the other hand, $F_D^{(n)}(c; \mathbf{a}; \mathbf{b}; \mathbf{t})$ has several integral representations, the most interesting one should be Euler's type representation, as an hypergeometric integral in one variable:

$$F_D^{(n)}(c; \mathbf{a}; \mathbf{b}; \mathbf{t}) = \int_0^1 u^{c-1} (1-u)^{b-c-1} (1-t_1 u)^{a_1-1} \cdots (1-t_n u)^{a_n-1} du, \tag{4}$$

also known as Picard's integral for $F_D^{(n)}(\cdot)$.

2.2. Integral Representations

Formulas (3) and (4) above allow us to use several interesting mathematical results related to Hypergeometric integrals, which are the focus of much recent work by Gelfand, Krapalov and Zelevinsky [7], named GKZ integrals. They are also known under the topic of A-Hypergeometric functions [8] during the last thirty years. The various hypergeometric functions in several variables, defined differently according to how variables are summed, and named as Horn, Lauricella, Wright, MacRobert functions etc., can now be integrated into a single approach. The introduction of Grobner basis in their study, by Saito, Sturmfels and Takayama [9], has lead to other important results.

Since some of the results obtained by our research group are highly mathematical we do not reproduce them here but they can be obtained by writing to the third author.

The trace of a square matrix is defined as the sum of its diagonal elements, and is sometimes used to measure

the total variance. So, let $T = \text{tr}(\mathbf{V})$, and its univariate density is under study in this article.

For the central Wishart distribution, we will show in the next two sections that when $\Sigma = \sigma^2 \mathbf{I}_p$, the trace, T , is a central Chi-square variable.

3. Central Wishart Distribution

3.1. Two Cases for Σ

Essentially there are two cases:

1) The matrix sigma is diagonal, $\Sigma = \sigma^2 \mathbf{I}_p$: There are several ways to determine the distribution of T :

Bartlett's classical decomposition of the Wishart matrix, $(1/\sigma^2)T \sim \chi_{np}^2$, is as follows: Let $\mathbf{V} = \mathbf{W}^T \mathbf{W}$ where $\mathbf{W} = (t_{ij})$ is upper-triangular $p \times p$ matrix with positive diagonal elements. Then the elements, t_{ij} , $1 \leq i \leq j \leq p$, are all independent, with the diagonal elements t_{ii}^2 being $(1/\sigma^2)t_{ii}^2 \sim \chi_{n-i+1}^2$, $i = 1, \dots, p$, while the off-diagonal elements t_{ij} being

$$(1/\sigma)t_{ij} \sim N(0,1).$$

Since we have $(1/\sigma^2)T = (1/\sigma^2) \sum_{i \leq j}^p t_{ij}^2$, the diagonal elements will give a chi square with $[n + (n+1-p)]p/2$ degree of freedom, while the off-diagonal give a chi square with $p(p-1)/2$ degree of freedom. Adding them together we then have a chi square with np degree of freedom, i.e. $(1/\sigma^2)T \sim \chi_{np}^2$.

Another approach: Consists in considering the latent roots on the diagonal matrix $\mathbf{\Omega}$ equivalent to \mathbf{V} , $\mathbf{\Omega} = \mathbf{B}\mathbf{V}\mathbf{B}^T \sim W_p(n, \sigma^2 \mathbf{I}_p)$, with \mathbf{B} is orthogonal matrix. These latent roots are ω_i , $i = 1, 2, \dots, p$ being independent, with $(1/\sigma^2)\omega_i$ being a chi square with n degree of freedom.

REMARK: In the more general case when $\Sigma = \text{Diag}(\sigma_{11}, \sigma_{22}, \dots, \sigma_{pp})$, then $\frac{v_{ii}}{\sigma_{ii}} \sim \chi_n^2$. The trace is then a linear combination of independent central Chi-squares, each with n degree of freedom.

PROPOSITION 1. Let T_i , $i = 1, \dots, k$ be the traces of k independent Wishart matrices $U_i \sim W_p(n_i, \mathbf{I}_p)$. Then we have $\sum_{i=1}^k T_i \sim \chi_N^2$, with $N = \sum_{i=1}^k n_i$. Furthermore, the product $\prod_{i=1}^k T_i$ and ratios T_i/T_j , $1 \leq i, j \leq k$, can have their densities expressed as G-functions.

PROOF: Immediate from the above results and from [10], where products and ratios of G-functions are presented.

QED.

2) The matrix sigma is not diagonal, $\Sigma \neq \mathbf{I}_p$:

Results are quite complicated for this case since it involves zonal polynomials, whose expressions are only known for simple cases ([11], p. 341).

For $p = 2$, the density is a mixture of gamma distributions, and various expressions of it are available in the statistical literature.

For $p > 2$, T has density

$$f(t) = |\lambda^{-1} \Sigma|^{-n/2} \cdot \sum_{k=0}^{\infty} \frac{1}{k!} g_{pn/2+k, 2\lambda}(t) \sum_{\kappa} (n/2)_{\kappa} C_{\kappa}(\mathbf{I}_p - \lambda \Sigma^{-1}),$$

where $g_{r, 2\lambda}(t) = \frac{t^{r-1} e^{-t/2\lambda}}{(2\lambda)^r \Gamma(r)}$, $t > 0$ and C_{κ} is the zonal polynomial.

We have the usual notations: $0 < \lambda < \infty$ is arbitrary (chosen to be $2\lambda_1 \lambda_p / (\lambda_1 + \lambda_p)$), where λ_1 and λ_p are respectively the largest and the smallest latent roots of matrix Σ , $\kappa = (k_1, k_2, \dots, k_p)$ is a partition of k with $k_1 \geq k_2 \geq \dots \geq k_p \geq 0$, C_{κ} is the zonal polynomial corresponding to κ , $(b)_{\kappa} = \prod_{i=1}^p \left(b - \frac{i-1}{2} \right)_{k_i}$ and $(x)_k = x(x+1) \cdots (x+k-1)$.

3.2. Other Expressions

Mathai and Pillai (1980) give another expression, quite similar:

$$f(t) = \left(2^{np/2} |\Sigma|^{n/2} \Gamma(np/2)\right)^{-1} t^{-1+np/2} e^{-t/(2\eta)} \cdot \sum_{k=0}^{\infty} \frac{\sum_{\kappa} \binom{n/2}{\kappa} C_{\kappa}(\mathbf{I}_p/\eta - \Sigma^{-1})(t/2)^k}{k!(np/2)_{\kappa}}.$$

However, using Mellin Transform methods [12] gives the following density function for T , which avoids the use of zonal polynomials

$$f(t) = \frac{t^{np/2-1} \exp\left(-\frac{t}{2\lambda_1}\right)}{2^{np/2} \Gamma(np/2) \prod_{j=1}^p \lambda_j^{n/2}} \Phi_2^{(p-1)}\left(\frac{n}{2}, \dots, \frac{n}{2}; \frac{np}{2}; t\right), \tag{5}$$

where $t = \left[\left(\frac{1}{\lambda_1} - \frac{1}{\lambda_2}\right) \frac{t}{2}, \dots, \left(\frac{1}{\lambda_1} - \frac{1}{\lambda_p}\right) \frac{t}{2}\right]$, and $\Phi_2^{(p-1)}$ is the Humbert hypergeometric function of the second type mentioned earlier.

