

Revisit the Wishart Distributionm

Emilly A. Obuya¹, Prakash C. Joshi¹, Thomas A. Gray¹, Thomas C. Keane¹ & Wayne E. Jones Jr.²

¹ ISES Labs, Department of Chemistry and Biochemistry, Russell Sage College, Troy, NY, USA

² Department of Chemistry, State University of New York at Binghamton, Vestal, NY, USA

Correspondence: Thomas C. Keane, Director, ISES Labs, Department of Chemistry and Biochemistry, Russell Sage College, Troy, NY 12180, USA.

Received: March 5, 2020 Accepted: April 23, 2020 Online Published: August 24, 2020

doi:10.5539/ijsp.v9n5p79

URL: <https://doi.org/10.5539/ijsp.v9n5p79>

Abstract

If $S_{p \times p}$ can be written as $S = X'X$, where $X_{n \times p}$ is a data matrix from $N_p(0, V)$, then S is said to have a Wishart distribution with scale matrix V of degree of freedom parameter n . We write $S \sim W_p(V, n)$. When $V = I$, the distribution is said to be in standard form. When $p=1$, the $W_1(\sigma^2, n)$ distribution is found to be $\sum_{i=1}^n x_i^2$, where the elements of x_i are identically independently distributed unit normal variables; being the $\sigma^2 \chi_{(n)}^2$ distribution.

Although Anderson (1984, p248~249) has presented two theorems for the Wishart distribution. In the following we give an alternative proof.

Keywords: Chi-Square distribution, lower triangle matrix, matrix of differentials, orthogonal matrix transformation, skew symmetric matrix, Stiefel manifold, wedge product, Wishart distribution

1. Introduction

It is easily shown that the sample mean vector of n independent samples from the multinormal population $N(\mu, V)$ is also a multinormal random variable with same mean parameter μ and V/n . This is a direct generalization of univariate mean and its implications for constructing significance tests and confidence regions for μ when V is known. In general, any symmetric positive definite matrix S of quadratic and bilinear forms which can be transformed to the sum $\sum_{i=1}^n y_i y_i'$ where p -component column vector y_i are independently distributed according to the distribution $N(0, V)$ is said to have the Wishart distribution. Due to the complexity in its proof, many authors simply omit the derivations with a simple statement like “generalized from bivariate distribution to a p -variate multivariate normal distribution”. For example, Graybill (1961, p200~202) uses the following procedure to derive the Wishart distribution. He starts with a bivariate random sample $(x_{1i}, x_{2i}), i = 1, 2, \dots, n$. By way of the bivariate normal distribution, he defines the joint distribution of $f(b_{11}, b_{12}, b_{22})$. Then he derives the joint moment-generating function of $f(b_{11}, b_{12}, b_{22})$ and shows that it can reach identical characteristics function. In this way, the author claims the result is the desired one, thereby concluding that “generalizes to a p -variate normal distribution and finds the Wishart distribution without giving any proof.” This is a strong motivation for us to revisit the distribution. The Wishart distribution was first derived by Fisher (1915) for $p=2$. Wishart (1928) gave a geometrical derivation of this distribution for general p . Ingram (1933) derived this distribution from its characteristic function. In 2004 Giri gave a more advanced mathematical proof of Wishart distribution. He not only used measure theory to derive the density function but also included the complex variable for the characteristics function. (See p228, theorem 6.4.3).

2. Orthogonal Matrix and Transformatrion

In this section, we define orthogonal matrix, orthogonal transformation, and Stiefel manifold.

Definition 2.1 If U is a $n \times n$ square matrix and

$$UU' = I_{n \times n} \tag{2.1}$$

That is, if

$$U' = U^{-1} \tag{2.2}$$

then U is called an orthogonal matrix. It is known from the theory of determinants that the determinant of the transposed matrix equals the determinant of the given matrix. By comparing the determinants of these matrices on both sides of equality (2.1), we obtain

$$|U||U'| = 1, |U|^2 = 1, |U| = \pm 1.$$

In other words, the determinant of an orthogonal matrix is ± 1 . Taking the transpose on both sides of (2.2), we obtain $(U^{-1})' = (U')' = U = (U^{-1})^{-1}$. So the inverse of an orthogonal matrix is again orthogonal. The product of orthogonal matrix is orthogonal. Since from

