

On Some Known Derivations and New Ones for the Wishart Distribution: A Didactic

Haruhiko Ogasawara^[0000–0002–5029–2086]

Otaru University of Commerce, Otaru, Japan
emt-hogasa@emt.otaru-uc.ac.jp

Abstract. The proofs of the probability density function (pdf) of the Wishart distribution tend to be complicated with geometric viewpoints, tedious Jacobians and not self-contained algebra. In this paper, some known proofs and simple new ones for uncorrelated and correlated cases are provided with didactic explanations. For the new derivation of the uncorrelated case, an elementary direct derivation of the distribution of the Bartlett-decomposed matrix is provided. In the derivation of the correlated case from the uncorrelated one, simple methods including a new one are shown.

Keywords: Jacobian · Multivariate normality · Probability density function (pdf) · Triangular matrix · Bartlett decomposition.

1 Introduction

The Wishart distribution has been often used for the matrix of the squares and cross products of random vectors. In multivariate analysis or more specifically structural equation modeling (SEM), a modified log-likelihood of this distribution (see e.g., Ogasawara, 2016, Equation (2.8)) has been used probably as a gold-standard discrepancy function for estimation even under non-normality though the distribution is given under multivariate normality. In SEM, variations of the distribution are also used as priors for covariance matrices (Liu, Qu, Zhang, & Wu, 2022; Zhang, 2021). The distribution has various extensions e.g., the inverted distribution (Anderson, 2003, Section 7.7), singular cases (Bodnar & Okhrin, 2008; Mathai & Provost, 2022; Srivastava, 2003), complex-valued ones (Srivastava & Khatri, 1979, Section 3.7; Mathai, Provost, & Haubold, 2022, Section 5.5), those with two different degrees of freedom (df's) (Ogasawara, 2023b), the joint distributions of the Wishart matrix and normal vectors (Yonenaga, 2022) and cases under arbitrary distributions (Hsu, 1940; Srivastava & Khatri, 1979, Lemma 3.2.3; Olkin, 2002, Section 2).

Asymptotic results associated with the Wishart distribution are also of practical use. In SEM, the asymptotic standard errors of the Wishart maximum likelihood estimators for structural parameters are often used under normality

or non-normality. In this situation, the large df is assumed. When the number of variables is also large under some condition as in high-dimensional data (see e.g., Yao, Zheng, & Bai, 2015), the limiting distribution of the eigenvalues of the Wishart matrix is given by the Marčenko and Pastur (1967, M-P) distribution (the author is indebted to an anonymous reviewer for this point). The M-P distribution gives a tool for the problems of the numbers of factors or components in SEM (Chen & Weng, 2023).

The probability density functions (pdf's) of the Wishart distribution were given by Fisher (1915, p. 510) and Wishart (1928) for the bivariate and general multivariate cases, respectively. The derivations tend to be involved with geometric viewpoints (see e.g., Anderson, 2003, Section 7.2) or not self-contained algebra as criticized by Ghosh and Sinha (2002) (for the references of derivations see Srivastava & Khatri, 1979, p. 73 and Anderson, 2003, pp. 256-257). Khatri (1963) showed a brief derivation using an integral of the unity over the constant quadratic forms having the chi-square density. Ghosh and Sinha (2002) gave a self-contained concise proof of the Wishart density though it is an indirect method. In spite of frequent use of the Wishart density and its variations in SEM, the derivation of the pdf seems to be often intractable for beginning students/researchers. Probably, many of them use the Wishart pdf as if referencing a cook book without understanding the derivation, which is an undesirable situation. A relatively concise derivation is to use the characteristic function and its inversion (Wishart & Bartlett, 1933; Wilks, 1962, Section 18.2). However, this method requires the Fourier integral theorem or Levy's inversion formula, which may be unfamiliar for beginners. In this paper, almost self-contained known proofs and new ones for the uncorrelated and correlated multivariate cases are shown with didactic explanations.

2 Proofs of the Wishart Distributions

2.1 The distribution of a lower-triangular matrix for the Wishart density

Suppose that in the random matrix $\mathbf{X} = \{X_{ij}\}$ ($i = 1, \dots, p; j = 1, \dots, n; p \leq n$), each column is multivariate normally distributed as $N_p(\mathbf{0}, \mathbf{I}_p)$ independent of the other columns with the population mean vector $\mathbf{0}$ and covariance matrix \mathbf{I}_p denoting the $p \times p$ identity matrix. That is, all the elements of \mathbf{X} are mutually independently distributed as standard normal.

Let $\mathbf{S} \equiv \mathbf{X}\mathbf{X}^T = \mathbf{T}\mathbf{T}^T$ be Bartlett-decomposed such that \mathbf{T} is a $p \times p$ lower-triangular matrix whose diagonal elements are positive. Define $\mathbf{s} = (s_{11}, s_{21}, s_{22}, \dots, s_{p1}, \dots, s_{pp})^T$ and $\mathbf{t} = (t_{11}, t_{21}, t_{22}, \dots, t_{p1}, \dots, t_{pp})^T$, where \mathbf{s} and \mathbf{t} are the $\{(p^2 + p)/2\} \times 1$ vectors of the non-duplicated elements of \mathbf{S} and the random elements of \mathbf{T} , respectively. Let $|\partial\mathbf{s}/\partial\mathbf{t}^T|_+$ (Srivastava & Khatri, 1979, p. 28) be the absolute value of the determinant of the Jacobian matrix for the transformation $\mathbf{S} \rightarrow \mathbf{T}$:

$$\frac{\partial\mathbf{s}}{\partial\mathbf{t}^T} = \left\{ \frac{\partial s_{ij}}{\partial t_{kl}} \right\} (p \geq i \geq j \geq 1; p \geq k \geq l \geq 1)$$

using the double subscript notation for the rows of the elements of \mathbf{S} and columns for those of \mathbf{t}^T in $\partial\mathbf{s}/\partial\mathbf{t}^T$. Then, the Jacobian of the transformation is given by $|\partial\mathbf{s}/\partial\mathbf{t}^T|_+$. For the proof of the Wishart distribution, the following lemmas are used.

Lemma 1 *Suppose that each of $2m$ variables X_{ik} and X_{jk} ($i \neq j; k = 1, \dots, m; m = 1, 2, \dots$) independently follows $N(0, 1) \equiv N_1(0, 1)$. Then, the distribution of $\sum_{k=1}^m X_{ik}X_{jk}$ is the same as that of $X_{il}\sqrt{\sum_{k=1}^m X_{jk}^2}$ ($i \neq j; l = 1, \dots, m$).*

Proof. When $m = 1$, the equal distribution of $X_{i1}X_{j1}$ and $X_{i1}\sqrt{X_{j1}^2} = X_{i1}|X_{j1}|$ is given by the symmetric distribution of $X_{i1}X_{j1}$ about zero. For general cases, consider the moment generating functions (mgf's). By definition, the mgf of $\sum_{k=1}^m X_{ik}X_{jk}$ is

$$\begin{aligned} & \mathbb{E} \left\{ \exp \left(t \sum_{k=1}^m X_{ik}X_{jk} \right) \right\} = \prod_{k=1}^m \mathbb{E} \left\{ \exp(tX_{ik}X_{jk}) \right\} \\ & = \prod_{k=1}^m \frac{1}{2\pi} \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} \exp \left(tx_{ik}x_{jk} - \frac{x_{jk}^2}{2} \right) dx_{jk} \exp \left(-\frac{x_{ik}^2}{2} \right) dx_{ik} \\ & = \prod_{k=1}^m \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} \frac{1}{\sqrt{2\pi}} \exp \left\{ -\frac{(x_{jk}-tx_{ik})^2}{2} \right\} dx_{jk} \exp \left\{ -\frac{(1-t^2)x_{ik}^2}{2} \right\} dx_{ik} \\ & = \prod_{k=1}^m \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} \exp \left\{ -\frac{(1-t^2)x_{ik}^2}{2} \right\} dx_{ik} \\ & = (1-t^2)^{-m/2} \quad (|t| < 1). \end{aligned}$$

On the other hand, the mgf of $X_{il}\sqrt{\sum_{k=1}^m X_{jk}^2}$ is

$$\begin{aligned} & \mathbb{E} \exp \left(tX_{il}\sqrt{\sum_{k=1}^m X_{jk}^2} \right) \\ & = \frac{1}{(2\pi)^{(m+1)/2}} \int_{-\infty}^{\infty} \cdots \int_{-\infty}^{\infty} \exp \left(tx_{il}\sqrt{\sum_{k=1}^m x_{jk}^2} - \frac{x_{il}^2}{2} - \frac{\sum_{k=1}^m x_{jk}^2}{2} \right) \\ & \quad \times dx_{il}dx_{j1} \cdots dx_{jm} \\ & = \frac{1}{(2\pi)^{m/2}} \int_{-\infty}^{\infty} \cdots \int_{-\infty}^{\infty} \frac{1}{(2\pi)^{1/2}} \exp \left\{ -\left(x_{il} - t\sqrt{\sum_{k=1}^m x_{jk}^2} \right)^2 / 2 \right\} dx_{il} \\ & \quad \times \exp \left\{ -(1-t^2)\sum_{k=1}^m x_{jk}^2 / 2 \right\} dx_{j1} \cdots dx_{jm} \\ & = \frac{1}{(2\pi)^{m/2}} \int_{-\infty}^{\infty} \cdots \int_{-\infty}^{\infty} \exp \left\{ -(1-t^2)\sum_{k=1}^m x_{jk}^2 / 2 \right\} dx_{j1} \cdots dx_{jm} \\ & = (1-t^2)^{-m/2} \quad (|t| < 1). \end{aligned}$$

It is found that the above two mgf's are the same, which shows the same distribution of $\sum_{k=1}^m X_{ik}X_{jk}$ and $X_{il}\sqrt{\sum_{k=1}^m X_{jk}^2}$ ($i \neq j; l = 1, \dots, m$). \square

The second proof using the pdf of the chi-distribution is given in the supplement to this paper (Ogasawara, 2023a).

Lemma 2 (Deemer & Olkin, 1951, Theorem 4.1; Srivastava & Khatri, 1979, Exercise 1.28 (i); Muirhead, 1982, Theorem 2.1.9; Anderson, 2003, p. 255). *The Jacobian of the transformation $\mathbf{S} \rightarrow \mathbf{T}$ is*

$$|\partial \mathbf{s} / \partial \mathbf{t}^T|_+ = 2^p \prod_{i=1}^p t_{ii}^{p-i+1}.$$

Proof. Deemer and Olkin (1951) derived the result as a special case of another general theorem. Muirhead (1982) used the exterior product while an essential standard proof was given by Anderson (2003). The derivation is given here by induction. When $p = 1$, $|\partial \mathbf{s} / \partial \mathbf{t}^T|_+ = ds_{11} / dt_{11} = dt_{11}^2 / dt_{11} = 2t_{11} > 0$ showing that the above result holds. Assume that the result holds when $p = p^*$ i.e., $|\partial \mathbf{s} / \partial \mathbf{t}^T|_+ = 2^{p^*} \prod_{i=1}^{p^*} t_{ii}^{p^*-i+1}$ ($p^* \geq 1$). When $p = p^* + 1$, the elements $s_{p^*+1,1}, s_{p^*+1,2}, \dots, s_{p^*+1,p^*+1}$ are added to \mathbf{s} at its end. Similarly,

$t_{p^*+1,1}, t_{p^*+1,2}, \dots, t_{p^*+1,p^*+1}$ are added to \mathbf{t}^T . Noting that $s_{ij} = \sum_{k=1}^j t_{ik}t_{jk}$ ($p \geq i \geq j \geq 1$), we find that $\partial \mathbf{s} / \partial \mathbf{t}^T$ is a lower-triangular matrix. Consequently, the added factor in $|\partial \mathbf{s} / \partial \mathbf{t}^T|_+$ when $p = p^* + 1$ over when $p = p^*$ is given by the product of the added diagonal elements:

$$\frac{\partial s_{p^*+1,1}}{\partial t_{p^*+1,1}} \frac{\partial s_{p^*+1,2}}{\partial t_{p^*+1,2}} \dots \frac{\partial s_{p^*+1,p^*}}{\partial t_{p^*+1,p^*}} \frac{\partial s_{p^*+1,p^*+1}}{\partial t_{p^*+1,p^*+1}} = t_{11}t_{22} \dots t_{p^*p^*} 2t_{p^*+1,p^*+1}.$$

That is, $|\partial \mathbf{s} / \partial \mathbf{t}^T|_+$ becomes

$$2^{p^*} \left(\prod_{i=1}^{p^*} t_{ii}^{p^*-i+1} \right) t_{11}t_{22} \dots t_{p^*p^*} 2t_{p^*+1,p^*+1} = 2^{p^*+1} \prod_{i=1}^{p^*+1} t_{ii}^{p^*+1-i+1},$$

which shows that the formula $|\partial \mathbf{s} / \partial \mathbf{t}^T|_+ = 2^p \prod_{i=1}^p t_{ii}^{p-i+1}$ holds when $p = p^* + 1$ indicating the required result. \square

In the following theorem for a known Wishart density, we use $\Gamma_p(n/2) \equiv \pi^{p(p-1)/4} \prod_{i=1}^p \Gamma\{(n-i+1)/2\}$ i.e., the p -variate Gamma function (Anderson, 2003, Definition 7.2.1; Subsection 7.2, Equation (18); see also DLMF, 2021, Section 35.3, <https://dlmf.nist.gov/35.3>), where $\Gamma(k) = \int_0^\infty v^{k-1} \exp(-v) dv$ ($k > 0$) is the usual gamma function.