4. Non-Central Wishart Distribution

4.1. The Non-Central Chi-Square Distribution

This distribution is present in many aspects of statistics. Its density is given in our table of densities but below is an alternate expression.

Let $I_{\nu}(w)$ be the modified first Bessel function of the 1st kind

$$I_{\nu}(w) = \sum_{m=0}^{\infty} \frac{1}{m! \Gamma(m + \nu + 1)} \left(\frac{w}{2}\right)^{2m + \nu}.$$

The associated Bessel density is:

$$f(\nu, \beta, \theta; x) = \left(\frac{2}{\beta}\right)^{\nu-1} \theta^{\nu} e^{-\frac{\beta^2}{4\theta} x^{(\nu-1)/2}} \exp(-\theta x) I_{\nu-1}(\beta\sqrt{x}). \tag{6}$$

A particular case of (6) is the non-central Chi-square density with n degrees of freedom and non-centrality parameter β , $Y \sim \chi_n^2(\beta)$, obtained when

$$\nu = n/2, \quad \theta = 1/2.$$

Its density is then [13]:

$$f(\beta, n; y) = \left(\frac{1}{2\beta \frac{n-2}{2}}\right) \exp\left(-\frac{\beta^2}{2}\right) \cdot y^{\frac{n}{4} - \frac{1}{2}} \exp\left(-\frac{y}{2}\right) I_{\frac{n}{2}-1}(\beta\sqrt{y}), \tag{7}$$

where $y > 0$.

Using the above functions Laha [14] proved that the reproductive property of the non-central Chi-square, *i.e.* the sum of m independent non-central Chi-square is itself a non-central Chi-square, with parameters β and \mathbf{n} being the corresponding sums of the related parameters.

The density of the product or quotient of two non-central chi square variables can be established in closed form using either Fourier transform [14] or Mellin Transform [13]. Following the latter we have:

PROPOSITION 2. Let X_i , $i = 1, 2$, be two independent non-central Chi-square random variables $\chi_{n_i}^2(\beta_i)$, with densities given by (7). Then the product $P = X_1 X_2$ has as density

$$f(p) = \frac{\exp\left\{-\left(\beta_1^2 + \beta_2^2\right)/2\right\}}{2^{(n_1+n_2)/2-1}} p^{(n_1+n_2)/2-1} \sum_{j=0}^{\infty} \left(\frac{\beta_2}{2}\right)^{2j} p^{j/2} \sum_{i=0}^j \frac{\left(\frac{\beta_1}{\beta_2}\right)^{2i}}{i!(j-i)! \Gamma(n_1/2+i) \Gamma(n_2/2+j-i)}, \tag{8}$$

where $K_\nu(x)$ is the modified Bessel function of the third kind and given by

$$K_\nu(x) = \frac{1}{2} \pi \sin(\nu\pi)^{-1} (I_{-\nu}(x) - I_\nu(x)).$$

While the ratio $R = \frac{X_1}{X_2}$ has as density

$$g(r) = \frac{\exp\left\{-\frac{\beta_1^2 + \beta_2^2}{2}\right\}}{(1+r)^{(n_1+n_2)/2}} \sum_{j=0}^{\infty} \left(\frac{\beta_2}{\sqrt{2}}\right)^{2j} \frac{\Gamma\left(\frac{n_1+n_2}{2} + j\right)}{(1+r)^j} \cdot \sum_{i=0}^j \frac{\left(\frac{\beta_1}{\beta_2}\right)^{2i}}{i!(j-i)!} \frac{r^i}{\Gamma\left(\frac{n_1}{2} + i\right)\Gamma\left(\frac{n_2}{2} + j - i\right)} r^{n_1/2-1}. \quad (9)$$

4.2. The Non-Central Wishart Distribution

As given by (2), its trace T can now be shown to be a non-central Chi-square in some cases.

First, a simple case is the linear non-central case where the non-centrality parameter is concentrated at one component, can be treated as the central case [15]. For a normal vector \mathbf{X} , this will happen when only the first component of $\boldsymbol{\mu}$ is different from 0 and for a normal $p \times m$ matrix \mathbf{X} , when only the first line of M is different from 0. More precisely, let

$$\mathbf{X} \sim W_p(n, \mathbf{I}_p, \boldsymbol{\Theta}),$$

where $\boldsymbol{\Theta} = \text{Diag}(\theta, 0, \dots, 0)$. Then there is a decomposition of $\mathbf{X} = \mathbf{T}\mathbf{T}^T$, where \mathbf{T} is lower triangular with independent elements such that only the first element is a non-central Chi-square, *i.e.*

$$t_{11}^2 \sim \chi_n^2(\theta), \quad t_{ii}^2 \sim \chi_{n-i+1}^2, \quad i = 2, \dots, p,$$

and $t_{ij} \sim N(0, 1)$, $1 \leq j < i \leq p$.

Hence we have

$$T \sim \chi_{np}^2(\theta). \quad (10)$$

The above result on Bessel function distributions now allows us to have the density of sums, product and ratios of the traces of independent linear non-central Wishart distributions. We have the following

THEOREM 1. Let $\mathbf{X}_j \sim W_p(n_j, \mathbf{I}_p, \boldsymbol{\Theta}_j)$, $j = 1, 2$ with $\boldsymbol{\Theta}_j = \text{Diag}(\theta_j, 0, \dots, 0)$ and \mathbf{X}_1 and \mathbf{X}_2 be independent, and let T_1 and T_2 be the two respective traces. The sum $S = T_1 + T_2$ is a non-central Chi-square, while for the product $T_1 T_2$ and ratio T_1/T_2 , their densities can be expressed in closed form, using (8) and (9).

PROOF. Applying (10), we have $T_1 \sim \chi_{n_1 p}^2(\theta_1)$, and similarly for $T_2 \sim \chi_{n_2 p}^2(\theta_2)$.

For $\mathbf{X}_1 + \mathbf{X}_2$, we have

$$\text{tr}(\mathbf{X}_1 + \mathbf{X}_2) = T_1 + T_2 \sim \chi_{(n_1+n_2)p}^2(\theta_1 + \theta_2).$$

For the product, we have $T_1 T_2$ having density given by (8) for its first component, while for the ratio T_1/T_2 , it has density (9) where $\beta_1 = \theta_1$ and $\beta_2 = \theta_2$, also for its first component. We then use the reproductive property of the non-central Chi-square.

QED.

4.3. Numerical Example

We can use (8) and (9) to graph the density of the product and quotient of the two traces. Some computer algebra software, Maple and Mathematica, for example, can do the computation in (8) (9) as an infinite series. But the computation, especially for (8), is very slow. Here, we approximate (8) by taking a large number of terms.

$$\tilde{f}(p; M) = \frac{\exp\left\{-\frac{(\beta_1^2 + \beta_2^2)}{2}\right\}}{2^{(n_1+n_2)/2-1}} p^{(n_1+n_2)/4-1} \sum_{j=0}^M \left(\frac{\beta_2}{2}\right)^{2j} \cdot p^{j/2} \sum_{i=0}^j \frac{\left(\frac{\beta_1}{\beta_2}\right)^{2i}}{i!(j-i)!} \frac{K_{[(n_1+n_2)/2+2i-j]/2}(\sqrt{p})}{\Gamma(n_1/2+i)\Gamma(n_2/2+j-i)},$$

with the value $M = 500$, this approximation seems to be very good. Let

$$\mathbf{X}_1 \sim W_2(2, \mathbf{I}_2, \boldsymbol{\Theta}_1) \text{ and } \mathbf{X}_2 \sim W_2(3, \mathbf{I}_2, \boldsymbol{\Theta}_2),$$

where $\boldsymbol{\Theta}_1 = \begin{pmatrix} 1/15 & 0 \\ 0 & 0 \end{pmatrix}$ and $\boldsymbol{\Theta}_2 = \begin{pmatrix} 1/10 & 0 \\ 0 & 0 \end{pmatrix}$. We get the following graphs of densities of $T_1 = \text{tr}(\mathbf{X}_1)$,

$T_2 = \text{tr}(\mathbf{X}_2)$, $P = T_1 T_2$ and $R = \frac{T_1}{T_2}$, where the horizontal scales are very different (Figure 1).