$$U' = U^{-1}, V' = V^{-1}$$

we have

$$(UV)' = V'U' = V^{-1}U^{-1} = (UV)^{-1}.$$

Definition 2.2 A linear transformation of the form

$$y_i = \sum_{j=1}^n u_{ij}x_j, i = 1, 2, \dots, n, \text{ such that } \sum_{i=1}^n y_i^2 = \sum_{j=1}^n x_j^2, \text{ where } U = (u_{ij}), ij = 1, \dots, n$$

is an orthogonal matrix. Alternatively, let $U = (u_{ij})$ be the matrix of a linear transformation. Since $U'U = I_{n \times n}$,

$$\sum_{i=1}^n y_i^2 = \underline{y}'I\underline{y} = \underline{x}'U'U\underline{x} = \sum_{i=1}^n x_i^2 \text{ where } \underline{y}' = (y_1, \dots, y_n)$$

and \underline{x}' are row vectors, and \underline{x} , and \underline{y} are column vectors.

Definition 2.3 The set of matrices $M_{pxr} = \{U : U \in R_{pxr}, U'U = I_r\}$. being the set of all orthogonal pxr matrices,

also known as the Stiefel manifold. We next prove a lemma that leads to our main result in section 3.

Lemma 2.1: Let X be a pxn , $n \geq p$, matrix of rank p and let $X = TU$ where T is a real pxp lower triangular matrix with distinct nonzero diagonal elements and U is a unique real $n \times p$ orthogonal matrix,

$U'U = I_p$, all are of functionally independent real variables. Let u_j be the j -th column of U and (du_j) its differential. By ignoring the sign, we have

$$(I) X=TU \text{ implies that } dX = \left\{ \prod_{j=1}^p |t_{jj}|^{n-j} \right\} dT dU, \text{ where}$$

$$dU = \prod_{j=1}^p \prod_{k=j+1}^n u_k' (du_j). \tag{2.3}$$

The surface area of the full Stiefel manifold $M_{p,r}$ or the total integral of the wedge product

$$\int_{j=1}^p \int_{k=j+1}^n u_k'(du_j) \text{ over } M_{p,r}$$

is given by (II)

$$\int_{M_{p,r}} \int_{j=1}^p \int_{k=j+1}^n u_k'(du_j) = \frac{2^p \pi^{pn/2}}{\Gamma_p(\frac{n}{2})}, \tag{2.4}$$

where $\text{Re}(\alpha) > \frac{p-1}{2}$, We used the notation $\int_{j=1}^p \int_{k=j+1}^n$

to mean the wedge product. In geometry, the surface generated by a straight-line intersecting a given line remains parallel to a plane, but not containing it, and

$$\Gamma_p(\alpha) = \int_{X>0} |X|^{\alpha-\frac{p+1}{2}} e^{-tr(X)} dX = \pi^{\frac{p(p-1)}{4}} \Gamma(\alpha)\Gamma(\alpha-\frac{1}{2})\dots\Gamma(\alpha-\frac{p-1}{2}) \tag{2.5}$$

The detailed derivation of equation (2.5) can be found in Mathai(1997, p57) example 1.24.

Proof. The detailed proof of part(I) can be found in

A. M. Mathai p.119, Theorem 2.14, or p 123, Theorem

2.15. We now assume $X = UT$, then $X'X=T'U'UT=T'T$, since $U'U=I$. using the result of part(I), we obtain the following identity:

$$\int_X e^{-tr(X'/X)} dX = \int_T \left\{ \prod_{j=1}^p |t_{jj}|^{n-j} \right\} e^{-\sum_{i \leq j} t_{ij}^2} dT \int_{M_{p,n}} \int_{j=1}^p \int_{k=j+1}^n u_k'(du_j) \tag{2.6}$$

where $-\infty < t_{ij} < \infty, i < j, 0 < t_{jj} < \infty,$

Left side of (2.6) is

$$\int_X e^{-tr(X'/X)} dX = \int_{-\infty}^{\infty} \dots \int_{-\infty}^{\infty} e^{-\sum_{ij} X_{ij}^2} \{ \prod dx_{ij} \} = \pi^{np/2}.$$