Theorem 1 *Under the condition that the n columns of \mathbf{X} independently follow $N_p(\mathbf{0}, \mathbf{I}_p)$, the pdf of the Wishart distributed \mathbf{S} is given by*

$$w_p(\mathbf{S} | \mathbf{I}_p, n) = \frac{\exp\{-\text{tr}(\mathbf{S})/2\} |\mathbf{S}|^{(n-p-1)/2}}{2^{np/2} \Gamma_p(n/2)} \quad (n \geq p).$$

Proof. Consider the case of $t_{ij} = X_{ij}$ and $t_{ii} = \sqrt{\sum_{k=i}^n X_{ik}^2}$ ($i = 1, \dots, p; j = 1, \dots, i-1$). Since X_{ij} ($i = 1, \dots, p; j = 1, \dots, n$) are mutually independent, t_{ij} ($i = 1, \dots, p; j = 1, \dots, i$) are independent. Note that $(\mathbf{T}\mathbf{T}^T)_{ii} = \sum_{j=1}^i t_{ij}^2 = (\mathbf{X}\mathbf{X}^T)_{ii}$ ($i = 1, \dots, p$) are independently chi-square distributed with n df, where $(\cdot)_{ij}$ is the (i, j) -th element of a matrix; and t_{ii} is chi-distributed with $n - i + 1$

df. Further, Lemma 1 shows that the distributions of the off-diagonal elements $(\mathbf{T}\mathbf{T}^\mathbf{T})_{ij} = \sum_{k=1}^j t_{ik}t_{jk}$ and $(\mathbf{X}\mathbf{X}^\mathbf{T})_{ij}$ ($p \geq i > j \geq 1$) using t_{jj} and t_{ij} ($i = 1, \dots, p; j = 1, \dots, i-1$) are the same. That is, the distribution of $\mathbf{S} = \mathbf{X}\mathbf{X}^\mathbf{T}$ and $\mathbf{T}\mathbf{T}^\mathbf{T}$ are the same when t_{ij} ($i = 1, \dots, p; j = 1, \dots, i$) are distributed as above. The pdf of the constructed t_{ij} 's ($p \geq i \geq j \geq 1$) denoted by $f_p(\mathbf{T})$ becomes

$$\begin{aligned} f_p(\mathbf{T}) &= \left[\prod_{i=1}^p \frac{t_{ii}^{n-i} \exp(-t_{ii}^2/2)}{2^{\{(n-i+1)/2\}-1} \Gamma\{(n-i+1)/2\}} \right] \\ &\quad \times \frac{1}{(\sqrt{2\pi})^{(p^2-p)/2}} \left\{ \prod_{p \geq i > j \geq 1} \exp(-t_{ij}^2/2) \right\} \\ &= \frac{\left\{ \prod_{i=1}^p t_{ii}^{n-i} \exp(-t_{ii}^2/2) \right\} \left\{ \prod_{p \geq i > j \geq 1} \exp(-t_{ij}^2/2) \right\}}{2^{\frac{(n+1)p}{2} - \frac{p(p+1)}{4} - p} \times 2^{\frac{p(p-1)}{4}} \pi^{\frac{p(p-1)}{4}} \prod_{i=1}^p \Gamma\{(n-i+1)/2\}} \\ &= \frac{\left(\prod_{i=1}^p t_{ii}^{n-i} \right) \exp\{-\text{tr}(\mathbf{T}\mathbf{T}^\mathbf{T})/2\}}{2^{\frac{np}{2} - p} \Gamma_p(n/2)}. \end{aligned}$$

In the above expression, the pdf of the chi-distributed t_{ii} with k df denoted by $f_\chi(t_{ii}|k)$ is given by that of the chi-square distributed $u = t_{ii}^2$ with k df i.e.,

$$f_{\chi^2}(u|k) = \frac{u^{(k/2)-1}}{2^{k/2} \Gamma(k/2)} \exp(-u/2) \text{ with the Jacobian } du/dt_{ii} = 2t_{ii}, \text{ yielding}$$

$$f_\chi(t_{ii}|k) = \frac{u^{(k/2)-1}}{2^{k/2} \Gamma(k/2)} \exp(-u/2) \frac{du}{dt_{ii}} = \frac{t_{ii}^{(n-i+1)-2+1} \exp(-t_{ii}^2/2)}{2^{(n-i+1)/2-1} \Gamma\{(n-i+1)/2\}}$$

as shown earlier, when $u = t_{ii}^2$ and $k = n - i + 1$.

Consider the transformation $\mathbf{T} \rightarrow \mathbf{S}$ in $\mathbf{S} = \mathbf{X}\mathbf{X}^\mathbf{T} = \mathbf{T}\mathbf{T}^\mathbf{T}$. The Jacobian $J(\mathbf{T} \rightarrow \mathbf{S})$ of this transformation is given by the reciprocal of $J(\mathbf{S} \rightarrow \mathbf{T})$ obtained in Lemma 2 as $J(\mathbf{T} \rightarrow \mathbf{S}) = 1/|\partial\mathbf{s}/\partial\mathbf{t}^\mathbf{T}|_+ = \left(2^p \prod_{i=1}^p t_{ii}^{p-i+1}\right)^{-1}$. Consequently, using $|\mathbf{S}|^{1/2} = |\mathbf{T}| = t_{11} \cdots t_{pp}$ the pdf of \mathbf{S} becomes

$$\begin{aligned} w_p(\mathbf{S}|\mathbf{I}_p, n) &= f_p(\mathbf{T})J(\mathbf{T} \rightarrow \mathbf{S}) \\ &= \frac{\left(\prod_{i=1}^p t_{ii}^{n-i}\right) \exp\{-\text{tr}(\mathbf{T}\mathbf{T}^\mathbf{T})/2\}}{2^{\frac{np}{2} - p} \Gamma_p(n/2) 2^p \prod_{i=1}^p t_{ii}^{p-i+1}} = \frac{\exp\{-\text{tr}(\mathbf{S})/2\} |\mathbf{S}|^{(n-p-1)/2}}{2^{np/2} \Gamma_p(n/2)}. \end{aligned}$$

□

Remark 1 The pdf of t_{ij} 's ($p \geq i \geq j \geq 1$) i.e., $f_p(\mathbf{T})$ given above using Lemma 1 is algebraically equal to those of Anderson (2003, Equation (6), p. 253, Corollary 7.2.1), Wijsman (1957, Equation (12)) and Kshirsagar (1959, Remarks). However, a typical derivation by e.g., Anderson is an indirect one using orthogonalization and the conditional normal density. Since Anderson's derivation

seems to give some complicated impressions for beginning students/researchers though it is almost self-contained, the corresponding didactic explanation of his derivation is given below. Anderson (2003, Equation (2), p. 252) defined the n -dimensional independent random vectors $\mathbf{v}_i \sim N_n(\mathbf{0}, \mathbf{I}_n)$ ($i = 1, \dots, p$) with

$$\mathbf{X} = \begin{pmatrix} \mathbf{v}_1^T \\ \vdots \\ \mathbf{v}_p^T \end{pmatrix}.$$

Then, the Gram-Schmidt sequential orthogonalization is employed (Anderson, 2003, Equation (3), p. 253) as

$$\mathbf{w}_i = \mathbf{v}_i - \sum_{j=1}^{i-1} \mathbf{w}_j \frac{\mathbf{w}_j^T \mathbf{v}_i}{\mathbf{w}_j^T \mathbf{w}_j} \quad (i = 2, \dots, p) \quad \text{and} \quad \mathbf{w}_1 = \mathbf{v}_1,$$

where he used the expression $\mathbf{v}_j^T \mathbf{w}_j$ for the denominator $\mathbf{w}_j^T \mathbf{w}_j$. Though $\mathbf{v}_j^T \mathbf{w}_j = \mathbf{w}_j^T \mathbf{w}_j$ ($j = 1, \dots, i$) as will become apparent, $\mathbf{w}_j^T \mathbf{w}_j$ may be more natural and appropriate. While he included the short derivation of the orthogonality among \mathbf{w}_i 's by induction, it is repeated here with some added explanations. When $i = 2$, we have

$$\mathbf{w}_2^T \mathbf{w}_1 = \{\mathbf{v}_2 - \mathbf{w}_1(\mathbf{w}_1^T \mathbf{w}_1)^{-1} \mathbf{w}_1^T \mathbf{v}_2\}^T \mathbf{w}_1 = \mathbf{v}_2^T \mathbf{w}_1 - \mathbf{v}_2^T \mathbf{w}_1 (\mathbf{w}_1^T \mathbf{w}_1)^{-1} \mathbf{w}_1^T \mathbf{w}_1 = 0$$

showing the orthogonality. Suppose that

$$\mathbf{w}_j^T \mathbf{w}_k = 0 \quad (j, k = 1, \dots, i-1; j \neq k)$$

hold. Then, we have

$$\begin{aligned} \mathbf{w}_k^T \mathbf{w}_i &= \mathbf{w}_k^T \left(\mathbf{v}_i - \sum_{j=1}^{i-1} \mathbf{w}_j \frac{\mathbf{w}_j^T \mathbf{v}_i}{\mathbf{w}_j^T \mathbf{w}_j} \right) = \mathbf{w}_k^T \mathbf{v}_i - \sum_{j=1}^{i-1} \mathbf{w}_k^T \mathbf{w}_j \frac{\mathbf{w}_j^T \mathbf{v}_i}{\mathbf{w}_j^T \mathbf{w}_j} \\ &= \mathbf{w}_k^T \mathbf{v}_i - \mathbf{w}_k^T \mathbf{w}_k \frac{\mathbf{w}_k^T \mathbf{v}_i}{\mathbf{w}_k^T \mathbf{w}_k} = 0 \quad (i = 2, \dots, p; k = 1, \dots, i-1), \end{aligned}$$

due to the assumption $\mathbf{w}_j^T \mathbf{w}_k = 0$ ($j, k = 1, \dots, i-1; j \neq k$), showing the required result $\mathbf{w}_j^T \mathbf{w}_k = 0$ ($j, k = 1, \dots, i; j \neq k$). Recall that $\mathbf{v}_j^T \mathbf{w}_j = \mathbf{w}_j^T \mathbf{w}_j$ ($j = 1, \dots, i$) mentioned earlier, which is obtained by $\mathbf{w}_j^T \mathbf{w}_k = 0$ ($j, k = 1, \dots, i; j \neq k$)

$$\text{and } \mathbf{w}_i = \mathbf{v}_i - \sum_{j=1}^{i-1} \mathbf{w}_j \frac{\mathbf{w}_j^T \mathbf{v}_i}{\mathbf{w}_j^T \mathbf{w}_j} \quad (i = 2, \dots, p).$$

The orthogonalization procedure is re-expressed by

$$\begin{aligned} \mathbf{w}_i &= \mathbf{v}_i - \sum_{j=1}^{i-1} \mathbf{w}_j \frac{\mathbf{w}_j^T \mathbf{v}_i}{\mathbf{w}_j^T \mathbf{w}_j} \\ &= \mathbf{v}_i - (\mathbf{w}_1, \dots, \mathbf{w}_{i-1}) \text{diag}\{(\mathbf{w}_1^T \mathbf{w}_1)^{-1}, \dots, (\mathbf{w}_{i-1}^T \mathbf{w}_{i-1})^{-1}\} (\mathbf{w}_1, \dots, \mathbf{w}_{i-1})^T \mathbf{v}_i \\ &\equiv \mathbf{v}_i - \mathbf{P}_{\mathbf{w}_{i-1}} \mathbf{v}_i = (\mathbf{I}_n - \mathbf{P}_{\mathbf{w}_{i-1}}) \mathbf{v}_i \equiv \mathbf{Q}_{\mathbf{w}_{i-1}} \mathbf{v}_i \quad (i = 2, \dots, p), \end{aligned}$$

where $\mathbf{P}_{\mathbf{W}_{i-1}} \equiv \mathbf{W}_{i-1}(\mathbf{W}_{i-1}^T \mathbf{W}_{i-1})^{-1} \mathbf{W}_{i-1}^T$ is the idempotent (i.e., $\mathbf{P}_{\mathbf{W}_{i-1}}^2 = \mathbf{P}_{\mathbf{W}_{i-1}}$) and symmetric projection matrix transforming or projecting \mathbf{v}_i onto the space spanned by the columns of $\mathbf{W}_{i-1} \equiv (\mathbf{w}_1, \dots, \mathbf{w}_{i-1})$ of full column rank by assumption; and $\mathbf{Q}_{\mathbf{W}_{i-1}} = \mathbf{I}_n - \mathbf{P}_{\mathbf{W}_{i-1}}$ is also an idempotent and symmetric projection matrix yielding the residual vector $\mathbf{v}_i - \mathbf{P}_{\mathbf{W}_{i-1}} \mathbf{v}_i$ or the projected vector on the space orthogonal to the column space of \mathbf{W}_{i-1} with $\mathbf{v}_i = \mathbf{P}_{\mathbf{W}_{i-1}} \mathbf{v}_i + \mathbf{Q}_{\mathbf{W}_{i-1}} \mathbf{v}_i$. Anderson (2003, p. 252) stated that “ \mathbf{w}_i is the vector from \mathbf{v}_i to the projection on $\mathbf{w}_1, \dots, \mathbf{w}_{i-1}$ ” with his Figure 7.1. He repeatedly stressed the equivalence of the column space of \mathbf{W}_{i-1} and that of $\mathbf{v}_1, \dots, \mathbf{v}_{i-1}$ in our expression.

Using the constructed $\mathbf{w}_1, \dots, \mathbf{w}_{i-1}$ by the Gram-Schmidt orthogonalization or projection, Anderson (2003, p. 252) defined

$$t_{ii} = \|\mathbf{w}_i\| = \sqrt{\mathbf{w}_i^T \mathbf{w}_i} \quad (i = 1, \dots, p)$$

and

$$t_{ij} = \mathbf{v}_i^T \mathbf{w}_j / \|\mathbf{w}_j\| \quad (i = 2, \dots, p; j = 1, \dots, i-1),$$

which may be uniformly expressed by $t_{ij} = \mathbf{v}_i^T \mathbf{w}_j / \|\mathbf{w}_j\| = (i = 2, \dots, p; j = 1, \dots, i)$ due to $\mathbf{v}_j^T \mathbf{w}_j = \mathbf{w}_j^T \mathbf{w}_j$ ($j = 1, \dots, i$) mentioned earlier. Then, noting that $\mathbf{w}_i = \mathbf{v}_i - \mathbf{P}_{\mathbf{W}_{i-1}} \mathbf{v}_i$, we have

$$\mathbf{v}_i = \mathbf{w}_i + \mathbf{P}_{\mathbf{W}_{i-1}} \mathbf{v}_i = \mathbf{w}_i + \sum_{j=1}^{i-1} \frac{\mathbf{w}_j \mathbf{w}_j^T}{\mathbf{w}_j^T \mathbf{w}_j} \mathbf{v}_i = \sum_{j=1}^i \frac{\mathbf{w}_j^T \mathbf{v}_i}{\|\mathbf{w}_j\|^2} \mathbf{w}_j = \sum_{j=1}^i \frac{t_{ij}}{\|\mathbf{w}_j\|} \mathbf{w}_j$$

and

$$\begin{aligned} (\mathbf{X}\mathbf{X}^T)_{ij} &= \mathbf{v}_i^T \mathbf{v}_j = \left\{ \sum_{k=1}^i \frac{t_{ik}}{\|\mathbf{w}_k\|} \mathbf{w}_k^T \right\} \sum_{k=1}^j \frac{t_{jk}}{\|\mathbf{w}_k\|} \mathbf{w}_k \\ &= \sum_{k=1}^j \frac{t_{ik}}{\|\mathbf{w}_k\|} \mathbf{w}_k^T \mathbf{w}_k \frac{t_{jk}}{\|\mathbf{w}_k\|} = \sum_{k=1}^j t_{ik} t_{jk} \quad (p \geq i \geq j \geq 1) \end{aligned}$$