In the case of planar non-centrality, *i.e.* $\boldsymbol{\Theta} = \text{Diag}(\theta_1, \theta_2, 0, \dots, 0)$, as remarked by Anderson [16], we run into an infinite series of Bessel functions and formulas become very complicated.

4.4. Case $\Sigma \neq \mathbf{I}_p$

We have the following argument, based on the Moment Generating Function (MGF) of T given by [17]:

For $\mathbf{X}_i \sim N_p(\boldsymbol{\mu}_i, \Sigma)$, $i = 1, \dots, n$, we let $\mathbf{X} = [\mathbf{X}_1 \mathbf{X}_2 \dots \mathbf{X}_n]$, $\mathbf{M} = [\boldsymbol{\mu}_1 \boldsymbol{\mu}_2 \dots \boldsymbol{\mu}_n]$, and $\boldsymbol{\Omega} = 1/2 \Sigma^{-1/2} \mathbf{M} \mathbf{M}^T \Sigma^{-1/2}$, we have

$$\mathbf{X} \mathbf{X}^T \sim W_p(n, \Sigma, \boldsymbol{\Theta}).$$

Here we set $\boldsymbol{\Theta} = 2\boldsymbol{\Omega}$, as in [17] where it is shown that the MGF of $T = \text{tr}(\mathbf{X} \mathbf{X}^T)$ is given by

$$M_T(t) = \text{etr}(-\boldsymbol{\Omega}) \prod_{j=1}^p (1 - 2t\lambda_j)^{-n/2} \exp\left\{b_{jj} (1 - 2t\lambda_j)^{-1}\right\},$$

where b_{jj} is the j -th diagonal element of $\mathbf{Q}^T \boldsymbol{\Omega} \mathbf{Q}$, \mathbf{Q} is the orthogonal matrix such that

$$\mathbf{Q}^T \boldsymbol{\Sigma} \mathbf{Q} = \boldsymbol{\Lambda} = \text{Diag}(\lambda_1, \lambda_2, \dots, \lambda_p).$$

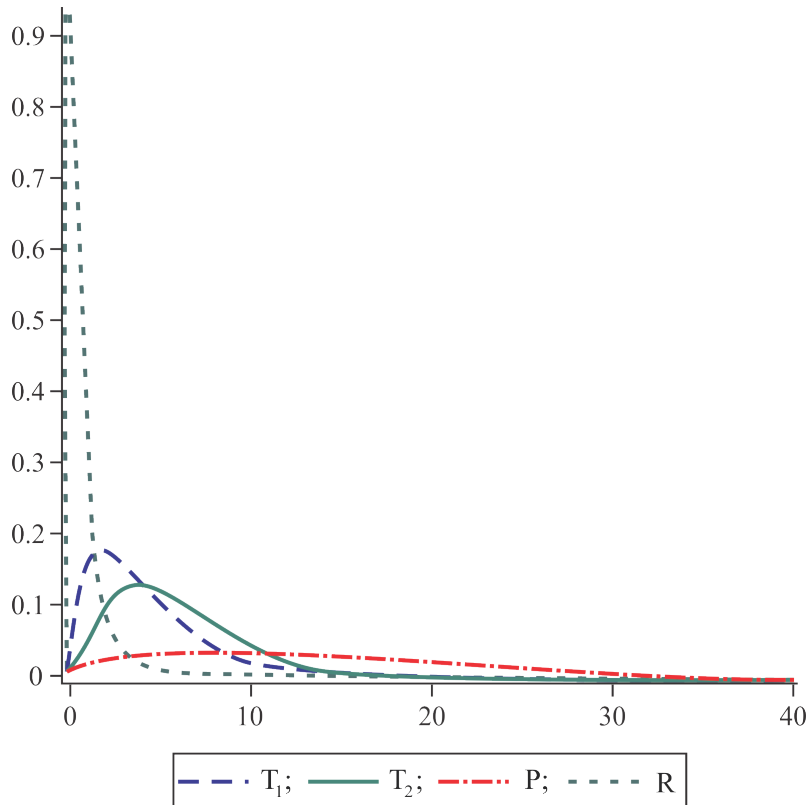


Figure 1. Density of $T_1 = \text{tr}(\mathbf{X}_1)$, $T_2 = \text{tr}(\mathbf{X}_2)$, $P = T_1 T_2$ and $R = T_1/T_2$.

Using $\text{tr}(-\mathbf{\Omega}) = \text{tr}(-\mathbf{Q}^T \mathbf{\Omega} \mathbf{Q}) = -\sum_{j=1}^p b_{jj}$ and, writing $\exp(-b_{jj}) \exp\left\{\frac{b_{jj}}{1-2t\lambda_j}\right\} = \exp\left\{\frac{2b_{jj}\lambda_j t}{1-2\lambda_j t}\right\}$, we have

$$M_T(t) = \prod_{j=1}^p \frac{\exp\left\{(2b_{jj})\lambda_j t(1-2\lambda_j t)^{-1}\right\}}{(1-2t\lambda_j)^{n/2}}.$$

Using the MGF of the Non-Central Chi-square in our table of densities we have the expression of the trace T in terms of a linear combination of non-central Chi square variables with n degrees of freedom each and non-centrality parameter $2b_{jj}$, the j -th diagonalelement of $\mathbf{Q}^T \mathbf{\Omega} \mathbf{Q}$, *i.e.*

$$T = \sum_{j=1}^p \lambda_j Y_j, \tag{11}$$

where $Y_j \sim \chi_n^2(2b_{jj})$, with Y_1, \dots, Y_p independent.

The density of T can be given under a variety of forms by inverting the MGF of T . [12], for example, gives 3 forms (see Section 4.7).

The density of a linear combination of non-central chi-square variables has been the subject of investigation by several authors, since it is associated with quadratic forms in normal variables. Ruben [18], Press [19] and Hartville [20] seemed to be among the first investigators. More recent is the work of Provost and Ruduik [21].

The approach using Laguerre expansions seems promising, as shown by some authors, including Castano-Martinez and Lopez-Blasquez [22]. But all the formulas obtained are quite complicated and we refer the readers to these articles. It should be mentioned that using the same MGF, Kourouklis and Moschopoulos [23] give this density as an infinite combination of gamma densities.