While part of the righthand side of (2.6) is given by

$$\int_T \left\{ \prod_{j=1}^p |t_{jj}|^{n-j} \right\} e^{-\sum_{i \leq j} t_{ij}^2} dT = 2^{-p} \Gamma_p(\frac{n}{2}), \text{ for } j=1, \dots, p \tag{2.7}$$

Where T is lower triangle matrix. Using the property

that all elements of T are functionally independent real

variable, equation (2.7) with multiple integral can be reduced to a single integral and these integrals are known as the Chi distribution. Similar situation applies to

normal distribution. They are listed as equations (2.8) and (2.9) below.

$$\int_0^{\infty} |t_{jj}|^{n-j} e^{-t_{jj}^2} dt_{jj} = 2^{-1} \Gamma(\frac{n}{2} - \frac{j-1}{2}), \quad n > j - 1 \tag{2.8}$$

and each of the p(p-1)/2 integrals

$$\int_{-\infty}^{\infty} e^{-t_{ij}^2} dt_{ij} = \sqrt{\pi}, i < j, \tag{2.9}$$

Finally, we derive our needed result,

$$M_{p,r} \int_{\Lambda} \prod_{j=1}^p \prod_{k=j+1}^n u_k' (du_j) = \frac{2^p \pi^{pn/2}}{\Gamma_p(\frac{n}{2})} \tag{2.10}$$

where $\text{Re}(\alpha) > \frac{p-1}{2}$ and $\Gamma_p(\frac{n}{2})$ are defined in equation (2.5).

3. Main Results

Theorem 3.1 Let X be a $pn, n \geq p$, random matrix having an np -variate real Gaussian density. Let $T = (t_{ij})$, with $t_{ij} = 0, i > j$ be a lower triangular matrix. Define U the same as in Lemma 2.1 with $U_1' U_1 = I_p$. Show that the joint density of T is given by,

$$G(T) = \frac{\prod_{j=1}^p t_{jj}^{n-j} \exp(-\frac{1}{2} \text{tr}(V^{-1} T T'))}{\frac{(\frac{n-2}{2})^p}{2} \frac{p(p-1)}{\pi} \frac{n}{4} |V|^{\frac{n}{2}} \prod_{j=1}^p \Gamma(\frac{n-j+1}{2})} \tag{3.1}$$

Proof: The random matrix $X = T U'$ having an np -variate real Gaussian density can be written as

$$f(X) = \frac{\exp(-\frac{1}{2} \text{tr}(V^{-1} X X'))}{(2\pi)^{\frac{np}{2}} |V|^{\frac{n}{2}}}, \text{ where } V = V' > 0 \tag{3.2}$$

Then

$$X X' = T U' U T' = T T' \text{ and } dX = \left\{ \prod_{j=1}^p |t_{jj}|^{n-j} \right\} dT \prod_{j=1}^p \prod_{k=j+1}^n u_k' (du_j) \tag{3.3}$$

The joint density of T and U is given by

$$g(U, T) dT dU = \frac{\exp(-\frac{1}{2} \text{tr}(V^{-1} T T'))}{(2\pi)^{\frac{np}{2}} |V|^{\frac{n}{2}}} \left\{ \prod_{j=1}^p |t_{jj}|^{n-j} \right\} \prod_{j=1}^p \prod_{k=j+1}^n u_k' dT (du_j) \tag{3.4}$$

By integrating U , we find the marginal distribution of T and write its density as

$$g(T) dT = \frac{\left\{ \prod_{j=1}^p |t_{jj}|^{n-j} \right\} \exp(-\frac{1}{2} \text{tr}(V^{-1} T T'))}{(2\pi)^{\frac{np}{2}} |V|^{\frac{n}{2}}} \frac{2^p \pi^{\frac{np}{2}}}{\Gamma_p(\frac{n}{2})}, \tag{3.5}$$

$$= \frac{\left\{ \prod_{j=1}^p |t_{jj}|^{n-j} \right\} \exp(-\frac{1}{2} \text{tr}(V^{-1} T T'))}{\frac{(\frac{n-2}{2})^p}{2} \frac{p(p-1)}{\pi} \frac{n}{4} |V|^{\frac{n}{2}} \prod_{j=1}^p \Gamma(\frac{n-j+1}{2})}$$

Where $\Gamma_p(\frac{n}{2})$ defined in equation (2.5)