(Anderson, 2003, p. 252). In $\mathbf{v}_i = \sum_{j=1}^i \frac{t_{ij}}{\|\mathbf{w}_j\|} \mathbf{w}_j$ ($i = 1, \dots, p$), $\mathbf{w}_j / \|\mathbf{w}_j\|$ ($j = 1, \dots, i-1$) is seen as the unit-norm vector representing the direction for the j -th coordinate in the $i-1$ coordinates given by $\mathbf{w}_1, \dots, \mathbf{w}_{i-1}$. He stated that “ $t_{ij}, j = 1, \dots, i-1$ are the first $i-1$ coordinates in the coordinate system with $\mathbf{w}_1, \dots, \mathbf{w}_{i-1}$ as the first coordinates axes” (p. 252). We also find that t_{ij} is $\|\mathbf{w}_j\|$ times the regression coefficient b_{ij} for \mathbf{v}_i on \mathbf{w}_j since

$$\begin{aligned} t_{ij} &= \mathbf{v}_i^T \mathbf{w}_j / \|\mathbf{w}_j\| = (\mathbf{v}_i^T \mathbf{w}_j / \mathbf{w}_j^T \mathbf{w}_j) \|\mathbf{w}_j\| \\ &= b_{ij} \|\mathbf{w}_j\| \quad (i = 2, \dots, p; j = 1, \dots, i-1). \end{aligned}$$

The properties of the normality of $t_{ij} = \mathbf{v}_i^T \mathbf{w}_j / \|\mathbf{w}_j\|$ ($i = 2, \dots, p; j = 1, \dots, i-1$) and their mutual independence shown by Anderson are based on the normality of the conditional distribution of the multivariate normal when \mathbf{w}_j ($j =$

$1, \dots, i-1$ are given and orthogonal transformation in $t_{ij} = \mathbf{v}_i^T \mathbf{w}_j / \|\mathbf{w}_j\|$ ($i = 2, \dots, p; j = 1, \dots, i-1$). That is, the standard normally-distributed variables $t_{ij} = \mathbf{v}_i^T \mathbf{w}_j / \|\mathbf{w}_j\|$ do not depend on $\mathbf{w}_1, \dots, \mathbf{w}_{i-1}$ indicating independence with $(\mathbf{w}_j / \|\mathbf{w}_j\|)^T \mathbf{w}_k / \|\mathbf{w}_k\| = \delta_{jk}$ ($j, k = 1, \dots, i-1$), where δ_{jk} is the Kronecker delta with $\delta_{jj} = 1$ and $\delta_{jk} = 0$ ($j \neq k$) (Anderson, 2003, Theorem 3.3.1).

The independent property of t_{ii} 's is given by $t_{ii} = \left\{ (\mathbf{X}\mathbf{X}^T)_{ii} - \sum_{j=1}^{i-1} t_{ij}^2 \right\}^{1/2}$. Although the same result as shown above by the didactic explanation of Anderson's derivation is directly given by Lemma 1, the two methods may be insightful with compensatory properties. [end of Remark 1]

2.2 The Wishart density for general correlated cases

For the correlated cases, four lemmas are provided. Lemma 3 is for three Jacobians in the product of two lower-triangular matrices, where the first Jacobian was used by Anderson (2003, Theorem 7.2.2) to derive the Wishart density for general correlated cases while the remaining two are given for generality with didactic purposes. Lemmas 4 and 5 are provided for the Jacobians in two alternative derivations of the general Wishart density. The proof of Lemma 6 associated with sufficient statistics is based on Ghosh and Sinha (2002).

Lemma 3 *Suppose that $\mathbf{A} = \mathbf{BC}$, where \mathbf{A} , \mathbf{B} and \mathbf{C} are $p \times p$ lower-triangular matrices. Consider the variable transformation from the non-zero elements of \mathbf{C} or \mathbf{B} to those of \mathbf{A} . Then, the Jacobians $J(\mathbf{C} \rightarrow \mathbf{A})$ and $J(\mathbf{B} \rightarrow \mathbf{A})$ are $\left| \prod_{i=1}^p b_{ii}^i \right|^{-1}$ and $\left| \prod_{i=1}^p c_{ii}^{p-i+1} \right|^{-1}$, respectively. When $\mathbf{B} = \mathbf{C}$, $J(\mathbf{B} \rightarrow \mathbf{A}) = \left| \prod_{i=1}^p \prod_{j=1}^i (b_{ii} + b_{jj}) \right|^{-1}$.*

Proof. Note that Anderson (2003, p. 254) gave $J(\mathbf{C} \rightarrow \mathbf{A})$. Since $a_{ij} = \sum_{k=j}^i b_{ik} c_{kj}$ ($p \geq i \geq j \geq 1$), we have

$$\begin{bmatrix} a_{11} \\ a_{21} \\ a_{22} \\ \vdots \\ a_{p1} \\ \vdots \\ a_{pm} \end{bmatrix} = \begin{bmatrix} b_{11} & 0 & 0 & \cdots & 0 & \cdots & 0 \\ * & b_{22} & 0 & \cdots & 0 & \cdots & 0 \\ * & * & b_{22} & \cdots & 0 & \cdots & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots & \vdots & \vdots \\ * & * & * & \cdots & b_{pp} & \cdots & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots & \vdots & \vdots \\ * & * & * & \cdots & * & \cdots & b_{pp} \end{bmatrix} \begin{bmatrix} c_{11} \\ c_{21} \\ c_{22} \\ \vdots \\ c_{p1} \\ \vdots \\ c_{pp} \end{bmatrix},$$

where the diagonal element of the lower-triangular matrix corresponding to the row for a_{ij} and the column for c_{ij} is b_{ii} ($p \geq i \geq j \geq 1$); the asterisks indicate

zero or non-zero elements; and

$$\begin{bmatrix} a_{11} \\ a_{21} \\ a_{22} \\ \vdots \\ a_{p1} \\ \vdots \\ a_{pp} \end{bmatrix} = \begin{bmatrix} c_{11} & 0 & 0 & \cdots & 0 & \cdots & 0 \\ * & c_{11} & 0 & \cdots & 0 & \cdots & 0 \\ * & * & c_{22} & \cdots & 0 & \cdots & 0 \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\ * & * & * & \cdots & c_{11} & \cdots & 0 \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\ * & * & * & \cdots & * & \cdots & c_{pp} \end{bmatrix} \begin{bmatrix} b_{11} \\ b_{21} \\ b_{22} \\ \vdots \\ b_{p1} \\ \vdots \\ b_{pp} \end{bmatrix},$$

where the corresponding diagonal element for a_{ij} and b_{ij} is c_{jj} ($p \geq i \geq j \geq 1$). Since the inverses of the Jacobian matrices for $J(\mathbf{C} \rightarrow \mathbf{A})$ and $J(\mathbf{B} \rightarrow \mathbf{A})$ on the right-hand sides of the above equations are lower-triangular, the Jacobians become the reciprocals of the absolute values of the determinants i.e., $\prod_{i=1}^p b_{ii}^i$ and $\prod_{i=1}^p c_{ii}^{p-i+1}$, respectively. The result when $\mathbf{B} = \mathbf{C}$ is obtained by the reciprocal of the determinant of the sum of the two lower-triangular matrices. \square

Lemma 4 Suppose that $\mathbf{A} = \mathbf{B}\mathbf{C}\mathbf{B}^T$, where \mathbf{A} and \mathbf{C} are $p \times p$ symmetric matrices; and \mathbf{B} is a lower-triangular matrix. Consider the variable transformation from the non-duplicated elements of \mathbf{C} to those of \mathbf{A} . Then, the Jacobian $J(\mathbf{C} \rightarrow \mathbf{A})$ is $|\mathbf{B}|_+^{-(p+1)}$.

Proof. Since the non-duplicated elements of \mathbf{A} using its diagonal and infra-diagonal elements are $a_{ij} = \sum_{k=1}^i \sum_{l=1}^j b_{ik}c_{kl}b_{jl}$ ($p \geq i \geq j \geq 1$), we have

$$\frac{\partial a_{ij}}{\partial c_{kl}} = b_{ik}b_{jl} \quad (p \geq i \geq j \geq 1; k = 1, \dots, i; l = 1, \dots, j),$$

which gives

$$\begin{bmatrix} a_{11} \\ a_{21} \\ a_{22} \\ \vdots \\ a_{p1} \\ \vdots \\ a_{pp} \end{bmatrix} = \begin{bmatrix} b_{11}b_{11} & 0 & 0 & \cdots & 0 & \cdots & 0 \\ * & b_{22}b_{11} & 0 & \cdots & 0 & \cdots & 0 \\ * & * & b_{22}b_{22} & \cdots & 0 & \cdots & 0 \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\ * & * & * & \cdots & b_{pp}b_{11} & \cdots & 0 \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\ * & * & * & \cdots & * & \cdots & b_{pp}b_{pp} \end{bmatrix} \begin{bmatrix} c_{11} \\ c_{21} \\ c_{22} \\ \vdots \\ c_{p1} \\ \vdots \\ c_{pp} \end{bmatrix},$$

where the diagonal element of the lower-triangular matrix for a_{ij} and c_{ij} is $\partial a_{ij} / \partial c_{ij} = b_{ii}b_{jj}$ ($p \geq i \geq j \geq 1$). Since $J(\mathbf{C} \rightarrow \mathbf{A})$ is the reciprocal of the absolute value of the determinant of the above lower-triangular matrix, we obtain $J(\mathbf{C} \rightarrow \mathbf{A}) = 1 / \left| \prod_{i=1}^p b_{ii}^{p+1} \right| = |\mathbf{B}|_+^{-(p+1)}$. \square

Lemma 5 Suppose that $\mathbf{A} = \mathbf{B}\mathbf{C}\mathbf{C}^T\mathbf{B}^T$, where \mathbf{A} is a $p \times p$ symmetric matrix; and \mathbf{B} and \mathbf{C} are lower-triangular matrices. Consider the variable transformation from the non-zero elements of \mathbf{C} to the non-duplicated elements of \mathbf{A} . Then, the Jacobian $J(\mathbf{C} \rightarrow \mathbf{A})$ is $|\mathbf{B}|_+^{-(p+1)} / \left| 2^p \prod_{i=1}^p c_{ii}^{p-i+1} \right|$.

Proof 1 The diagonal and infra-diagonal elements of \mathbf{A} are employed for its non-duplicated ones without loss of generality. Then, define $\mathbf{a} =$

$(a_{11}, a_{21}, a_{22}, \dots, a_{p1}, \dots, a_{pp})^T$ and $\mathbf{c} = (c_{11}, c_{21}, c_{22}, \dots, c_{p1}, \dots, c_{pp})^T$ with the elements lexicographically ordered. Since \mathbf{B} , \mathbf{C} and \mathbf{BC} are lower-triangular, the Jacobian matrix $\partial\mathbf{a}/\partial\mathbf{c}^T = \{\partial a_{ij}/\partial c_{kl}\}$ ($p \geq i \geq j \geq 1; p \geq k \geq l \geq 1$) becomes lower-triangular. This can be shown by

$$\begin{aligned} \frac{\partial a_{ij}}{\partial c_{kl}} &= \{\mathbf{B}(\mathbf{E}_{kl}\mathbf{C}^T + \mathbf{C}\mathbf{E}_{lk})\mathbf{B}^T\}_{ij} = (\mathbf{B}\mathbf{E}_{kl}\mathbf{C}^T\mathbf{B}^T)_{ij} + (\mathbf{B}\mathbf{C}\mathbf{E}_{lk}\mathbf{B}^T)_{ij} \\ &= b_{ik}(\mathbf{BC})_{jl} + (\mathbf{BC})_{il}b_{jk} \quad (p \geq i \geq j \geq 1; p \geq k \geq l \geq 1), \end{aligned}$$

where \mathbf{E}_{ij} is the matrix of an appropriate size, whose (i, j) th element is 1 with the remaining ones being 0. The right-hand side of the last equation in the above expression vanishes when $i < k$ or $\{i = k\} \cap \{j < l\}$. This condition indicates the lower-triangular form of $\partial\mathbf{a}/\partial\mathbf{c}^T = \{\partial a_{ij}/\partial c_{kl}\}$. Then, the diagonal elements are

$$\frac{\partial a_{ij}}{\partial c_{ij}} = \{\mathbf{B}(\mathbf{E}_{ij}\mathbf{C}^T + \mathbf{C}\mathbf{E}_{ji})\mathbf{B}^T\}_{ij} = (\mathbf{B}\mathbf{E}_{ij}\mathbf{C}^T\mathbf{B}^T)_{ij} = b_{ii}c_{jj}b_{jj} \quad (p \geq i > j \geq 1)$$

and

$$\frac{\partial a_{ii}}{\partial c_{ii}} = \{\mathbf{B}(\mathbf{E}_{ii}\mathbf{C}^T + \mathbf{C}\mathbf{E}_{ii})\mathbf{B}^T\}_{ii} = 2b_{ii}^2c_{ii} \quad (i = 1, \dots, p).$$

Since the determinant of the Jacobian matrix for $J(\mathbf{A} \rightarrow \mathbf{C})$ is

$$\begin{aligned} \prod_{i=1}^p \prod_{j=1}^i \frac{\partial a_{ij}}{\partial c_{ij}} &= \left(\prod_{i=1}^p \prod_{j=1}^{i-1} \frac{\partial a_{ij}}{\partial c_{ij}} \right) \prod_{i=1}^p \frac{\partial a_{ii}}{\partial c_{ii}} = 2^p \prod_{i=1}^p \prod_{j=1}^i b_{ii}c_{jj}b_{jj} \\ &= 2^p \left(\prod_{i=1}^p b_{ii}^i \right) \prod_{j=1}^p c_{jj}^{p-j+1} b_{jj}^{p-j+1} = 2^p \prod_{i=1}^p b_{ii}^{p+1} c_{ii}^{p-i+1} \\ &= 2^p |\mathbf{B}|^{p+1} \prod_{i=1}^p c_{ii}^{p-i+1}, \end{aligned}$$

the Jacobian $J(\mathbf{C} \rightarrow \mathbf{A})$ is the reciprocal of the absolute value of the above quantity:

$$J(\mathbf{C} \rightarrow \mathbf{A}) = |\mathbf{B}|_+^{-(p+1)} / \left| 2^p \prod_{i=1}^p c_{ii}^{p-i+1} \right|,$$

which is the required result. \square

Proof 2 The transformation $\mathbf{A} = \mathbf{BCC}^T\mathbf{B}^T$ is seen in two steps. In the first step, the transformation $\mathbf{C} \rightarrow \mathbf{CC}^T$ is considered, whose Jacobian is given by Lemma 2 as $J(\mathbf{C} \rightarrow \mathbf{CC}^T) = 1 / \left| 2^p \prod_{i=1}^p c_{ii}^{p-i+1} \right|$. The second step is for the transformation $\mathbf{CC}^T \rightarrow \mathbf{A} = \mathbf{BCC}^T\mathbf{B}^T$ with the Jacobian $J(\mathbf{CC}^T \rightarrow \mathbf{A}) = |\mathbf{B}|_+^{-(p+1)}$, which is given by Lemma 4. Then, the Jacobian $J(\mathbf{C} \rightarrow \mathbf{A})$ is the product of the two Jacobians due to the chain rule, which gives the required result. \square