4.5. A Simulation Study

Simulation for the density of the trace of non-central Wishart matrix. Following 4.4, let the covariance matrix be

$$\mathbf{\Sigma} = \begin{pmatrix} 3 & 2 & 0 \\ 2 & 4 & -2 \\ 0 & -2 & 5 \end{pmatrix}, \text{ positive, definite. And the four means be:}$$

$$\boldsymbol{\mu}_1 = (1/2, 2, 1)^T, \quad \boldsymbol{\mu}_2 = (1/3, 1, 2)^T, \quad \boldsymbol{\mu}_3 = (3, 1, 3)^T, \quad \boldsymbol{\mu}_4 = (3, 2, 2)^T.$$

With the above means and covariance matrix, we have:

$$\mathbf{Q} = \frac{1}{3} \begin{pmatrix} -1 & 2 & -2 \\ -2 & 1 & 2 \\ 2 & 2 & 1 \end{pmatrix}, \quad \mathbf{\Lambda} = \begin{pmatrix} 7 & 0 & 0 \\ 0 & 4 & 0 \\ 0 & 0 & 1 \end{pmatrix},$$

$$\mathbf{\Lambda}^{1/2} = \begin{pmatrix} \sqrt{7} & 0 & 0 \\ 0 & 2 & 0 \\ 0 & 0 & 1 \end{pmatrix}, \quad \mathbf{M}\mathbf{M}^T = \begin{pmatrix} 3 & 2 & 1 \\ 2 & 3 & 2 \\ 1 & 2 & 2 \end{pmatrix},$$

and the matrix $\mathbf{Q}^T \mathbf{\Theta} \mathbf{Q}$ is computed using:

$$\mathbf{Q}^T \mathbf{\Theta} \mathbf{Q} = \mathbf{Q}^T 2\mathbf{\Omega} \mathbf{Q} = \mathbf{Q}^T (\mathbf{\Sigma}^{-1/2} \mathbf{M}\mathbf{M}^T \mathbf{\Sigma}^{-1/2}) \mathbf{Q}.$$

$$\mathbf{Q}^T \mathbf{\Theta} \mathbf{Q} = \begin{pmatrix} 0.3020 & -0.5471 & -0.2286 \\ -0.5471 & 10.2809 & 1.4383 \\ -0.2286 & 1.4383 & 3.1235 \end{pmatrix}.$$

We finally have: $\lambda_1 = 7$, $\lambda_2 = 4$, $\lambda_3 = 1$ and $2b_{11} = 0.3020$, $2b_{22} = 10.2809$, $2b_{33} = 3.1235$. Now, two approaches are used to obtain the density of $\text{tr}(\mathbf{X}\mathbf{X}^T)$:

1) Direct approach 1: We use

$$T = \text{tr}(\mathbf{X}\mathbf{X}^T) = \text{tr}([\mathbf{X}_1 \mathbf{X}_2 \mathbf{X}_3, \mathbf{X}_4][\mathbf{X}_1 \mathbf{X}_2 \mathbf{X}_3 \mathbf{X}_4]^T).$$

We use Matlab command `mnrnd`($\boldsymbol{\mu}_i, \boldsymbol{\Sigma}$) to generate 4 normal vectors $\mathbf{X}_i^{(j)} \sim N_3(\boldsymbol{\mu}_i, \boldsymbol{\Sigma})$, $i = 1, \dots, 4$, from which we obtain a value of the trace $T(j) = \text{tr}(\mathbf{X}(j)\mathbf{X}(j)^T)$. Doing this operation 10,000 times we have the density of T given by **Figure 2**.

2) Approach using non-central Chi-squares: We use

$$T = \text{tr}(\mathbf{X}\mathbf{X}^T) = 7Y_1 + 4Y_2 + Y_3,$$

where $Y_1 \sim \chi_4^2(0.3020)$, $Y_2 \sim \chi_4^2(10.2809)$, and $Y_3 \sim \chi_4^2(3.1235)$.

We use Matlab routine `ncx2rnd`(4, ψ) to generate observations $y_i^{(j)}$ from $\chi_4^2(\psi)$, and compute $T(j) = \lambda_1 y_1^{(j)} + \lambda_2 y_2^{(j)} + \lambda_3 y_3^{(j)}$ 10000 times and we have **Figure 3**.

We can see that the two graphs are very close to each other.

4.6. Modified Traces

The influence of $\boldsymbol{\Sigma}$ on T is through the coefficients λ_i . If we remove these coefficients we have the modi-

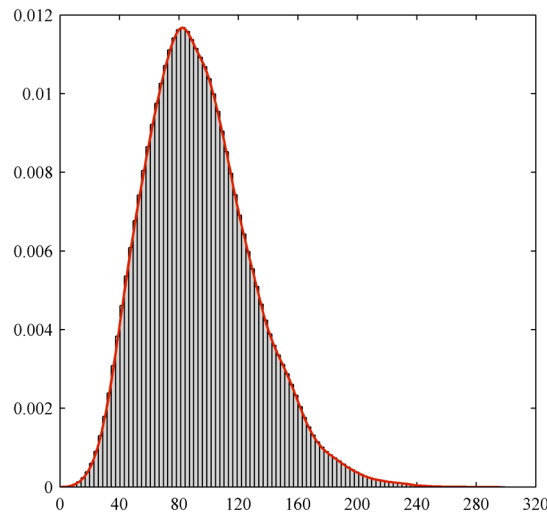


Figure 2. $T = \text{tr}(\mathbf{X}\mathbf{X}^T)$ using Matlab.mndvn.

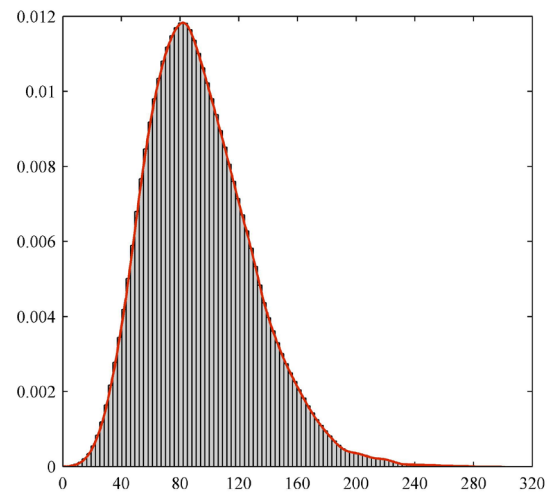


Figure 3. $T = \lambda_1 y_1 + \lambda_2 y_2 + \lambda_3 y_3$ using Matlab.ncx2rnd.

fied trace Z .

PROPOSITION 3: Let $V_j \sim W_p(n_j, \Sigma_j, \Theta_j)$, $j=1,2$ with non-diagonal Θ_j , be independent and T_j , $j=1,2$ be their traces. Let Z_i , $j=1,2$ be the “modified traces” obtained from T_i by taking $\lambda_1 = \lambda_2 = \dots = \lambda_p = 1$. Then the sum $Z_1 + Z_2$, and of the product, $Z_1 Z_2$ and ratio Z_1/Z_2 , can be obtained in closed form.

PROOF: Using (11) with $\Sigma = I_p$ and

$$\lambda_1 = \dots = \lambda_p = 1.$$

Thus, we have

$$Z_j \sim \chi_{n_j p}^2 \left(\sum_{i=1}^p 2b_{ii} \right) = \chi_{n_j p}^2 \left(\text{tr}(\Theta_j) \right), \quad j=1,2,$$

and their sum $Z_1 + Z_2$ is itself a non-central Chi-square, with, as parameters the corresponding sums of individual parameters, *i.e.*

$$Z_1 + Z_2 \sim \chi_{(n_1+n_2)p}^2 \left(\text{tr}(\Theta_1) + \text{tr}(\Theta_2) \right).$$

For the product: $Z_1 Z_2$, we have the same distribution as the product of two non-central Chi-squared random variables, *i.e.* its density is given by (8). Similarly for the ratio, using (9).