Theorem 3.2 Let $S = |TT'| = \prod_{j=1}^p t_{jj}^2$, we wish to find the density of $S = S^>0$, denoted by $H(S)$, as

$$H(S) = C_1 |S|^{\frac{n-p-1}{2}} \text{Exp}(-\frac{1}{2} \text{tr}(V^{-1}S)), \tag{3.6}$$

with $\int_S H(S) dS = 1$ integrating S by using a matrix-variate gamma integral and compare with standard form

$$\Gamma_p(\alpha) = \int_{X>0} |X|^{\alpha-\frac{p+1}{2}} e^{-\text{tr}(X)} dX. \text{ Replace } \alpha = \frac{n}{2}, V=I, \text{ and include } \frac{1}{2}$$

in front of trace, we get $C_1 = \{2^{\frac{np}{2}} \Gamma_p(\frac{n}{2}) |V|^{\frac{n}{2}}\}^{-1}$, The density of $S = |TT'|$ can be written as

$$H(S) = \frac{|S|^{\frac{n-p-1}{2}} \exp(-\frac{1}{2} \text{tr}(V^{-1}S))}{2^{\frac{np}{2}} (\pi)^{\frac{p(p-1)}{4}} |V|^{\frac{n}{2}} \prod_{j=1}^p \Gamma(\frac{n-j+1}{2})}, \tag{3.7}$$

where S is positive definite.

4. Concluding Remarks

In proving Lemma 2.1, we use an important transformation

$$X = TU' \text{ where } T \text{ is lower or upper triangular matrix and } U'$$

is the orthogonal matrix. It is nature to ask “is there a promise that such matrix exists? Or how could we find it.” The book by Ostle and Mensing (1975) entitled “statistics in research” (p196~202) answers this question positively. In early days, Doolittle Method is a very important and useful method. The technique can also help decompose a positive-definite real variables matrix into lower or upper triangle matrix. Finally, we give some numerical examples to demonstrate that it is not difficult to find some orthogonal matrices.

$$\text{when } p=2 \ U = \begin{pmatrix} \cos \theta & \pm \sin \theta \\ \sin \theta & \mp \cos \theta \end{pmatrix}, \text{ when } p=3 \ U = \begin{pmatrix} \sqrt{2}/2 & 0 & \sqrt{2}/2 \\ \sqrt{2}/2 & 0 & -\sqrt{2}/2 \\ 0 & 1 & 0 \end{pmatrix}, \text{ when } p=4 \ U = \frac{1}{6} \begin{pmatrix} 3 & 3 & 3 & 3 \\ 3 & -5 & 1 & 1 \\ 3 & 1 & -5 & 1 \\ 3 & 1 & 1 & -5 \end{pmatrix},$$

It is straight forward to check that $U'U = I_p$ for $p=2,3,4$.

References

Anderson, T. W. (1984). *An Introduction to Multivariate Statistical Analysis* (2nd ed.). John Wiley & Sons, New York.
 Cullen, C. G. (1990). *Matrices and Linear Transformations*. (2nd ed.). Dover Publications, New York.
 Fisher, R. A. (1915). Frequency distribution of the values of the correlation coefficient in samples from an indefinitely large population. *Biometrika*, 10, 507-521. <https://doi.org/10.2307/2331838>
 Giri, N. C. (2004). *Multivariate Statistical Analysis* (2nd ed.). Mercel Dekker, New York.
 Graybill, F. A. (1961). *An Introduction to Linear Statistical Models*. McGraw-Hill Book Company.
 Ingham, A. E. (1933). An integral that occurs in statistics. *Proc. Cambridge phil. Soc.*, 29, 271-276. <https://doi.org/10.1017/S0305004100011075>

Mathai, A. M. (1997). *Jacobians of Matrix Transformations and Functions of Matrix Argument*. World Scientific Publishing Co. <https://doi.org/10.1142/3438>

Ostle, B., & Mensing, R. W. (1975). *Statistics in Research* (3rd ed.). The Iowa State University Press.

Wishart, J. (1928). The generalized product moment distribution from a normal multivariate distribution. *Biometrika* 20A, 32-52. <https://doi.org/10.1093/biomet/20A.1-2.32>

Copyrights

Copyright for this article is retained by the author(s), with first publication rights granted to the journal.

This is an open-access article distributed under the terms and conditions of the Creative Commons Attribution license (<http://creativecommons.org/licenses/by/4.0/>).