Suppose that each column of a $p \times n$ matrix \mathbf{Y} follows $N_p(\mathbf{0}, \Sigma)$ with positive definite Σ independent of the other columns. Recall \mathbf{X} in Theorem 1. Let $\Sigma = \mathbf{BB}^T$ be the Cholesky decomposition, where \mathbf{B} is a fixed lower-triangular matrix

whose diagonal elements are positive for identification and convenience. Then, each column of $\mathbf{Y} = \mathbf{B}\mathbf{X}$ independently follows $N_p(\mathbf{0}, \boldsymbol{\Sigma})$. Define $\mathbf{S}_\boldsymbol{\Sigma} \equiv \mathbf{Y}\mathbf{Y}^\top = \mathbf{B}\mathbf{X}\mathbf{X}^\top\mathbf{B}^\top = \mathbf{B}\mathbf{S}\mathbf{B}^\top$, where $\mathbf{S} = \mathbf{S}_{\mathbf{I}_p} = \mathbf{X}\mathbf{X}^\top = \mathbf{T}\mathbf{T}^\top$, and the $\{p(p+1)/2\} \times 1$ vector $\mathbf{s}_\boldsymbol{\Sigma} \equiv (s_{\boldsymbol{\Sigma}11}, s_{\boldsymbol{\Sigma}21}, s_{\boldsymbol{\Sigma}22}, \dots, s_{\boldsymbol{\Sigma}p1}, \dots, s_{\boldsymbol{\Sigma}pp})^\top$ with $\mathbf{S}_\boldsymbol{\Sigma} = \{s_{\boldsymbol{\Sigma}ij}\}$ ($i, j = 1, \dots, p$).

Lemma 6 Define positive definite $\boldsymbol{\Sigma}_i = \mathbf{B}_i\mathbf{B}_i^\top$ and $\mathbf{S}_{\boldsymbol{\Sigma}_i} = \mathbf{B}_i\mathbf{S}\mathbf{B}_i^\top$ ($i = 1, 2$), where \mathbf{S} is as before. Denote the pdf's of $\mathbf{S}_{\boldsymbol{\Sigma}_i}$ at $\mathbf{S}_\boldsymbol{\Sigma}$ by $g_{\boldsymbol{\Sigma}=\boldsymbol{\Sigma}_i}(\mathbf{S}_\boldsymbol{\Sigma})$ ($i = 1, 2$). Then,

$$\frac{g_{\boldsymbol{\Sigma}=\boldsymbol{\Sigma}_1}(\mathbf{S}_\boldsymbol{\Sigma})}{g_{\boldsymbol{\Sigma}=\boldsymbol{\Sigma}_2}(\mathbf{S}_\boldsymbol{\Sigma})} = \frac{\phi_{p,n}(\mathbf{Y}|\mathbf{0}, \boldsymbol{\Sigma}_1)}{\phi_{p,n}(\mathbf{Y}|\mathbf{0}, \boldsymbol{\Sigma}_2)},$$

where $\phi_{p,n}(\mathbf{Y}|\mathbf{0}, \boldsymbol{\Sigma}_i) = \prod_{j=1}^n \phi_p\{(\mathbf{Y})_{\cdot j}|\mathbf{0}, \boldsymbol{\Sigma}_i\}$; $(\mathbf{Y})_{\cdot j}$ is the j -th column of \mathbf{Y} ; and

$$\phi_p\{(\mathbf{Y})_{\cdot j}|\mathbf{0}, \boldsymbol{\Sigma}_i\} = \frac{\exp\{-(\mathbf{Y})_{\cdot j}^\top \boldsymbol{\Sigma}_i^{-1} (\mathbf{Y})_{\cdot j}/2\}}{(2\pi)^{n/2} |\boldsymbol{\Sigma}_i|^{1/2}} \quad (i = 1, 2; j = 1, \dots, n).$$

Proof. The derivation is given by the factorization theorem for the sufficient statistic corresponding to $\mathbf{S}_\boldsymbol{\Sigma}$ for $\boldsymbol{\Sigma}$ as used by Ghosh and Sinha (2002, Equation (8)):

$$\phi_{p,n}(\mathbf{Y}|\mathbf{0}, \boldsymbol{\Sigma}_i) = g_{\boldsymbol{\Sigma}=\boldsymbol{\Sigma}_i}(\mathbf{S}_\boldsymbol{\Sigma})h(\mathbf{Y}) \quad (i = 1, 2),$$

which gives the required result. \square

The Wishart density for general correlated cases (see e.g., Srivastava & Khatri, 1979, Theorem 3.2.1; Anderson, 2003, Theorem 7.2.2) is derived in different ways.

Theorem 2 Let each column of a $p \times n$ matrix \mathbf{Y} follows $N_p(\mathbf{0}, \boldsymbol{\Sigma})$ with positive definite $\boldsymbol{\Sigma}$ independent of the other columns. Then, the pdf of $\mathbf{S}_\boldsymbol{\Sigma} = \mathbf{Y}\mathbf{Y}^\top$ is

$$w_p(\mathbf{S}_\boldsymbol{\Sigma}|\boldsymbol{\Sigma}, n) = \frac{\exp\{-\text{tr}(\boldsymbol{\Sigma}^{-1}\mathbf{S}_\boldsymbol{\Sigma})/2\} |\mathbf{S}_\boldsymbol{\Sigma}|^{(n-p-1)/2}}{2^{np/2} |\boldsymbol{\Sigma}|^{n/2} \Gamma_p(n/2)}.$$

Proof 1 Consider the transformation $\mathbf{T} \rightarrow \mathbf{S}_\boldsymbol{\Sigma} = \mathbf{B}\mathbf{T}\mathbf{T}^\top\mathbf{B}^\top$. The Jacobian is given by Lemma 5, when $\mathbf{A} = \mathbf{S}_\boldsymbol{\Sigma}$, $\mathbf{B} = \mathbf{B}$ and $\mathbf{C} = \mathbf{T}$ with added restrictions $b_{ii} > 0$ and $t_{ii} > 0$ ($i = 1, \dots, p$) as

$$J(\mathbf{T} \rightarrow \mathbf{S}_\boldsymbol{\Sigma}) = |\mathbf{B}|^{-(p+1)} / \left(2^p \prod_{i=1}^p t_{ii}^{p-i+1}\right) = |\boldsymbol{\Sigma}|^{-(p+1)/2} / \left(2^p \prod_{i=1}^p t_{ii}^{p-i+1}\right).$$

The pdf of \mathbf{T} denoted by $f_p(\mathbf{T})$ was given by Theorem 1. Then, we have

$$\begin{aligned}
 w_p(\mathbf{S}_\Sigma|\Sigma, n) &= f_p(\mathbf{T})J(\mathbf{T} \rightarrow \mathbf{S}_\Sigma) \\
 &= \frac{\exp\{-\text{tr}(\mathbf{T}\mathbf{T}^\text{T})/2\} \prod_{i=1}^p t_{ii}^{n-i}}{2^{(np/2)-p}\Gamma_p(n/2)} \frac{|\Sigma|^{-(p+1)/2}}{2^p \prod_{i=1}^p t_{ii}^{p-i+1}} \\
 &= \frac{\exp\{-\text{tr}(\mathbf{T}\mathbf{T}^\text{T})/2\} |\Sigma|^{-(p+1)/2} \prod_{i=1}^p t_{ii}^{n-p-1}}{2^{np/2}\Gamma_p(n/2)} \\
 &= \frac{\exp\{-\text{tr}(\mathbf{B}^{-1}\mathbf{S}_\Sigma\mathbf{B}^{\text{T}-1})/2\} |\Sigma|^{-(p+1)/2} |\mathbf{B}^{-1}\mathbf{S}_\Sigma\mathbf{B}^{\text{T}-1}|^{(n-p-1)/2}}{2^{np/2}\Gamma_p(n/2)} \\
 &= \frac{\exp\{-\text{tr}(\Sigma^{-1}\mathbf{S}_\Sigma)/2\} |\mathbf{S}_\Sigma|^{(n-p-1)/2}}{2^{np/2} |\Sigma|^{n/2} \Gamma_p(n/2)},
 \end{aligned}$$

where $\text{tr}(\mathbf{B}^{-1}\mathbf{S}_\Sigma\mathbf{B}^{\text{T}-1}) = \text{tr}(\mathbf{B}^{\text{T}-1}\mathbf{B}^{-1}\mathbf{S}_\Sigma) = \text{tr}\{(\mathbf{B}\mathbf{B}^\text{T})^{-1}\mathbf{S}_\Sigma\} = \text{tr}(\Sigma^{-1}\mathbf{S}_\Sigma)$ and $|\mathbf{B}^{-1}\mathbf{S}_\Sigma\mathbf{B}^{\text{T}-1}| = |\mathbf{S}_\Sigma||\Sigma|^{-1}$ are used. The last expression gives the required result. \square

Proof 2 Employ the two-step transformation $\mathbf{T} \rightarrow \mathbf{S} = \mathbf{T}\mathbf{T}^\text{T} \rightarrow \mathbf{S}_\Sigma = \mathbf{B}\mathbf{S}\mathbf{B}^\text{T}$. The first step was used by Theorem 1. The Jacobian $J(\mathbf{T} \rightarrow \mathbf{S} = \mathbf{T}\mathbf{T}^\text{T})$ in the first step is given by Lemma 2 by taking the reciprocal of the last result of the lemma while $J(\mathbf{S} \rightarrow \mathbf{S}_\Sigma = \mathbf{B}\mathbf{S}\mathbf{B}^\text{T})$ is obtained by Lemma 4. That is,

$$\begin{aligned}
 w_p(\mathbf{S}_\Sigma|\Sigma, n) &= f_p(\mathbf{T})J(\mathbf{T} \rightarrow \mathbf{S})J(\mathbf{S} \rightarrow \mathbf{S}_\Sigma) \\
 &= \frac{\exp\{-\text{tr}(\mathbf{S})/2\} |\mathbf{S}|^{(n-p-1)/2}}{2^{np/2}\Gamma_p(n/2)} J(\mathbf{S} \rightarrow \mathbf{S}_\Sigma) \\
 &= \frac{\exp\{-\text{tr}(\mathbf{S})/2\} |\mathbf{S}|^{(n-p-1)/2}}{2^{np/2}\Gamma_p(n/2)} |\mathbf{B}|^{-(p+1)} \\
 &= \frac{\exp\{-\text{tr}(\Sigma^{-1}\mathbf{S}_\Sigma)/2\} |\Sigma^{-1}\mathbf{S}_\Sigma|^{(n-p-1)/2} |\Sigma|^{-(p+1)/2}}{2^{np/2}\Gamma_p(n/2)} \\
 &= \frac{\exp\{-\text{tr}(\Sigma^{-1}\mathbf{S}_\Sigma)/2\} |\mathbf{S}_\Sigma|^{(n-p-1)/2}}{2^{np/2} |\Sigma|^{n/2} \Gamma_p(n/2)}.
 \end{aligned}$$

\square

Proof 3 (Anderson, 2003, Theorem 7.2.2) Anderson used an alternative two-step transformation $\mathbf{T} \rightarrow \mathbf{T}^* = \mathbf{B}\mathbf{T} \rightarrow \mathbf{S}_\Sigma = \mathbf{T}^*\mathbf{T}^{*\text{T}}$. The Jacobian $J(\mathbf{T} \rightarrow \mathbf{T}^*)$ is given by the first result of Lemma 3 while $J(\mathbf{T}^* \rightarrow \mathbf{S}_\Sigma)$ is given by the reciprocal

of the last result in Lemma 2 when $\mathbf{T} = \mathbf{T}^*$. That is,

$$\begin{aligned}
w_p(\mathbf{S}_\Sigma|\Sigma, n) &= f_p(\mathbf{T})J(\mathbf{T} \rightarrow \mathbf{T}^*)J(\mathbf{T}^* \rightarrow \mathbf{S}_\Sigma) \\
&= \frac{\exp\{-\text{tr}(\mathbf{T}\mathbf{T}^\text{T})/2\} \prod_{i=1}^p t_{ii}^{n-i}}{2^{(np/2)-p} \Gamma_p(n/2)} \left(\prod_{i=1}^p b_{ii}^i\right)^{-1} \left(2^p \prod_{i=1}^p t_{ii}^{*p-i+1}\right)^{-1} \\
&= \frac{\exp\{-\text{tr}(\mathbf{T}\mathbf{T}^\text{T})/2\} \prod_{i=1}^p (t_{ii}^*/b_{ii})^{n-i}}{2^{np/2} \Gamma_p(n/2) \left(\prod_{i=1}^p b_{ii}^i\right) \prod_{i=1}^p t_{ii}^{*p-i+1}} = \frac{\exp\{-\text{tr}(\mathbf{T}\mathbf{T}^\text{T})/2\} \prod_{i=1}^p t_{ii}^{*n-p-1}}{2^{np/2} \left(\prod_{i=1}^p b_{ii}^i\right) \Gamma_p(n/2)} \\
&= \frac{\exp\{-\text{tr}(\Sigma^{-1}\mathbf{S}_\Sigma)/2\} |\mathbf{S}_\Sigma|^{(n-p-1)/2}}{2^{np/2} |\Sigma|^{n/2} \Gamma_p(n/2)}.
\end{aligned}$$

□

Proof 4 Use Theorem 1 and Lemma 6 when $\Sigma_1 = \mathbf{I}_p$ and $\Sigma_2 = \mathbf{B}_2 \mathbf{B}_2^\text{T} = \Sigma^{1/2}(\Sigma^{1/2})^\text{T} = \Sigma$. Then, we have

$$\begin{aligned}
w_p(\mathbf{S}_\Sigma|\Sigma, n) &= w_p(\mathbf{S}_\Sigma|\mathbf{I}_p, n) \frac{\phi_{p,n}(\mathbf{Y}|\mathbf{0}, \Sigma)}{\phi_{p,n}(\mathbf{Y}|\mathbf{0}, \mathbf{I}_p)} \\
&= \frac{\exp\{-\text{tr}(\mathbf{S}_\Sigma)/2\} |\mathbf{S}_\Sigma|^{(n-p-1)/2} \exp\{-\text{tr}(\mathbf{Y}\mathbf{Y}^\text{T}\Sigma^{-1})/2\} / \{(2\pi)^{pn/2} |\Sigma|^{n/2}\}}{2^{np/2} \Gamma_p(n/2)} \frac{\exp\{-\text{tr}(\mathbf{Y}\mathbf{Y}^\text{T})/2\} / (2\pi)^{pn/2}}{\exp\{-\text{tr}(\Sigma^{-1}\mathbf{S}_\Sigma)/2\} |\mathbf{S}_\Sigma|^{(n-p-1)/2}} \\
&= \frac{\exp\{-\text{tr}(\Sigma^{-1}\mathbf{S}_\Sigma)/2\} |\mathbf{S}_\Sigma|^{(n-p-1)/2}}{2^{np/2} |\Sigma|^{n/2} \Gamma_p(n/2)}.
\end{aligned}$$