QED.

Glueck and Muller [24] also relate the trace of any type of Wishart, singular or nonsingular, central or non-central, true or pseudo, to a weighted sum of non-central Chi-squared random variables and constants. However, the expression of this density is not given, although computational methods are presented, either approximate or permitting to prescribe a degree of accuracy.

4.7. Some Expressions of the Density of T

For some values of the parameters, there can be closed form expression for the density of T . For example, [12] gives this density when n is even (or the sample size $N = n + 1$ is odd). The formula is however complicated, with reference to other works.

When the general case we have an expression similar to (5), but preceded by the non-central factor:

$$f(t) = \frac{\text{etr}(-\Omega) \sum_{\kappa=0}^{\infty} \sum_{\kappa} \left\{ \prod_{i=1}^p \frac{b_{ii}^{k_i}}{k_i!} \right\} t^{m-1} \exp\left(-\frac{t}{2\lambda_1}\right)}{2^m \Gamma(m) \prod_{j=1}^p \lambda_j^{m_j}} \cdot \Phi_2^{(p-1)}(m_2, \dots, m_p; m; t), \quad (12)$$

with $t = \left[\left(\frac{1}{\lambda_1} - \frac{1}{\lambda_2} \right) \frac{t}{2}, \dots, \left(\frac{1}{\lambda_1} - \frac{1}{\lambda_p} \right) \frac{t}{2} \right]$, $0 < t < \infty$, and $f(t) = 0$, $t \leq 0$.

where $m_i = k_i + \frac{n}{2}$, $i=1, \dots, p$, $m = m_1 + \dots + m_p$, $\kappa = (k_1, \dots, k_p)$, $k = \sum_{i=1}^p k_i$.

In terms of zonal polynomials, we have Formula (14) of [12] using common zonal polynomials, or a more compact formula, using Davis expended zonal polynomials $C_{\phi}^{\kappa, \lambda}(X, Y)$.

$$f(t) = \text{etr}(-\Omega) |\Sigma|^{-n/2} \sum_{\kappa, \lambda: \phi}^{\infty} \theta_{\phi}^{\kappa, \lambda}(n/2) \phi C_{\phi}^{\kappa, \lambda}(qI_p - \Sigma^{-1}, \Omega \Sigma^{-1}) \frac{e^{-qt/2} t^{np/2+f-1}}{2^{np/2+f} \Gamma(np/2+f)}. \quad (13)$$

For $t > 0$ and $f(t) = 0$, $t \leq 0$.

5. Moments of the Trace

The trace of V is present in the expressions of several of its moments, and moments are frequently easier to obtain than densities themselves. For example, the r -th cumulant of T , K_r , which is the coefficient of $\frac{t^r}{r!}$ in

the expansion of $\log M_T(t)$, where $M_T(t)$ is the moment generating function of T is found in [17] to be $r!2^r \left\{ \frac{n}{2r} \text{tr}(\Sigma^r) + \text{tr}(\Omega \Sigma^r) \right\}$, which gives the mean of the trace $K_1 = E(T) = 2 \left\{ \frac{n}{2} \text{tr}(\Sigma) + \text{tr}(\Omega \Sigma) \right\}$, a result also found in [25]. Saw [26], and Shah and Khatri [27] proved several other results on moments of the trace of the non-central Wishart.

Some results are unexpected. For example, for $S \sim W_p(n, \Sigma)$, and A is a $p \times p$ constant matrix: $E(SAS) = n\Sigma A^T \Sigma + n \text{tr}(\Sigma A) \Sigma + n^2 \Sigma A \Sigma$, and using zonal polynomials, $E(\text{tr}(S)^k) = 2^k \sum_{\kappa} \binom{n}{2}_{\kappa} C_{\kappa}(\Sigma)$, $k = 0, 1, 2, \dots$ ([15], p. 98 and p. 106).

Several other equalities can be found in the same reference.

Letic and Massam [28], on the other hand, computed the moments of the Wishart matrix V , of the form $E(Q(V))$, where $Q(V)$ is an invariant polynomial in the entries of the matrix V , i.e. depending only on the eigenvalues of V . Finally, there are several results available in the literature on the expectation of $\text{tr}_j(V)$, which is the sum of all principal minors of order j of matrix V . For example, for $V \sim W_p(n, \Sigma)$, we have:

$$E(\text{tr}_p V) = E(|V|) = n^{(p)} |\Sigma| \text{ and } E(\text{tr}_j(V)) = n^{(j)} \text{tr}_j(\Sigma),$$

where $n^{(j)} = n(n-1)\dots(n-j+1)$.

6. The Wishartness of Certain Quadratic Forms

For $X \sim N_p(\mu, \Sigma)$, it is of interest to look for more general quadratic forms that could also be Wishart.

There are 3 cases of Quadratic forms:

- $X^T A X$ is a unidimensional random variable when $X \sim N_p(\mu, \Sigma)$. We are interested at $X^T A X$ being a non-central Chi-square variable.
- $X A X^T$ is a matrix if $X = [X_1 X_2 \dots X_n]$, where $X_i \sim N_p(\mu, \Sigma)$. We are interested in the condition for this matrix to be Wishart.
- Similarly, $X A X^T$ is a random matrix, possibly Wishart when X is a $p \times n$ normal random matrix.

PROPOSITION 4. ([15], p. 256-257) Let $S = X A X^T$, where $X \sim N_{p,n}(M, \Sigma \otimes I_n)$. Then the necessary and sufficient condition for S to be distributed as $W_p(t, \Sigma, \Sigma^{-1} M A M^T)$ is that A is idempotent of rank $t \geq p$. A similar condition applies to $X \sim N_{p,n}(0, \Sigma \otimes I_n)$ to be $W_p(t, \Sigma)$.

The trace T of S in these cases can be studied as previously. We will not elaborate on this point.

7. Sphericity Testing Criterion

In this section we limit ourselves to the vector case, i.e. of $X \sim N_p(\mu, \Sigma)$. Understandably, as seen from what precedes, the case $\Sigma = \sigma^2 I_p$ is important, and this test would permit us to accept, or not, that the matrix is diagonal with same diagonal value.

7.1. Sphericity Test

An interesting property of the Gamma distribution in shape parameter α and scale parameter β , $Ga(\alpha, \beta)$, is that, for a random sample of observations the distribution of the arithmetic mean to the geometric mean is independent of the parameters [29]. An application in telecommunication is given by [30].

Let $X \sim N_p(\mu, \Sigma)$ and let $\{X_1, \dots, X_N\}$ be a sample from this distribution. Let

$$V = \sum_{i=1}^N (X_i - \bar{X})(X_i - \bar{X})^T, \quad n = N - 1.$$

In testing the hypothesis $H_0 : \Sigma = \sigma^2 I_p$ (or the p components are equally variable), called sphericity, we can use:

- 1) The classical likelihood ratio criterion (LRC), $W(n, p)$, first used by Mauchly [31]

$$W(n, p) = \frac{|V|^{1/p}}{[\text{tr}(V)/p]}. \tag{14}$$

The LRC above is hence the ratio of the geometric mean of the eigenvalues to their arithmetic mean. The null-distribution is the distribution of this criterion under H_0 is the density of the product of betas of the first kind,

$$\prod_{j=1}^{p-1} X_j, \tag{15}$$

where $X_j \sim \text{beta}_1^I\left(n - \frac{j}{2}, -\frac{n}{2} + j\left(\frac{1}{2} + \frac{1}{p}\right)\right)$,

as shown by [32]. This product can be shown to have a G-function density, namely

$$f(w) = K(n, p) G_{p-1}^{p-1} \left[w \begin{matrix} \frac{n-2}{2} + \frac{1}{p}, \dots, \frac{n-2}{2} + \frac{p-1}{p} \\ \frac{n-2}{2} - \frac{1}{2}, \dots, \frac{n-2}{2} - \frac{p-1}{2} \end{matrix} \right], \tag{16}$$

where $0 \leq w \leq 1$, and

$$K(n, p) = \prod_{i=1}^{p-1} \frac{\Gamma\left(\frac{n}{2} + \frac{i}{p}\right)}{\Gamma\left(\frac{n-i}{2}\right)}.$$

2) The product of 2 independent beta products [16]:

$$W(n, p) = W_1^* W_2^*,$$

with $W_1^* = \prod_{j=2}^p X_j$ and $W_2^* = p^p \prod_{i=2}^p Y_i^{i-1} (1 - Y_i)$, where $X_i \sim \text{beta}_1^I((n+1-i)/2, (i-1)/2)$ are mutually independent from each other, while $Y_i \sim \text{beta}_1^I(n(i-1)/2, n/2)$ are also mutually independent.

- $\lambda_1^* = (W_1^*)^{n/2}$ is the test criterion for H_1 : The matrix is diagonal, and
 - $\lambda_2^* = (W_2^*)^{n/2}$ is the criterion for H_2 : The diagonal elements are equal.
- Their product is

$$W_1^* W_2^* = p^p \left(\prod_{j=2}^p X_j \right) \prod_{i=2}^p Y_i^{i-1} (1 - Y_i).$$

Since these two tests are in fact independent the product of the two criteria gives the above sphericity test criterion. [32] has adopted a simulation approach to deal with this product.

7.2. Bartlett's Test

In univariate statistics, using p different samples to test that the variances of p independent normal populations are equal, we have Bartlett's test for homogeneity, based on $L^{N/2} = \prod_{i=1}^p \left(\frac{s_i^2}{S^2} \right)^{n_i}$, where $S^2 = \sum_{i=1}^p \frac{n_i}{N} s_i^2$,

with $N = \sum_{i=1}^p n_i$.

When the samples have the same size, Glaser [33] has shown that the null distribution of U^n is a product of independent betas: $\prod_{j=1}^{n-1} X_j$, where

$$X_j \sim \text{beta}_1^I(\alpha, j/n),$$

n is sample size and α is shape parameter of the Gamma distribution. U is the ratio of the geometric mean

to the arithmetic mean.

But when these sizes are different [33] shows that the distribution of Bartlett’s statistic, which now is the adjusted ratio of weighted geometric mean of the sample variances to their weighted arithmetic mean, can be obtained with incomplete beta functions.

Gleser ([34]) considered the two criteria λ_1^* and λ_2^* above and discussed the interesting relationship between Bartlett’s test and the sphericity test, which become equivalent under a change of variables. In the case of non-normality, Hartley’s test or Levene’s test, can be used to the same purpose.

Accepting the hypothesis of sphericity allows us to proceed on to other topics, such as analysis of variance using repeated measures. A generalization of this test to g covariance matrices is possible, and is often known as the Mendoza test.

7.3. Non-Null Distribution

When $\Sigma \neq \sigma^2 \mathbf{I}_p$, we have the non-null distribution of λ .

1) Khatri and Srivastava [35] gives the following expression for the density of the LRC, using both zonal polynomials and Meijer functions:

$$p(\lambda^*) = \frac{\pi^{(p-1)(p+2)/4}}{p^{(np-2)/2} 2^{-(p-1)/2} \Gamma_p(n/2)} (\lambda^*)^{(n-p-1)/2} \cdot \sum_{k=0}^{\infty} \sum_{\kappa} \frac{\Gamma_p((np/2)+k) C_{\kappa}(\mathbf{I}_p - q\Sigma^{-1}) \lambda_1}{p^k k! \sum_{i=1}^p \lambda_i} \cdot \mathbf{G}_{\begin{matrix} p & 0 \\ p & p \end{matrix}} \left[\lambda^* \left| \begin{matrix} \frac{p-1+k}{2} + \frac{k}{p}, \dots, \frac{p-1}{2} + \frac{p+k-1}{p} \\ \frac{p-1}{2} + k_1, \dots, \frac{p-1}{2} + k_p \end{matrix} \right. \right],$$

where C_{κ} is the zonal polynomial associated with κ .

2) [36] propose a convenient strictly numerical method to approximate the power and test size under non-sphericity.

REMARKS. The non-null density in testing diagonality, as given in [37], is a multiple infinite series involving Meijer functions:

$$f(y) = A \sum_{m=0}^{\infty} \sum_{a_1=0}^m \dots \sum_{a_p=0}^m \sum_{j_1=0}^{a_1} \dots \sum_{j_p=0}^{a_p} \frac{L(a_1, \dots, a_p; j_1, \dots, j_p)}{m!} \cdot \mathbf{G}_{\begin{matrix} p+1 & 0 \\ p+1 & p+1 \end{matrix}} \left[y \left| \begin{matrix} \frac{n-1}{2} + a_s - j_s, s=1, \dots, p, \frac{n-1}{2} \\ \frac{n-s}{2}, s=1, \dots, p, \frac{n-1}{2} + m \end{matrix} \right. \right],$$

where A , $L(a_1, \dots, a_p, j_1, \dots, j_p)$ is defined in [37].

Again, here, the computation of the values of this expression is very complicated and we refer the reader to the original paper.

8. Distribution of Ratios of Latent Roots to the Trace

This distribution has attracted renewed interest lately due to its uses in Physics, on random matrices. Krishnaiah and Shurmann [38] were among the first authors to investigate the distributions of these ratios.

The Simulation approach: These distributions have been mentioned by [39] and Johnstone [40] in the context of random matrices. There, the limit theorems are those of Tracy-Widom, or TW, and Wigner. In particular, Nadler reports that, under the hypothesis that $p, n \rightarrow \infty$, with $p/n \rightarrow c$, than an approximate explicit expression of the distribution of this ratio l_1/T , where l_1 is the largest latent root, can be derived, taking into consideration the second derivative of the TW distribution. Computation and simulation methods are used to derive numerical results.

1) Central case:

a) When the matrix sigma is diagonal:

Let $\{\lambda_1 > \lambda_2 > \dots > \lambda_p\}$ be the latent roots of the sample covariance matrix S . The ratio of two latent roots λ_1/λ_p is also called “condition number” in regression and is associated with collinearity. Troskie [41] gives a

very complicated formula based on change of variable technique for the densities of these ratios, which we do not reproduce here. For the ratio of the largest to the trace $\lambda_1/\text{tr}(\mathbf{S})$, this density is not even tractable. However, [42] gives a relation between the exact null-distribution of the j -th largest root and the distribution of the ratio of this root to the trace, but only for $p \leq 2$.

b) When the matrix sigma is not diagonal:

This is even more complex and no result is available on this case. The only resort is by simulation, as in Pham-Gia and Turkkan (2010). Simulation of random matrices, using the appropriate technique, can be very accurate, as shown by several articles by Pham-Gia and Turkkan [39] [43].