□

3 Remarks and Conclusion

For the general correlated cases, four proofs are shown in Theorem 2. The one-step first proof uses $f_p(\mathbf{T})$ with $J(\mathbf{T} \rightarrow \mathbf{S}_\Sigma)$ given by Lemma 5, where $\mathbf{S}_\Sigma = \mathbf{B}\mathbf{T}\mathbf{T}^\text{T}\mathbf{B}^\text{T}$ with lower-triangular \mathbf{B} and \mathbf{T} is seen as a two-fold Bartlett (Cholesky) decomposition or a usual Bartlett (1933) $\mathbf{S}_\Sigma = \mathbf{B}\mathbf{T}(\mathbf{B}\mathbf{T})^\text{T}$ in terms of lower-triangular $\mathbf{B}\mathbf{T}$. The two-step second proof uses $f_p(\mathbf{T})$ with $J(\mathbf{T} \rightarrow \mathbf{S} = \mathbf{T}\mathbf{T}^\text{T})$ and $J(\mathbf{S} \rightarrow \mathbf{S}_\Sigma = \mathbf{B}\mathbf{S}\mathbf{B}^\text{T})$ obtained by Lemmas 2 and 4, respectively. Anderson (2003)'s two-step third proof uses $f_p(\mathbf{T})$ with $J(\mathbf{T} \rightarrow \mathbf{T}^* = \mathbf{B}\mathbf{T})$ and $J(\mathbf{T}^* \rightarrow \mathbf{S}_\Sigma)$ given by Lemmas 3 and 2, respectively. Among the four proofs, the first and fourth ones are relatively simple. The remaining two-step proofs seem to be comparable. It is found that in order to derive the final Jacobian by Proofs 2 and 3, Lemma 2 is firstly and secondly used, respectively. When only the pdf of $\mathbf{S} (= \mathbf{S}_{\Sigma=\mathbf{I}_p})$ is focused on, Proof 2 may be the simplest though the same result is immediately obtained from the pdf of \mathbf{S}_Σ substituting $\Sigma = \mathbf{I}_p$.

In each of the four proofs, $f_p(\mathbf{T})$ is used. Two derivations for $f_p(\mathbf{T})$ were shown. The first method using Lemma 1 is much simpler than that used by Anderson (2003) as detailed in Remark 1. The author believes that this simplification will reduce the difficulties frequently encountered when beginning

students/researchers master the Wishart density. Note that when the Wishart density for $w_p(\mathbf{S}|\mathbf{I}_p, n)$ is given, $f_p(\mathbf{T})$ is obtained using $J(\mathbf{S} \rightarrow \mathbf{T})$ in Lemma 2 as easily as the transformation $J(\mathbf{T} \rightarrow \mathbf{S})$, which is the reversed problem (see Bartlett, 1933; Muirhead, 1982, Theorem 3.2.14).

Remark 2 Lemma 1 gave the justification of $\mathbf{X}\mathbf{X}^T = \mathbf{T}\mathbf{T}^T$ with mutually independent normal t_{ij} ($p \geq i > j \geq 1$) and chi-distributed t_{ii} ($i = 1, \dots, p$). While the chi-square distribution of $(\mathbf{T}\mathbf{T}^T)_{ii}$ is obvious, the distribution of $(\mathbf{T}\mathbf{T}^T)_{ij}$ ($i > j$) is that of the product sum of p pairs of independent normals (the product-sum normal for short). The pdf and mgf of the product-sum normal in the case of a possibly correlated single pair was given by Craig (1936) (see also Ogasawara, 2023a, Remarks S1-S4). For current developments of this issue, see e.g., Seijas-Macías and Oliveira (2012), Seijas-Macías, Oliveira, Oliveira, and Leiva (2020), and Gaunt (2022).

Remark 3 As addressed earlier, the complicated property found in many of the proofs of the Wishart density seems to be due partially to the associated Jacobians in e.g., Srivastava and Khatri (1979, Section 3.2) and Anderson (2003, Section 7.2). The proof of the Wishart density in Theorem 1 is similar to that in Srivastava and Khatri (1979, Section 3.2). Though the Jacobian in Lemma 2 was also used by Srivastava and Khatri, we did not use the Jacobian of $\mathbf{X} \rightarrow \{\mathbf{T}, \mathbf{V}^*\}$ in $\mathbf{X} = \mathbf{T}\mathbf{V}^*$, where \mathbf{V}^* is a $p \times n$ semi-orthonormal matrix with $\mathbf{V}^*\mathbf{V}^{*\top} = \mathbf{I}_p$ (see Srivastava & Khatri, 1979, Exercise 1.33). Instead, we used the marginal chi and normal distributions for \mathbf{T} as in Anderson (2003).

As shown earlier, in the three proofs of the Wishart density $w_p(\mathbf{S}_\Sigma|\Sigma, n)$, the Bartlett-like Cholesky decomposition $\Sigma = \mathbf{B}\mathbf{B}^T$ is used for non-stochastic Σ . Though this factorization gives simple results, other factorizations can also be used with $\Sigma = \mathbf{B}\mathbf{G}\mathbf{G}^T\mathbf{B}^T = \mathbf{B}\mathbf{G}(\mathbf{B}\mathbf{G})^T = \mathbf{D}\mathbf{D}^T$, where $\mathbf{G}\mathbf{G}^T = \mathbf{G}^T\mathbf{G} = \mathbf{I}_p$ and $\mathbf{D} = \mathbf{B}\mathbf{G}$. For illustration, Proof 5 using $\mathbf{D} = \Sigma^{1/2}$ with $(\Sigma^{1/2})^2 = \Sigma$ will be shown in Appendix A for didactic purposes with associated remarks. The concise derivation of Khatri (1963) will be explained in Appendix B. The Bartlett decomposition $\mathbf{S} = \mathbf{T}\mathbf{T}^T$ can also be replaced by other ones with the same number of random variables. The case called the exchanged Bartlett decomposition will be shown in Appendix C.

Conclusion Among Proofs 1 to 4 of the Wishart distribution given earlier and Proofs 5 to 7 to be shown in the appendix for expository purposes, Proof 4 using our Lemma 1 for the equivalence of the distributions of the product-sum normal and the product of the chi and standard normal as well as Lemma 6 for the factorization theorem given by Ghosh and Sinha (2002) is the simplest. Since Proof 4 uses elementary and self-contained methods, the proof may be understood by beginning students/researchers without much difficulty.

References

- Anderson, T. W. (2003). *An introduction to multivariate statistical analysis* (3rd ed.). New York: Wiley.

- Bartlett, M. S. (1933). On the theory of statistical regression. *Proceedings of the Royal Society of Edinburgh*, 53, 260–283. doi: <https://doi.org/10.1017/s0370164600015637>
- Bodnar, T., & Okhrin, Y. (2008). Properties of the singular, inverse and generalized inverse partitioned Wishart distributions. *Journal of Multivariate Analysis*, 99(10), 2389–2405. doi: <https://doi.org/10.1016/j.jmva.2008.02.024>
- Chen, Y.-L., & Weng, L.-J. (2023). On Horn’s approximation to the sampling distribution of eigenvalues from random correlation matrices in parallel analysis. *Current Psychology*(online published). doi: <https://doi.org/10.1007/s12144-023-04635-9>
- Craig, C. C. (1936). On the frequency function of xy . *The Annals of Mathematical Statistics*, 7(1), 1–15. doi: <https://doi.org/10.1214/aoms/1177732541>
- Deemer, W. L., & Olkin, I. (1951). The jacobians of certain matrix transformations useful in multivariate analysis: Based on lectures of P. L. Hsu at the University of North Carolina, 1947. *Biometrika*, 38(3/4), 345–367. doi: <https://doi.org/10.2307/2332581>
- DLMF. (2021). *NIST Digital Library of Mathematical Functions*. National Institutes of Standards and Technology, U. S. Department of Commerce. Release 1.1.3 of 2021-09-15. Retrieved from <http://dlmf.nist.gov/> (F. W. J. Olver, A. B. Olde Daalhuis, D. W. Lozier, B. I. Schneider, R. F. Boisvert, C. W. Clark, B. R. Miller, B. V. Saunders, H. S. Cohl, M. A. McClain (Eds))
- Fisher, R. A. (1915). Frequency distribution of the values of the correlation coefficient in samples from an indefinitely large population. *Biometrika*, 10(4), 507–521. doi: <https://doi.org/10.2307/2331838>
- Gaunt, R. E. (2022). The basic distributional theory for the product of zero mean correlated normal random variables. *Statistica Neerlandica*, 76(4), 450–470. doi: <https://doi.org/10.1111/stan.12267>
- Ghosh, M., & Sinha, B. K. (2002). A simple derivation of the Wishart distribution. *The American Statistician*, 56(2), 100–101. doi: <https://doi.org/10.1198/000313002317572754>
- Hsu, P. L. (1940). An algebraic derivation of the distribution of rectangular coordinates. *Proceedings of the Edinburgh Mathematical Society*, 6(3), 185–189. doi: https://doi.org/10.1007/978-1-4684-9324-5_14
- Khatri, C. G. (1963). (no title). *Journal of the Indian Statistical Association*, 1(Queries section), 52–54.
- Kshirsagar, A. M. (1959). Bartlett decomposition and Wishart distribution. *The Annals of Mathematical Statistics*, 30(1), 239–241. doi: <https://doi.org/10.1214/aoms/1177706379>
- Kshirsagar, A. M. (1972). *Multivariate analysis*. New York: Marcel Dekker. doi: <https://doi.org/10.2307/1267507>
- Liu, H., Qu, W., Zhang, Z., & Wu, H. (2022). A new Bayesian structural equation modeling approach with priors on the covariance matrix parameter. *Journal of Behavioral Data Science*, 2(2), 23–46. doi:

- <https://doi.org/10.35566/jbds/v2n2/p2>
- Magnus, J. R., & Neudecker, H. (1986). Symmetry, 0-1 matrices and Jacobians: A review. *Econometric Theory*, 2(2), 157–190. doi: <https://doi.org/10.1017/s0266466600011476>
- Magnus, J. R., & Neudecker, H. (1999). *Matrix differential calculus with applications in statistics and econometrics* (Rev. ed.). New York: Wiley. doi: <https://doi.org/10.1002/9781119541219>
- Marčenko, V. A., & Pastur, L. A. (1967). Distribution of eigenvalues for some sets of random matrices. *Mathematics of the USSR-Sbornik*, 1(4), 457–483. doi: <https://doi.org/10.1070/sm1967v001n04abeh001994>
- Mathai, A. M., & Provost, S. B. (2022). On the singular gamma, Wishart, and beta matrix-variate density functions. *Canadian Journal of Statistics*, 50(4), 1143–1165. doi: <https://doi.org/10.1002/cjs.11710>
- Mathai, A. M., Provost, S. B., & Haubold, H. J. (2022). *Multivariate statistical analysis in the real and complex domains*. Cham, Switzerland: Springer Nature. doi: <https://doi.org/10.1007/978-3-030-95864-0>
- Muirhead, R. J. (1982). *Aspects of multivariate statistical theory*. New York: Wiley. doi: <https://doi.org/10.2307/2288369>
- Ogasawara, H. (2007). Asymptotic expansions of the distributions of estimators in canonical correlation analysis under nonnormality. *Journal of Multivariate Analysis*, 98(9), 1726–1750. doi: <https://doi.org/10.1016/j.jmva.2006.12.001>
- Ogasawara, H. (2016). Bias correction of the Akaike information criterion in factor analysis. *Journal of Multivariate Analysis*, 149, 144–159. doi: <https://doi.org/10.1016/j.jmva.2016.04.003>
- Ogasawara, H. (2022). *A stochastic derivation of the surface area of the (n-1)-sphere*. Preprint at ResearchGate. doi: <https://doi.org/10.13140/RG.2.2.28827.95528>
- Ogasawara, H. (2023a). Supplement to the paper “on some known derivations and new ones for the Wishart distribution: A didactic”. To appear in *Economic Review (Otaru University of Commerce)*, 74(2 & 3, <https://www.otaru-uc.ac.jp/~emt-hogasa/>, <https://barrel.repo.nii.ac.jp/>).
- Ogasawara, H. (2023b). The Wishart distribution with two different degrees of freedom. *Statistics and Probability Letters*, 200(109866, online published). doi: <https://doi.org/10.1016/j.spl.2023.109866>
- Olkin, I. (2002). The 70th anniversary of the distribution of random matrices: A survey. *Linear Algebra and Its Applications*, 354, 231–243. doi: [https://doi.org/10.1016/s0024-3795\(01\)00314-7](https://doi.org/10.1016/s0024-3795(01)00314-7)
- Schuberth, F. (2021). The Henseler-Ogasawara specification of composites in structural equation modeling: A tutorial. *Psychological methods*(online published). doi: <https://doi.org/10.1037/met0000432>
- Seijas-Macías, A., & Oliveira, A. (2012). An approach to distribution of the product of two normal variables. *Discussiones Mathematicae Probability and Statistics*, 32(1-2), 87–99. doi: <https://doi.org/10.7151/dmps.1146>