2) Non-central case:

This case is naturally more complicated than the previous one and the simulation approach seems to be the only recourse.

Example

We give below the simulation results related to the ratio $\lambda_1/\text{tr}(\mathbf{W})$. Where $\lambda_1 > \lambda_2 > \dots > \lambda_p$ are latent roots of matrix $\mathbf{W} \sim W_p(n, \Sigma)$ with 10,000 generated observations in cases:

1) $p = 4, n = 8, \Sigma = \text{Diag}(2, 1, 3, 4)$.

2) $p = 1000, n = 2p = 2000$, and $\Sigma = \text{Diag}(\lambda_1, \lambda_2, \dots, \lambda_{1000})$, where $\lambda_i = i, i = 1, \dots, 1000$.

The simulation results are given by Figure 4 & Figure 5.

9. Conclusions

We have gathered here several important research results related to the trace of a Wishart matrix, and also indicated some potential research topics. Moreover, we have established several connections among these results and proved a few original results. The two main important applications of the trace are the sphericity test and the distribution of the ratio of a latent root to the trace. The lack of results in the second topic clearly shows that research efforts should be made there, as already pointed out by some researchers. Matrix simulation can clearly supply several useful answers.

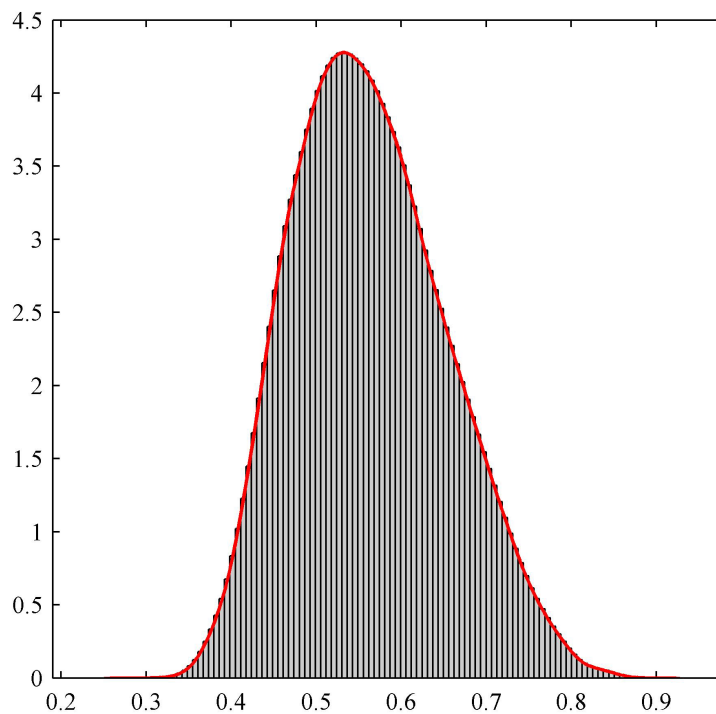


Figure 4. Density function $\frac{\lambda_1}{\text{tr}(\mathbf{W})}$ for $p = 4, n = 2p$.

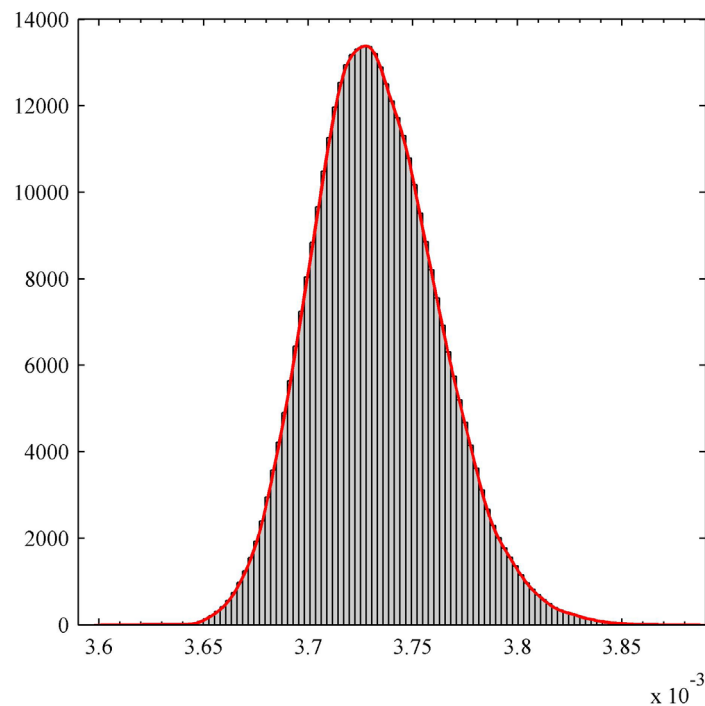


Figure 5. Density function $\frac{\lambda_1}{\text{tr}(\mathbf{W})}$ for $p=1000, n=2p$.

Finally, as shown in our table of densities, the trace can be further investigated by considering the Gamma random matrix, of which the Wishart is only a special case.

References

- [1] Tourneret, J.Y., Ferrari, A. and Letac, G. (2005) The Non-Central Wishart Distribution: Properties and Application to Speckle Imaging. 2005 *IEEE/SP 13th Workshop on Statistical Signal Processing*, Novosibirsk, 17-20 July 2005, 924-929.
- [2] Díaz-García, J.A. and Gutiérrez-Jáimez, R. (2011) On Wishart Distribution: Some Extensions. *Linear Algebra and its Applications*, **435**, 1296-1310.
- [3] Pham-Gia, T. and Turkkan, N.T. (2011) Distributions of Ratios: from Random Variables to Random Matrices. *Open Journal of Statistics*, **1**, 93-104.
- [4] Chamayou, J.F. and Wesolowski, J. (2009) Lauricella and Humbert Functions through Probabilistic Tools. *Integral transforms and Special Functions*, **20**, 529-538. <http://dx.doi.org/10.1080/10652460802645750>
- [5] Mathai, A.M. and Pederzoli, G. (1995) Hypergeometric Functions of Many Matrix Variables and Distributions of Generalized Quadratic Forms. *American Journal of Mathematical and Management Sciences*, **15**, 343-354.
- [6] Pham-Gia, T. and Turkkan, T.N. (1999) System Availability in a Gamma Alternating Renewal Process. *Naval Research Logistics*, **46**, 822-844. [http://dx.doi.org/10.1002/\(SICI\)1520-6750\(199910\)46:7<822::AID-NAV5>3.0.CO;2-D](http://dx.doi.org/10.1002/(SICI)1520-6750(199910)46:7<822::AID-NAV5>3.0.CO;2-D)
- [7] Gelfand, I.M., Kapronov, W.M. and Zelevinsky, A.V. (1991) Hypergeometric Functions, Toric Varieties and Newton Polyhedra. In: Kashiwara, I.M. and Miwa, T., Eds., *ICM 90 Satellite Conference Proceedings*, Springer-Verlag, Tokyo, 104-121.
- [8] Aomoto, K. and Kita, M. (2011) Theory of Hypergeometric Functions. Springer, New York. <http://dx.doi.org/10.1007/978-4-431-53938-4>
- [9] Saito, M., Sturmfels, B. and Takayama, N. (2000) Grobner Deformations of Hypergeometric Differential Equations. Springer, New York. <http://dx.doi.org/10.1007/978-3-662-04112-3>
- [10] 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>
- [11] Muirhead, R. (1982) Aspects of Multivariate Statistical Analysis. Wiley, New York.