- Seijas-Macías, A., Oliveira, A., Oliveira, T. A., & Leiva, V. (2020). Approximating the distribution of the product of two normally distributed random variables. *Symmetry*, *12*(8), 1201. doi: <https://doi.org/10.3390/sym12081201>
- Srivastava, M. S. (2003). Singular Wishart and multivariate beta distributions. *The Annals of Statistics*, *31*(5), 1537–1560. doi: <https://doi.org/10.1214/aos/1065705118>
- Srivastava, M. S., & Khatri, C. G. (1979). *An introduction to multivariate statistics*, 350. Amsterdam: Elsevier.
- Wijsman, R. A. (1957). Random orthogonal transformations and their use in some classical distribution problems in multivariate analysis. *The Annals of Mathematical Statistics*, *28*(2), 415–423. doi: <https://doi.org/10.1214/aoms/1177706969>
- Wilks, S. S. (1962). *Mathematical statistics*. New York: Wiley. doi: <https://doi.org/10.2307/2311277>
- Wishart, J. (1928). The generalised product moment distribution in samples from a normal multivariate population. *Biometrika*, *20A*, 32–52. doi: <https://doi.org/10.2307/2331939>
- Wishart, J., & Bartlett, M. S. (1933). The generalised product moment distribution in a normal system. *Mathematical Proceedings of the Cambridge Philosophical Society*, *29*(2), 260–270. doi: <https://doi.org/10.1017/s0305004100011063>
- Yao, J., Zheng, S., & Bai, Z. D. (2015). *Sample covariance matrices and high-dimensional data analysis*. New York: Cambridge University Press.
- Yonenaga, K. (2022). *Functionals of a Wishart matrix and a normal vector and its application to linear discriminant analysis*. Doctoral dissertation, Graduate School of Economics and Business, Hokkaido University, Japan.
- Yu, X., Schuberth, F., & Henseler, J. (2023). Specifying composites in structural equation modeling: A refinement of the Henseler–Ogasawara specification. *Statistical Analysis and Data Mining: The ASA Data Science Journal*(online published). doi: <https://doi.org/10.1002/sam.11608>
- Zhang, Z. (2021). A note on Wishart and inverse Wishart priors for covariance matrix. *Journal of Behavioral Data Science*, *1*(2), 119–126. doi: <https://doi.org/10.35566/jbds/v1n2/p2>

Appendix A An alternative proof of the Wishart density for correlated cases

Let $\Sigma^{1/2}$ be a symmetric matrix-square-root of Σ satisfying $(\Sigma^{1/2})^2 = \Sigma$. Then, we have $(\mathbf{Y})_{\cdot j} = (\Sigma^{1/2}\mathbf{X})_{\cdot j} \sim N_p(\mathbf{0}, \Sigma)$ as $(\mathbf{B}\mathbf{X})_{\cdot j} \sim N_p(\mathbf{0}, \Sigma)$ ($j = 1, \dots, n$), which gives $\mathbf{S}_\Sigma \equiv \mathbf{Y}\mathbf{Y}^T = \Sigma^{1/2}\mathbf{X}\mathbf{X}^T\Sigma^{1/2} = \Sigma^{1/2}\mathbf{S}\Sigma^{1/2}$, where \mathbf{S}_Σ is redefined using $\Sigma^{1/2}$. Let $\mathbf{s}_\Sigma = (s_{\Sigma 11}, s_{\Sigma 21}, s_{\Sigma 22}, \dots, s_{\Sigma p1}, \dots, s_{\Sigma pp})^T$ using redefined $\mathbf{S}_\Sigma = \{s_{\Sigma ij}\}$ ($i, j = 1, \dots, p$). Then, $\mathbf{D}_p\mathbf{s}_\Sigma = \text{vec}(\mathbf{S}_\Sigma)$ follows, where \mathbf{D}_p of full column rank is the $p^2 \times \{p(p+1)/2\}$ duplication matrix consisting of 0's and 1's

(Magnus & Neudecker, 1999, Chapter 3, Section 8); and $\text{vec}(\cdot)$ is the vectorizing operator stacking the columns of a matrix in parentheses sequentially with the first column on the top. Using the formula $\text{vec}(\mathbf{ABC}) = (\mathbf{C}^T \otimes \mathbf{A})\text{vec}(\mathbf{B})$ (see Magnus & Neudecker, 1999, Chapter 2, Theorem 2), where \otimes denotes the direct or Kronecker product, we obtain

$$\mathbf{D}_p \mathbf{s}_\Sigma = \text{vec}(\mathbf{S}_\Sigma) = \text{vec}(\Sigma^{1/2} \mathbf{S} \Sigma^{1/2}) = (\Sigma^{1/2} \otimes \Sigma^{1/2}) \text{vec}(\mathbf{S}) = (\Sigma^{1/2} \otimes \Sigma^{1/2}) \mathbf{D}_p \mathbf{s}.$$

Pre-multiplying the above equation by $(\mathbf{D}_p^T \mathbf{D}_p)^{-1} \mathbf{D}_p^T \equiv \mathbf{D}_p^+$, which is the left- or Moore-Penrose generalized inverse of \mathbf{D}_p with $\mathbf{D}_p^+ \mathbf{D}_p = \mathbf{I}_{p(p+1)/2}$ (see Magnus & Neudecker, 1999, Chapter 3, Section 8), we have

$$\mathbf{s}_\Sigma = \mathbf{D}_p^+ (\Sigma^{1/2} \otimes \Sigma^{1/2}) \mathbf{D}_p \mathbf{s}.$$

The Jacobian of the transformation $\mathbf{S}_\Sigma \rightarrow \mathbf{S}$ or equivalently $\mathbf{s}_\Sigma \rightarrow \mathbf{s}$ is given by $|\mathbf{D}_p^+ (\Sigma^{1/2} \otimes \Sigma^{1/2}) \mathbf{D}_p|_+ = |\Sigma|^{(p+1)/2}$, which is derived using the following lemma.

Lemma 7 (Magnus & Neudecker, 1986, Equation (7.11)). *Let \mathbf{A} be a $p \times p$ positive definite matrix with distinct eigenvalues. Then, $|\mathbf{D}_p^+(\mathbf{A} \otimes \mathbf{A}) \mathbf{D}_p| = |\mathbf{A}|^{p+1}$.*

Proof. While Magnus and Neudecker (1986) used Shur's theorem for the existence of a non-singular matrix \mathbf{V} satisfying $\mathbf{V}^{-1} \mathbf{A} \mathbf{V} = \mathbf{M}$, where \mathbf{M} is an upper-triangular matrix for a general square matrix \mathbf{A} , we use a familiar special case of the theorem as $\mathbf{L}^T \mathbf{A} \mathbf{L} = \Lambda$ when $\mathbf{A} = \mathbf{L} \Lambda \mathbf{L}^T$ with $\mathbf{L} \mathbf{L}^T = \mathbf{L}^T \mathbf{L} = \mathbf{I}_p$ and $\Lambda = \text{diag}(\lambda_1, \dots, \lambda_p)$ ($\lambda_1 > \dots > \lambda_p > 0$), where the columns of \mathbf{L} and λ_i ($i = 1, \dots, p$) are the eigenvectors and eigenvalues of \mathbf{A} , respectively. Note that

$$\begin{aligned} & \mathbf{D}_p^+ (\mathbf{L}^T \otimes \mathbf{L}^T) \mathbf{D}_p \mathbf{D}_p^+ (\mathbf{A} \otimes \mathbf{A}) \mathbf{D}_p \mathbf{D}_p^+ (\mathbf{L} \otimes \mathbf{L}) \mathbf{D}_p \\ &= \mathbf{D}_p^+ (\mathbf{L}^T \otimes \mathbf{L}^T) (\mathbf{A} \otimes \mathbf{A}) (\mathbf{L} \otimes \mathbf{L}) \mathbf{D}_p \\ &= \mathbf{D}_p^+ \{ (\mathbf{L}^T \mathbf{A} \mathbf{L}) \otimes (\mathbf{L}^T \mathbf{A} \mathbf{L}) \} \mathbf{D}_p = \mathbf{D}_p^+ (\Lambda \otimes \Lambda) \mathbf{D}_p, \end{aligned}$$

where $\mathbf{D}_p \mathbf{D}_p^+ (\mathbf{A} \otimes \mathbf{A}) = (\mathbf{A} \otimes \mathbf{A}) \mathbf{D}_p \mathbf{D}_p^+$ and $\mathbf{D}_p \mathbf{D}_p^+ \mathbf{D}_p = \mathbf{D}_p$ (Magnus & Neudecker, 1999, Chapter 3, Theorem 13) are used, followed by the transformation given by $(\mathbf{A} \otimes \mathbf{B})(\mathbf{C} \otimes \mathbf{D}) = (\mathbf{AC}) \otimes (\mathbf{BD})$ when multiplications are defined.

Note that $\mathbf{D}_p^+ (\mathbf{L}^T \otimes \mathbf{L}^T) \mathbf{D}_p = \{ \mathbf{D}_p^+ (\mathbf{L} \otimes \mathbf{L}) \mathbf{D}_p \}^{-1}$ since

$$\mathbf{D}_p^+ (\mathbf{L}^T \otimes \mathbf{L}^T) \mathbf{D}_p \mathbf{D}_p^+ (\mathbf{L} \otimes \mathbf{L}) \mathbf{D}_p = \mathbf{D}_p^+ (\mathbf{L}^T \otimes \mathbf{L}^T) (\mathbf{L} \otimes \mathbf{L}) \mathbf{D}_p = \mathbf{D}_p^+ \mathbf{D}_p = \mathbf{I}_{p(p+1)/2}.$$

Consequently, we can write as

$$\begin{aligned} & \mathbf{D}_p^+ (\mathbf{L}^T \otimes \mathbf{L}^T) \mathbf{D}_p \mathbf{D}_p^+ (\mathbf{A} \otimes \mathbf{A}) \mathbf{D}_p \mathbf{D}_p^+ (\mathbf{L} \otimes \mathbf{L}) \mathbf{D}_p \\ & \equiv \mathbf{B}^{-1} \mathbf{D}_p^+ (\mathbf{A} \otimes \mathbf{A}) \mathbf{D}_p \mathbf{B} = \mathbf{D}_p^+ (\Lambda \otimes \Lambda) \mathbf{D}_p, \end{aligned}$$

which shows that the eigenvalues of $\mathbf{D}_p^+(\mathbf{A} \otimes \mathbf{A})\mathbf{D}_p$ are the same as those of $\mathbf{D}_p^+(\mathbf{A} \otimes \mathbf{A})\mathbf{D}$ (see e.g., Magnus & Neudecker, 1999, Chapter 1, Theorem 5). Employ the double subscript notation as used earlier for the row numbers i and j ($p \geq i \geq j \geq 1$) and column numbers k and l ($p \geq k \geq l \geq 1$) of the $\{p(p+1)/2\} \times \{p(p+1)/2\}$ matrix $\mathbf{D}_p^+(\mathbf{A} \otimes \mathbf{A})\mathbf{D}_p$. These numbers correspond to the subscripts of the elements of e.g., the $\{p(p+1)/2\} \times 1$ vector $\mathbf{s} = (s_{11}, s_{21}, s_{22}, \dots, s_{p1}, \dots, s_{pp})^\top$.

Consider $(\mathbf{A} \otimes \mathbf{A})\mathbf{D}_p$, where the (k, k) th columns of $(\mathbf{A} \otimes \mathbf{A})\mathbf{D}_p$ ($k = 1, \dots, p$) are unchanged from the corresponding ones of $\mathbf{A} \otimes \mathbf{A}$ while the (k, l) th columns ($p \geq k > l \geq 1$) of $(\mathbf{A} \otimes \mathbf{A})\mathbf{D}_p$ are combined ones as the sum of the (k, l) - and (l, k) -th columns of $\mathbf{A} \otimes \mathbf{A}$ such that e.g.,

$$(\mathbf{A} \otimes \mathbf{A})\mathbf{D}_2 = \text{diag}(\lambda_1^2, \lambda_1 \lambda_2, \lambda_2 \lambda_1, \lambda_2^2) \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix} = \begin{pmatrix} \lambda_1^2 & 0 & 0 \\ 0 & \lambda_1 \lambda_2 & 0 \\ 0 & \lambda_2 \lambda_1 & 0 \\ 0 & 0 & \lambda_2^2 \end{pmatrix}$$

when $p = 2$. For the second transformation $\mathbf{D}_p^+(\mathbf{A} \otimes \mathbf{A})\mathbf{D}_p$, noting that $\mathbf{D}_p^+ = (\mathbf{D}_p^\top \mathbf{D}_p)^{-1} \mathbf{D}_p^\top$ consists of 1's, 1/2's and 0's as $\mathbf{D}_2^+ = \begin{pmatrix} 1 & 0 & 0 & 0 \\ 0 & 1/2 & 1/2 & 0 \\ 0 & 0 & 0 & 1 \end{pmatrix}$, we find that $\mathbf{D}_p^+(\mathbf{A} \otimes \mathbf{A})\mathbf{D}_p$ is the $\{p(p+1)/2\} \times \{p(p+1)/2\}$ diagonal matrix whose diagonal elements are λ_i^2 ($i = 1, \dots, p$) and $\lambda_i \lambda_j$ ($p \geq i > j \geq 1$) as $\mathbf{D}_2^+(\mathbf{A} \otimes \mathbf{A})\mathbf{D}_2 = \text{diag}(\lambda_1^2, \lambda_2 \lambda_1, \lambda_2^2)$. Then, we have

$$\begin{aligned} |\mathbf{D}_p^+(\mathbf{A} \otimes \mathbf{A})\mathbf{D}_p| &= |\mathbf{D}_p^+(\mathbf{A} \otimes \mathbf{A})\mathbf{D}_p| \\ &= \left(\prod_{i=1}^p \lambda_i^2 \right) \prod_{p \geq i > j \geq 1} \lambda_i \lambda_j = \left(\prod_{i=1}^p \lambda_i \right)^{p+1} = |\mathbf{A}|^{p+1}. \end{aligned}$$

□

Proof 5 of the Wishart density in Theorem 2 The Jacobian of the transformation $\mathbf{S}_\Sigma \rightarrow \mathbf{S}$ or equivalently $\mathbf{s}_\Sigma \rightarrow \mathbf{s}$ is given by Lemma 7 as $|\mathbf{D}_p^+(\Sigma^{1/2} \otimes \Sigma^{1/2})\mathbf{D}_p|_+ = |\Sigma|^{(p+1)/2}$. Consequently, $J(\mathbf{s} \rightarrow \mathbf{s}_\Sigma)$ becomes $|\Sigma|^{-(p+1)/2}$. Then, the pdf of \mathbf{S}_Σ is obtained by that of $\mathbf{S} = \Sigma^{-1/2} \mathbf{S}_\Sigma \Sigma^{-1/2}$ in Theorem 1 and $J(\mathbf{s} \rightarrow \mathbf{s}_\Sigma) = |\Sigma|^{-(p+1)/2}$ as

$$\begin{aligned} w_p(\mathbf{S}_\Sigma | \Sigma, n) &= \frac{\exp\{-\text{tr}(\mathbf{S})/2\} |\mathbf{S}|^{(n-p-1)/2}}{2^{np/2} \Gamma_p(n/2)} |\Sigma|^{-(p+1)/2} \\ &= \frac{\exp\{-\text{tr}(\Sigma^{-1/2} \mathbf{S}_\Sigma \Sigma^{-1/2})/2\} |\Sigma^{-1/2} \mathbf{S}_\Sigma \Sigma^{-1/2}|^{(n-p-1)/2} |\Sigma|^{-(p+1)/2}}{2^{np/2} \Gamma_p(n/2)} \\ &= \frac{\exp\{-\text{tr}(\Sigma^{-1} \mathbf{S}_\Sigma)/2\} |\mathbf{S}_\Sigma|^{(n-p-1)/2}}{2^{np/2} |\Sigma|^{n/2} \Gamma_p(n/2)}. \end{aligned}$$

□

Remark 4 When Lemma 7 for the Jacobian of $\mathbf{S}_\Sigma \rightarrow \mathbf{S}$ is given, Theorem 2 for the Wishart density for general correlated cases was immediately obtained. Conversely, when the Wishart densities for \mathbf{S} and \mathbf{S}_Σ are available, the Jacobian is easily given by comparing two densities using $\mathbf{S} = \Sigma^{-1/2} \mathbf{S}_\Sigma \Sigma^{-1/2}$, which was employed by Anderson (2003, Theorem 7.3.3).

Appendix B On Khatri (1963)'s self-contained concise derivation

Khatri (1963) is referred to only by Kshirsagar (1972, p. 59) and, Srivastava and Khatri (1979, p. 76) to the author's knowledge. The derivation depends on the integral $\pi^{k/2} q^{(k/2)-1} / \Gamma(k/2) = \int_{\mathbf{x}^T \mathbf{x} = q} dx_1 \cdots dx_k$, where q is a positive constant and x_i 's with $\mathbf{x} = (x_1, \dots, x_k)^T$ independently follow the standard normal. This integral is typically obtained in a proof of the chi-square distribution with k df using the surface area $S_{k-1} = 2\pi^{k/2} r^{k-1} / \Gamma(k/2)$ of the $(k-1)$ -sphere with the radius $r = q^{1/2}$ in the k -dimensional Euclidian space and $dr = \{1/(2q^{1/2})\}dq$:

$$\begin{aligned} & \left\{ \prod_{i=1}^k (1/\sqrt{2\pi}) \exp(-x_i^2/2) \Big|_{\mathbf{x}^T \mathbf{x} = q} \right\} \int_{\mathbf{x}^T \mathbf{x} = q} dx_1 \cdots dx_k \\ &= \frac{1}{(2\pi)^{k/2}} \exp\left(-\frac{q}{2}\right) \frac{2\pi^{k/2} r^{k-1} dr}{\Gamma(k/2)} = \frac{1}{(2\pi)^{k/2}} \exp\left(-\frac{q}{2}\right) \frac{2\pi^{k/2} q^{(k-1)/2}}{\Gamma(k/2)} \frac{1}{2q^{1/2}} \\ &= \frac{1}{2^{k/2} \Gamma(k/2)} q^{(k/2)-1} \exp\left(-\frac{q}{2}\right), \end{aligned}$$

yielding

$$\int_{\mathbf{x}^T \mathbf{x} = q} dx_1 \cdots dx_k = \frac{2\pi^{k/2} q^{(k-1)/2}}{\Gamma(k/2)} \frac{1}{2q^{1/2}} = \frac{\pi^{k/2} q^{(k/2)-1}}{\Gamma(k/2)}.$$

Khatri (1963, p. 53) stated that $\int_{\mathbf{x}^T \mathbf{x} = q} dx_1 \cdots dx_k = \pi^{k/2} q^{k/2} / \Gamma(k/2)$ using our notation, where $q^{k/2}$ rather than $q^{(k/2)-1}$ is probably a typo since otherwise the correct factor $|\mathbf{S}|^{(n-p-1)/2}$ corresponding to $q^{(k/2)-1}$ when $k = n - p + 1$ in his subsequent expression of the Wishart density does not follow. An alternative short derivation of $\int_{\mathbf{x}^T \mathbf{x} = q} dx_1 \cdots dx_k$ was given by Ogasawara (2022) as follows. Suppose that the pdf of the chi-square with k df, which is equal to that of the gamma with the shape parameter $k/2$ and the scale parameter 2, is obtained by a different method using e.g., the property of the distribution that the sum of the independent gamma distributed variables with the same scale parameter becomes the gamma with the shape parameter being the sum of those of the gammas and the same scale. Note that the beta integral or the moment generating function can be used for the derivation of this property. Then, we have

$$\left\{ \prod_{i=1}^k (1/\sqrt{2\pi}) \exp(-x_i^2/2) \Big|_{\mathbf{x}^T \mathbf{x} = q} \right\} \int_{\mathbf{x}^T \mathbf{x} = q} dx_1 \cdots dx_k = \frac{q^{(k/2)-1} \exp(-q/2)}{2^{k/2} \Gamma(k/2)},$$

which gives

$$\int_{\mathbf{x}^T \mathbf{x} = q} dx_1 \cdots dx_k = \frac{q^{(k/2)-1} \exp(-q/2) / \{2^{k/2} \Gamma(k/2)\}}{\prod_{i=1}^k (1/\sqrt{2\pi}) \exp(-x_i^2/2) |_{\mathbf{x}^T \mathbf{x} = q}} = \frac{\pi^{k/2} q^{(k/2)-1}}{\Gamma(k/2)}.$$

We find that this derivation without using the area of the $(k-1)$ -sphere is similar to that by [Anderson \(2003\)](#) mentioned in Remark 4.

Proof 6 of the Wishart density in Theorem 2 (Khatri, 1963) Khatri's 1.5-page short derivation is due partially to his concise description. Since the article is less well documented with no title, the citations mentioned earlier using the same incorrect page numbers and several possible typos including the above one for important points and other minor errors, the corrected proof is provided with some added explanations. The derivation consists of a p -step variable transformation with p Jacobians canceling most of them after multiplication.

Define the $p \times n$ matrix \mathbf{X}_Σ , where each column independently follows $N_p(\mathbf{0}, \Sigma)$.

$$\text{Partition } \mathbf{S}_\Sigma = \mathbf{X}_\Sigma \mathbf{X}_\Sigma^T = \begin{bmatrix} \mathbf{S}_{p-1} & \mathbf{s}_{p-1} \\ \mathbf{s}_{p-1}^T & s_{pp} \end{bmatrix} = \begin{bmatrix} \mathbf{X}_{p-1} \mathbf{X}_{p-1}^T & \mathbf{X}_{p-1} \mathbf{x}_p \\ \mathbf{x}_p^T \mathbf{X}_{p-1}^T & \mathbf{x}_p^T \mathbf{x}_p \end{bmatrix}, \text{ where e.g.,}$$

s_{pp} is temporarily used in place of $s_{\Sigma pp}$ for simplicity. Define the $n \times n$ matrix

$$\mathbf{P}_n = \begin{bmatrix} \mathbf{X}_{p-1} \\ \mathbf{Y}_{n-p+1} \end{bmatrix}, \text{ where the } (n-p+1) \times n \text{ submatrix } \mathbf{Y}_{n-p+1} \text{ is chosen such}$$

that $\mathbf{Y}_{n-p+1} \mathbf{X}_{p-1}^T = \mathbf{O}$ and $\mathbf{Y}_{n-p+1} \mathbf{Y}_{n-p+1}^T = \mathbf{I}_{n-p+1}$. Then, we have $\mathbf{P}_n \mathbf{P}_n^T =$

$$\begin{bmatrix} \mathbf{S}_{p-1} & \mathbf{O} \\ \mathbf{O} & \mathbf{I}_{n-p+1} \end{bmatrix}, \text{ which gives } |\mathbf{P}_n|_+ = |\mathbf{P}_n \mathbf{P}_n^T|^{1/2} = |\mathbf{S}_{p-1}|^{1/2}. \text{ Consider the}$$

variable transformation from \mathbf{x}_p to $\mathbf{P}_n \mathbf{x}_p$ with $(\mathbf{s}_{p-1}^T, \mathbf{z}_{n-p+1}^T)^T = \mathbf{P}_n \mathbf{x}_p$, where $\mathbf{z}_{n-p+1} \equiv \mathbf{Y}_{n-p+1} \mathbf{x}_p$ and $J(\mathbf{x}_p \rightarrow \mathbf{P}_n \mathbf{x}_p) = |\mathbf{P}_n|_+^{-1} = |\mathbf{S}_{p-1}|^{-1/2}$. Since

$$\begin{aligned} s_{pp} &= \mathbf{x}_p^T \mathbf{x}_p = (\mathbf{s}_{p-1}^T, \mathbf{z}_{n-p+1}^T) \mathbf{P}_p^T \mathbf{P}_p^{-1} (\mathbf{s}_{p-1}^T, \mathbf{z}_{n-p+1}^T)^T \\ &= (\mathbf{s}_{p-1}^T, \mathbf{z}_{n-p+1}^T) \begin{bmatrix} \mathbf{S}_{p-1}^{-1} & \mathbf{O} \\ \mathbf{O} & \mathbf{I}_{n-p+1} \end{bmatrix} \begin{bmatrix} \mathbf{s}_{p-1} \\ \mathbf{z}_{n-p+1} \end{bmatrix} = \mathbf{s}_{p-1}^T \mathbf{S}_{p-1}^{-1} \mathbf{s}_{p-1} + \mathbf{z}_{n-p+1}^T \mathbf{z}_{n-p+1}, \end{aligned}$$

we have $\mathbf{z}_{n-p+1}^T \mathbf{z}_{n-p+1} = s_{pp} - \mathbf{s}_{p-1}^T \mathbf{S}_{p-1}^{-1} \mathbf{s}_{p-1} = |\mathbf{S}_\Sigma| / |\mathbf{S}_{p-1}|$.

Using the multivariate normal density, the joint marginal density of \mathbf{X}_{p-1} , when a random matrix \mathbf{S}_Σ at Σ is a fixed one, becomes

$$\begin{aligned} f_{\mathbf{X}_{p-1}}(\mathbf{X}_{p-1}) &\equiv f_{\mathbf{X}_{p-1}} \\ &= \int_{-\infty}^{\infty} \cdots \int_{-\infty}^{\infty} (2\pi)^{-np/2} |\Sigma|^{-n/2} \exp\{-\text{tr}(\Sigma^{-1} \mathbf{S}_\Sigma)/2\} \\ &\quad \times J\{\mathbf{x}_p \rightarrow (\mathbf{s}_{p-1}^T, \mathbf{z}_{n-p+1}^T)^T\} dz_1 \cdots dz_{n-p+1} \\ &= \int_{-\infty}^{\infty} (2\pi)^{-np/2} |\Sigma|^{-n/2} \exp\{-\text{tr}(\Sigma^{-1} \mathbf{S}_\Sigma)/2\} |\mathbf{S}_{p-1}|^{-1/2} d\mathbf{z}_{n-p+1}, \end{aligned}$$

where the integrand does not include \mathbf{z}_{n-p+1} . Then, the above integral becomes

$$\begin{aligned}
 f_{\mathbf{X}_{p-1}} &= (2\pi)^{-np/2} |\boldsymbol{\Sigma}|^{-n/2} \exp\{-\text{tr}(\boldsymbol{\Sigma}^{-1} \mathbf{S}_{\Sigma})/2\} |\mathbf{S}_{p-1}|^{-1/2} \\
 &\quad \times \int_{\mathbf{z}_{n-p+1}^T \mathbf{z}_{n-p+1} = |\mathbf{S}_{\Sigma}|/|\mathbf{S}_{p-1}|} d\mathbf{z}_{n-p+1} \\
 &= (2\pi)^{-np/2} |\boldsymbol{\Sigma}|^{-n/2} \exp\{-\text{tr}(\boldsymbol{\Sigma}^{-1} \mathbf{S}_{\Sigma})/2\} |\mathbf{S}_{p-1}|^{-1/2} \\
 &\quad \times \frac{\pi^{(n-p+1)/2} (|\mathbf{S}_{\Sigma}|/|\mathbf{S}_{p-1}|)^{\{(n-p+1)/2\}-1}}{\Gamma\{(n-p+1)/2\}} \\
 &= \frac{\pi^{(n-p+1)/2} |\mathbf{S}_{\Sigma}|^{(n-p-1)/2} \exp\{-\text{tr}(\boldsymbol{\Sigma}^{-1} \mathbf{S}_{\Sigma})/2\}}{(2\pi)^{np/2} |\boldsymbol{\Sigma}|^{n/2} \Gamma\{(n-p+1)/2\}} \frac{1}{|\mathbf{S}_{p-1}|^{(n-p)/2}},
 \end{aligned}$$

where Khatri's (p. 54) expression $|\mathbf{S}_{p-1}|^{(n-p-2)/2}$ using our notation in place of $|\mathbf{S}_{p-1}|^{(n-p)/2}$ is incorrect. Define $\mathbf{X}_{p-i}\{(p-i) \times n\}$, $\mathbf{Y}_{n-p+i}\{(n-p+i) \times n\}$, $\mathbf{S}_{p-i}\{(p-i) \times (p-i)\}$, $\mathbf{s}_{p-i}\{(p-i) \times 1\}$ and $\mathbf{z}_{n-p+i}\{(n-p+i) \times 1\}$ ($i = 2, \dots, p-1$) similarly to those when $i = 1$, respectively. Then, using these matrices and vectors in similar manners, we have the successive transformations as

$$\begin{aligned}
 f_{\mathbf{X}_1} &= \frac{\pi^{(n-p+1)/2} |\mathbf{S}_{\Sigma}|^{(n-p-1)/2} \exp\{-\text{tr}(\boldsymbol{\Sigma}^{-1} \mathbf{S}_{\Sigma})/2\}}{(2\pi)^{np/2} |\boldsymbol{\Sigma}|^{n/2} \Gamma\{(n-p+1)/2\}} \frac{1}{|\mathbf{S}_{p-1}|^{(n-p)/2}} \\
 &\quad \times \prod_{i=2}^{p-1} \int_{\mathbf{z}_{n-p+i}^T \mathbf{z}_{n-p+i} = |\mathbf{S}_{p-i+1}|/|\mathbf{S}_{p-i}|} d\mathbf{z}_{n-p+i} \\
 &= \frac{\pi^{(n-p+1)/2} |\mathbf{S}_{\Sigma}|^{(n-p-1)/2} \exp\{-\text{tr}(\boldsymbol{\Sigma}^{-1} \mathbf{S}_{\Sigma})/2\}}{(2\pi)^{np/2} |\boldsymbol{\Sigma}|^{n/2} \Gamma\{(n-p+1)/2\}} \frac{1}{|\mathbf{S}_{p-1}|^{(n-p)/2}} \\
 &\quad \times \prod_{i=2}^{p-1} \frac{\pi^{(n-p+i)/2} |\mathbf{S}_{p-i+1}|^{(n-p+i-2)/2} / |\mathbf{S}_{p-i}|^{\{(n-p+i-1)/2\}}}{\Gamma\{(n-p+i)/2\}} \\
 &= \frac{\pi^{[(n-p)(p-1) + \{p(p-1)/2\} - np]/2} |\mathbf{S}_{\Sigma}|^{(n-p-1)/2} \exp\{-\text{tr}(\boldsymbol{\Sigma}^{-1} \mathbf{S}_{\Sigma})/2\}}{2^{np/2} |\boldsymbol{\Sigma}|^{n/2} \prod_{i=1}^{p-1} \Gamma\{(n-p+i)/2\}} |\mathbf{S}_1|^{-(n-2)/2} \\
 &= \frac{|\mathbf{S}_{\Sigma}|^{(n-p-1)/2} \exp\{-\text{tr}(\boldsymbol{\Sigma}^{-1} \mathbf{S}_{\Sigma})/2\}}{2^{np/2} |\boldsymbol{\Sigma}|^{n/2} \pi^{p(p-1)/4} \prod_{i=1}^p \Gamma\{(n-p+i)/2\}} \times \frac{(\mathbf{X}_1 \mathbf{X}_1^T)^{-(n-2)/2}}{\pi^{n/2} / \Gamma(n/2)}.
 \end{aligned}$$