- [12] Mathai, A.M. and Pillai, K.C.S. (1982) Further Results on the Trace of a Non-Central Wishart Matrix. *Communications in Statistics—Theory and Methods*, **11**, 1077-1086. <http://dx.doi.org/10.1080/03610928208828294>
- [13] Kotz, S. and Srinivasan, R. (1969) Distribution of Product and Quotient of Bessel Function Variates. *Annals of the Institute of Statistical Mathematics*, **21**, 201-210. <http://dx.doi.org/10.1007/BF02532244>
- [14] Laha, R.G. (1954) On Some Properties of the Bessel Function Distributions. *Bulletin of the Calcutta Mathematical Society*, **46**, 59-71.
- [15] Gupta, A.K. and Nagar, D.K. (2000) Matrix Variate Distributions. Chapman and Hall/CRC, Boca Raton.
- [16] Anderson, T.W. (1982) An Introduction to Multivariate Statistical Analysis. John Wiley and Sons, New York.
- [17] Mathai, A.M. (1980) Moments of the Trace of a Non-Central Wishart Matrix. *Communications in Statistics—Theory and Methods*, **9**, 795-801.
- [18] Ruben, J. (1962) Probability Content of Regions under Spherical Normal Distribution, IV: The Distribution of Homogeneous and Non-Homogeneous Quadratic Functions in Normal Variables. *The Annals of Mathematical Statistics*, **33**, 542-570. <http://dx.doi.org/10.1214/aoms/1177704580>
- [19] Press, S.J. (1966) Linear Combinations of Non-Central Chi-Square Variates. *The Annals of Mathematical Statistics*, **37**, 480-487.
- [20] Hartville, D.A. (1971) On the Distribution of Linear Combinations of Non Central Chi Squares. *The Annals of Mathematical Statistics*, **42**, 809-811.
- [21] Provost, S. and Ruduik, E. (1996) The Exact Distribution of Indefinite Quadratic Forms in Noncentral Normal Vectors. *Annals of the Institute of Statistical Mathematics*, **48**, 381-394. <http://dx.doi.org/10.1007/BF00054797>
- [22] Castano-Martinez, A. and Lopez-Blaquez, F. (2006) Distribution of a Sum of Weighted Non Central Chi-Square Variables. *TEST*, **14**, 397-115.
- [23] Kourouklis, S. and Moschopoulos, P.G. (1985) On the Distribution of the Trace of a Non-Central Wishart Matrix. *Metron*, **43**, 85-92.
- [24] Glueck, D.H. and Muller, K.E. (1998) On the Trace of a Wishart. *Communications in Statistics—Theory and Methods*, **27**, 2137-2141. <http://dx.doi.org/10.1080/03610929808832218>
- [25] De Waal, D.J. (1972) On the Expected Values of the Elementary Symmetric Functions of a Non-Central Wishart Matrix. *The Annals of Mathematical Statistics*, **43**, 344-347.
- [26] Saw, J.G. (1973) Expectation of Elementary Symmetric Functions of a Wishart Matrix. *Annals of Statistics*, **1**, 580-582.
- [27] Shah, B.K. and Khatri, C.G. (1974) Proofs of Conjectures about the Expected Values of the Elementary Symmetric Functions of a Non-Central Wishart Matrix. *Annals of Statistics*, **2**, 833-636.
- [28] Letac, G. and Massam, H. (2004) All Invariant Moments of the Wishart Distribution. *Scandinavian Journal of Statistics*, **31**, 295-318.
- [29] Glaser, R.E. (1976) The Ratio of the Geometric Mean to the Arithmetic Mean from a Random Sample from a Gamma Distribution. *Journal of the American Statistical Association*, **71**, 480-487. <http://dx.doi.org/10.1080/01621459.1976.10480373>
- [30] Cheng, J., Wag, N. and Tellambura, C. (2010) Probability Density Function of Logarithmic Ratio of Arithmetic Mean to Geometric Mean for Nakagami- m Fading Power. *Proceedings of the 25th Biennial Symposium on Communications*, Kingston, 12-14 May 2010, 348-351. <http://dx.doi.org/10.1109/BSC.2010.5472954>
- [31] Mauchly, J.W. (1940) Significance Test for Sphericity of a Normal n -Variate Distribution. *The Annals of Mathematical Statistics*, **11**, 204-209. <http://dx.doi.org/10.1214/aoms/1177731915>
- [32] Pham-Gia, T. and Turkkan, T.N. (2010) Testing Sphericity Using Small Samples. *Statistics*, **44**, 601-616.
- [33] Glaser, R.E. (1980) A Characterization of Bartlett's Statistic Involving Incomplete Beta Functions. *Biometrika*, **67**, 53-58.
- [34] Gleser, L. (1966) A Note on Sphericity Test. *The Annals of Mathematical Statistics*, **37**, 464-467. <http://dx.doi.org/10.1214/aoms/1177699529>
- [35] Khatri, C.G. and Srivastava, M.S. (1971) On Exact Non-Null Distributions of Likelihood Ratio Criteria for Sphericity Test and Equality of Two Covariance Matrices. *Sankhyā*, **33**, 201-206.
- [36] Muller, K.E. and Barton, C.N. (1989) Approximate Power for Repeated-Measures ANOVA Lacking Sphericity. *Journal of the American Statistical Association*, **84**, 549-555.
- [37] Mathai, A.M. and Tan, W.Y. (1977) The Non-Null Distribution of the Likelihood Ratio Criterion for Testing the Hypothesis That the Covariance Matrix Is Diagonal. *Canadian Journal of Statistics*, **5**, 63-74.

- [38] Krishnaiah, P.R. and Shurmann, F.J. (1974) On the Evaluation of Some Distributions That Arise in Simultaneous Tests for the Equality of the Latent Roots of the Covariance Matrix. *Journal of Multivariate Analysis*, **4**, 265-282. [http://dx.doi.org/10.1016/0047-259X\(74\)90033-5](http://dx.doi.org/10.1016/0047-259X(74)90033-5)
- [39] Nadler, B. (2011) On the Distribution of the Ratio of the Largest Eigenvalue to the Trace of a Wishart Matrix. Unpublished Manuscript on the Internet.
- [40] Johnstone, I. (2001) On the Distribution of the Largest Eigenvalue in Principal Components Analysis. *Annals of Statistics*, **29**, 295-327.
- [41] Troskie, C.G. and Conradie, W.J. (1986) The Distribution of the Ratios of the Characteristic Roots (Condition Numbers) and Their Applications in Principal Component or Ridge Regression. *Linear Algebra and Its Applications*, **82**, 255-279. [http://dx.doi.org/10.1016/0024-3795\(86\)90156-4](http://dx.doi.org/10.1016/0024-3795(86)90156-4)
- [42] Davis, A.W. (1972) On the Ratios of the Individual Latent Roots to the Trace of a Wishart Matrix. *Journal of Multivariate Analysis*, **2**, 440-443. [http://dx.doi.org/10.1016/0047-259X\(72\)90037-1](http://dx.doi.org/10.1016/0047-259X(72)90037-1)
- [43] Pham-Gia, T. and Turkkan, T.N. (2009) Testing a Covariance Matrix: Exact Null Distribution of Its Likelihood Criterion. *Journal of Statistical Computation and Simulation*, **79**, 1331-1340.