Noting that $(\mathbf{X}_1 \mathbf{X}_1^T)^{-(n-2)/2} = |\mathbf{S}_1|^{-(n-2)/2} = s_{11}^{-(n-2)/2}$ is a fixed quantity, the last step is the integral with respect to the row vector \mathbf{X}_1 :

$$\begin{aligned}
 w_p(\mathbf{S}_{\Sigma} | \boldsymbol{\Sigma}, n) &= f_{\mathbf{X}_1} \int_{\mathbf{X}_1 \mathbf{X}_1^T = \mathbf{s}_{11}} d\mathbf{X}_1 = f_{\mathbf{X}_1} \pi^{n/2} s_{11}^{(n/2)-1} / \Gamma(n/2) \\
 &= \frac{|\mathbf{S}_{\Sigma}|^{(n-p-1)/2} \exp\{-\text{tr}(\boldsymbol{\Sigma}^{-1} \mathbf{S}_{\Sigma})/2\}}{2^{np/2} |\boldsymbol{\Sigma}|^{n/2} \pi^{p(p-1)/4} \prod_{i=1}^p \Gamma\{(n-p+i)/2\}}.
 \end{aligned}$$

□

Appendix C The exchanged Bartlett decomposition

The Bartlett decomposition $\mathbf{S} = \mathbf{T}\mathbf{T}^T$ has been used in this paper as well as in literatures. Let $\mathbf{S} = \mathbf{U}\mathbf{U}^T$, where $\mathbf{U} (\neq \mathbf{T}^T)$ is the upper-triangular matrix

whose non-zero elements are random variables. Note that \mathbf{U} can be obtained by rotating \mathbf{T} as $\mathbf{U} = \mathbf{T}\mathbf{V}$ using an orthonormal matrix \mathbf{V} . Define the upper-triangular matrix \mathbf{C} satisfying $\boldsymbol{\Sigma} = \mathbf{C}\mathbf{C}^T$ with $c_{ii} > 0$ ($i = 1, \dots, p$), where \mathbf{C} is obtained by $\mathbf{C} = \mathbf{B}\mathbf{V}^*$ and \mathbf{V}^* is another orthonormal matrix. Recall that the Cholesky decomposition $\boldsymbol{\Sigma} = \mathbf{B}\mathbf{B}^T$ was used earlier. The form $\boldsymbol{\Sigma} = \mathbf{C}\mathbf{C}^T$ is also called the exchanged (reversed) Cholesky or upper-lower (UL) decomposition in this paper.

Remark 5 Consider the distribution of u_{ij} ($i = 1, \dots, p; j = i, \dots, p$), which are assumed to be mutually independent. As in the case of the usual Bartlett, Lemma 1 shows that when u_{ii} is chi-distributed with $n - p + i$ df ($i = 1, \dots, p$) and u_{ij} is standard normal ($i = 1, \dots, p; j = i + 1, \dots, p$), the distribution of $\mathbf{S} = \mathbf{X}\mathbf{X}^T (= \mathbf{T}\mathbf{T}^T)$ is the same as that of $\mathbf{U}\mathbf{U}^T$. Note that t_{ii} is chi-distributed with $n - i + 1$ df rather than $n - p + i$. The joint pdf of \mathbf{U} denoted by $f_p(\mathbf{U})$ becomes

$$\begin{aligned} f_p(\mathbf{U}) &= \left[\prod_{i=1}^p \frac{u_{ii}^{n-p+i-1} \exp(-u_{ii}^2/2)}{2^{\{(n-p+i)/2\}-1} \Gamma\{(n-p+i)/2\}} \right] \\ &\quad \times \frac{1}{(\sqrt{2\pi})^{(p^2-p)/2}} \left\{ \prod_{1 \leq i < j \leq p} \exp(-u_{ij}^2/2) \right\} \\ &= \frac{\left\{ \prod_{i=1}^p u_{ii}^{n-p+i-1} \exp(-u_{ii}^2/2) \right\} \left\{ \prod_{1 \leq i < j \leq p} \exp(-u_{ij}^2/2) \right\}}{2^{\frac{(n-p)p}{2} + \frac{p(p+1)}{4} - p} \times 2^{\frac{p(p-1)}{4}} \pi^{\frac{p(p-1)}{4}} \prod_{i=1}^p \Gamma\{(n-p+i)/2\}} \\ &= \frac{\left(\prod_{i=1}^p u_{ii}^{n-p+i-1} \right) \exp\{-\text{tr}(\mathbf{U}\mathbf{U}^T)/2\}}{2^{\frac{np}{2} - p} \Gamma_p(n/2)}. \end{aligned}$$

Proof 7 of the Wishart density in Theorem 2 Consider the one-step transformation from \mathbf{U} to $\mathbf{S}_\Sigma = \mathbf{C}\mathbf{X}\mathbf{X}^T\mathbf{C}^T = \mathbf{C}\mathbf{S}\mathbf{C}^T = \mathbf{C}\mathbf{U}\mathbf{U}^T\mathbf{C}^T$, where it is found that $\mathbf{C}(\mathbf{X})_j \stackrel{\text{i.i.d.}}{\sim} N_p(\mathbf{0}, \boldsymbol{\Sigma})$ ($j = 1, \dots, n$). Redefine the vector of the non-duplicated elements in \mathbf{S}_Σ as $\mathbf{s}_\Sigma = (s_{\Sigma 11}, \dots, s_{\Sigma 1p}, s_{\Sigma 22}, \dots, s_{\Sigma 2p}, \dots, s_{\Sigma pp})^T$ whose elements are lexicographically ordered. Similarly, define the $\{p(p+1)/2\} \times 1$ vectors \mathbf{c} and \mathbf{u} using the corresponding elements of \mathbf{C} and \mathbf{U} , respectively.

The proof is similar to Proof 1 of Lemma 5. Since \mathbf{C} , \mathbf{U} and $\mathbf{C}\mathbf{U}$ are upper-triangular, the Jacobian matrix $\partial \mathbf{s}_\Sigma / \partial \mathbf{u}^T = \{\partial s_{\Sigma ij} / \partial u_{kl}\}$ ($1 \leq i \leq j \leq p; 1 \leq k \leq l \leq p$) becomes upper-triangular, whose diagonal elements are

$$\frac{\partial s_{\Sigma ij}}{\partial u_{ij}} = \{\mathbf{C}(\mathbf{E}_{ij}\mathbf{U}^T + \mathbf{U}\mathbf{E}_{ji})\mathbf{C}^T\}_{ij} = (\mathbf{C}\mathbf{E}_{ij}\mathbf{U}^T\mathbf{C}^T)_{ij} = c_{ii}u_{jj}c_{jj} \quad (1 \leq i < j \leq p)$$

and

$$\frac{\partial s_{\Sigma ii}}{\partial u_{ii}} = \{\mathbf{C}(\mathbf{E}_{ii}\mathbf{U}^T + \mathbf{U}\mathbf{E}_{ii})\mathbf{C}^T\}_{ii} = 2c_{ii}^2u_{ii} \quad (i = 1, \dots, p).$$

Since the determinant of the Jacobian matrix or $J(\mathbf{S}_\Sigma \rightarrow \mathbf{U})$ becomes

$$\begin{aligned} \prod_{i=1}^p \prod_{j=i}^p \frac{\partial s_{\Sigma ij}}{\partial u_{ij}} &= \left(\prod_{i=1}^p \prod_{j=i+1}^p \frac{\partial s_{\Sigma ij}}{\partial u_{ij}} \right) \prod_{i=1}^p \frac{\partial s_{\Sigma ii}}{\partial u_{ii}} = 2^p \prod_{i=1}^p \prod_{j=i}^p c_{ii} u_{jj} c_{jj} \\ &= 2^p \left(\prod_{i=1}^p c_{ii}^{p-i+1} \right) \prod_{j=1}^p u_{jj}^j c_{jj}^j = 2^p \prod_{i=1}^p c_{ii}^{p+1} u_{ii}^i = 2^p |\mathbf{C}|^{p+1} \prod_{i=1}^p u_{ii}^i \\ &= 2^p |\Sigma|^{(p+1)/2} \prod_{i=1}^p u_{ii}^i, \end{aligned}$$

$J(\mathbf{U} \rightarrow \mathbf{S}_\Sigma)$ is given by the reciprocal of the above quantity.

The Wishart density is given by $f_p(\mathbf{U})$ and $J(\mathbf{U} \rightarrow \mathbf{S}_\Sigma)$:

$$\begin{aligned} w_p(\mathbf{S}_\Sigma | \Sigma, n) &= f_p(\mathbf{U}) J(\mathbf{U} \rightarrow \mathbf{S}_\Sigma) \\ &= \frac{\exp\{-\text{tr}(\mathbf{U}\mathbf{U}^T)/2\} \prod_{i=1}^p u_{ii}^{n-p+i-1} |\Sigma|^{-(p+1)/2}}{2^{(np/2)-p} \Gamma_p(n/2)} \frac{2^p \prod_{i=1}^p u_{ii}^i}{2^{np/2} \Gamma_p(n/2)} \\ &= \frac{\exp\{-\text{tr}(\mathbf{U}\mathbf{U}^T)/2\} |\Sigma|^{-(p+1)/2} \prod_{i=1}^p u_{ii}^{n-p-1}}{2^{np/2} \Gamma_p(n/2)} \\ &= \frac{\exp\{-\text{tr}(\mathbf{C}^{-1} \mathbf{S}_\Sigma \mathbf{C}^{T-1})/2\} |\Sigma|^{-(p+1)/2} |\mathbf{C}^{-1} \mathbf{S}_\Sigma \mathbf{C}^{T-1}|^{(n-p-1)/2}}{2^{np/2} \Gamma_p(n/2)} \\ &= \frac{\exp\{-\text{tr}(\Sigma^{-1} \mathbf{S}_\Sigma)/2\} |\mathbf{S}_\Sigma|^{(n-p-1)/2}}{2^{np/2} |\Sigma|^{n/2} \Gamma_p(n/2)} \end{aligned}$$

as expected. \square

Remark 6 Though $\mathbf{U} \neq \mathbf{T}^T$ as noted earlier, \mathbf{U} is obtained by reversing the row indexes of \mathbf{T} followed by the similar reversal of the column ones. When $p = 3$, this transformation proceeds as

$$\mathbf{T} = \begin{bmatrix} t_{11} & 0 & 0 \\ t_{21} & t_{22} & 0 \\ t_{31} & t_{32} & t_{33} \end{bmatrix} \rightarrow \begin{bmatrix} t_{31} & t_{32} & t_{33} \\ t_{21} & t_{22} & 0 \\ t_{11} & 0 & 0 \end{bmatrix} \rightarrow \begin{bmatrix} t_{33} & t_{32} & t_{31} \\ 0 & t_{22} & t_{21} \\ 0 & 0 & t_{11} \end{bmatrix} \equiv \begin{bmatrix} u_{11} & u_{12} & u_{13} \\ 0 & u_{22} & u_{23} \\ 0 & 0 & u_{33} \end{bmatrix} = \mathbf{U}.$$

The above example indicates other decompositions $\mathbf{S} = \mathbf{T}^* \mathbf{T}^{*T} = \mathbf{U}^* \mathbf{U}^{*T}$ with the unchanged distribution of $\mathbf{S} = \mathbf{X}\mathbf{X}^T$, where $\mathbf{T}^*(\mathbf{U}^*)$ is a lower (upper)-triangular matrix defined with the non-zero elements on and below (above) the minor diagonals. Note that \mathbf{T}^* and \mathbf{U}^* are obtained by \mathbf{T} and \mathbf{U} by reversing

the row or column indexes. When $p = 3$, \mathbf{T}^* and \mathbf{U}^* are $\begin{bmatrix} 0 & 0 & t_{11} \\ 0 & t_{22} & t_{21} \\ t_{33} & t_{32} & t_{31} \end{bmatrix} \equiv$

$$\begin{bmatrix} 0 & 0 & t_{33}^* \\ 0 & t_{22}^* & t_{23}^* \\ t_{31}^* & t_{32}^* & t_{33}^* \end{bmatrix} \text{ and } \begin{bmatrix} u_{13} & u_{12} & u_{11} \\ u_{23} & u_{22} & 0 \\ u_{33} & 0 & 0 \end{bmatrix} \equiv \begin{bmatrix} u_{11}^* & u_{12}^* & u_{13}^* \\ u_{21}^* & u_{22}^* & 0 \\ u_{31}^* & 0 & 0 \end{bmatrix}, \text{ respectively.}$$

Actually, we have infinitely many transformations with the unchanged distribution of \mathbf{S} , including the above ones, using various orthonormal $p \times p$ matrices denoted by \mathbf{V} 's since each column of $\mathbf{V}\mathbf{X}$ independently follows $N_p(\mathbf{0}, \mathbf{I}_p)$ (see

e.g., [Anderson, 2003](#), Theorem 3.3.1). In other words, the distributions of $\mathbf{V}\mathbf{X}$ and \mathbf{X} are the same. Then, $\mathbf{S} = \mathbf{X}\mathbf{X}^T$ can be replaced by $\mathbf{S} = \mathbf{V}\mathbf{X}\mathbf{X}^T\mathbf{V}^T$. Note that one of the decomposed matrices e.g., \mathbf{T} , \mathbf{T}^* , \mathbf{U} and \mathbf{U}^* are given by other ones using \mathbf{V} as $\mathbf{T} = \mathbf{V}\mathbf{U}^*$. This indeterminacy of transformation is similar to the rotational indeterminacy in orthogonal rotation in factor analysis and canonical correlation analysis or more generally transformations in structural equation modeling ([Ogasawara, 2007](#); [Schuberth, 2021](#); [Yu, Schuberth, & Henseler, 2023](#)).