Wishart distributions: Advances in theory with Bayesian application

Andriëtte Bekker[1], Janet van Niekerk[1,*], Mohammad Arashi[2,*]

Abstract

In this paper, we generalize the Wishart distribution utilizing a fresh approach that leads to the hypergeometric Wishart generator distribution with the Wishart generator and the Wishart as special cases. Important statistical characteristics are derived. The significance of this generator distribution is further demonstrated by assuming a special case as a prior for the underlying matrix variate normal model.

Keywords: Bayesian estimation, Frobenius norm, Hypergeometric Wishart, Matrix-variate convergence, Matrix-variate normal, Wishart distribution.

AMS 2000 Classification codes: 62F15, 62H05

1. Introduction

The Wishart distribution, which represents the sum of squares (and cross products) of n draws from a multivariate normal model, and its generalizations are among the most prominent probability distributions in multivariate statistical analysis, arising naturally in applied research and as a basis for theoretical models. The reader is referred to [3] and [14] for a more extensive study regarding the theoretical as well as the practical uses of the Wishart distribution.

---

[*]Corresponding author

Email address: janet.vanniekerk@up.ac.za (Janet van Niekerk)

[1]Department of Statistics, Faculty of Natural and Agricultural Sciences,University of Pretoria, Pretoria, South Africa

[2]Department of Statistics, School of Mathematical Sciences, Shahrood University of Technology, Shahrood, Iran. Department of Statistics, Faculty of Natural and Agricultural Sciences, University of Pretoria, Pretoria, South Africa

Various generalizations and extensions have been proposed for the Wishart distribution because of its importance in matrix theory. To mention a few: Sutradhar and Ali [26] generalized the Wishart distribution for the multivariate elliptical models, while Teng et al. [27] considered matrix-variate elliptical models in their study. Wong and Wang [32] defined the Laplace–Wishart distribution, while Letac and Massam [17] defined the normal quasi-Wishart distribution. In the context of graphical models, Rovetaro [24] defined the hyper-inverse Wishart and Wang and West [31] extended the inverse Wishart distribution for using hyper-Markov properties (see [8]), while Bryc [5] proposed the compound Wishart and q-Wishart in graphical models. Abul-Magd et al. [1] proposed a generalization to Wishart–Laguerre ensembles. Adhikari [2] generalized the Wishart distribution for probabilistic structural dynamics, and Díaz-García and Gutiérrez-Jáimez [9] extended the Wishart distribution for real normed division algebras. Munilla and Cantet [22] also formulated a special structure for the Wishart distribution to apply in modeling the maternal animal. In MIMO systems, Lopez-Martinez et al. [19] applied the complex Wishart distribution and Kumar et al. [16] used the eigenvalue distribution of the sum of Wishart random variates.

The above generalizations and applications justify the study of new models within the Wishart ensemble. In this paper, we propose a construction methodology for creating new matrix variate distributions. The approach is based on a composition of a known statistical distribution combined with a Borel measurable function of the trace operator over the matrix space. Bekker and Roux [4] studied the normal-Wishart prior for the normal model, using this as a springboard for the application of the new developed model as a prior for the covariance matrix of the matrix variate normal model.

The paper is organized as follows: In Section 2, the hypergeometric Wishart generator distribution is proposed with the focus on its characteristics in Section 3. In Section 4, some special cases are highlighted to set the platform for the application as a prior in Section 5. A specific member of this new distribution (derived by [28]) is compared to well-known prior models for the covariance

matrix of the matrix variate normal model. The simulation study illustrates that this prior can outperform the inverse Wishart and Wishart priors, using different measures. Finally, an application with real data is presented.

2. Hypergeometric Wishart generator distribution

The key idea in this section is to propose a generalized form based on a combination of a generalized matrix-variate hypergeometric distribution [28] and the shape generator h. From this form emanates what we call a Wishart generator distribution. Note this latter form contains the cases presented in [6] and [9].

In what follows, we use the following notations and conventions. First, $\boldsymbol{a} = (a_1, \ldots, a_s)$, $\boldsymbol{b} = (b_1, \ldots, b_t)$, and $_sF_t(\cdot; \cdot)$ is the hypergeometric function of matrix argument; see Eq. (1.6.1) on p. 34 of [14]. Furthermore, $C_\kappa(\cdot)$ is the zonal polynomial of order κ [21], $C_\phi^{\kappa,\tau}(\cdot, \cdot)$ is the invariant polynomial [7], $\mathbb{C}$ is the set of complex numbers, S_m is the space of positive definite $m \times m$ matrices, Γ_m is the multivariate gamma function [14] and Γ is Euler's gamma function.

Definition 1. *A random matrix $\boldsymbol{X} \in S_m$ is said to have the hypergeometric Wishart generator distribution (HWGD) with parameters $a_1, \ldots, a_s \in \mathbb{C}$, $b_1, \ldots, b_t \in \mathbb{C}$, $(s \leq t)$, $\boldsymbol{\Omega}, \boldsymbol{\Sigma} \in S_m$, degrees of freedom $n \geq m$ and shape generator $h \neq 1$, if it has the following density function*

$$f(\boldsymbol{X}) = \ell_{n,m} |\boldsymbol{\Sigma}|^{-n/2} |\boldsymbol{X}|^{n/2-(m+1)/2} \, _sF_t(\boldsymbol{a}; \boldsymbol{b}; \boldsymbol{\Omega}\boldsymbol{X}) \, h\{\mathrm{tr}(\boldsymbol{\Sigma}^{-1}\boldsymbol{X})\} \qquad (1)$$

where, from Theorem 1 on p. 480 of [27] and Definition 7.3.1 on p. 258 of [21],

$$
\begin{aligned}
\ell_{n,m}^{-1} &= |\boldsymbol{\Sigma}|^{-n/2} \int_{S_m} |\boldsymbol{X}|^{n/2-(m+1)/2} \, _sF_t(\boldsymbol{a}; \boldsymbol{b}; \boldsymbol{\Omega}\boldsymbol{X}) \, h\{\mathrm{tr}(\boldsymbol{\Sigma}^{-1}\boldsymbol{X})\} d\boldsymbol{X} \\
&= |\boldsymbol{\Sigma}|^{-n/2} \sum_{k=0}^{\infty} \sum_{\kappa} \frac{(a_1)_\kappa \cdots (a_s)_\kappa}{(b_1)_\kappa \cdots (b_t)_\kappa} \frac{1}{k!} \int_{S_m} |\boldsymbol{X}|^{n/2-(m+1)/2} h\{\mathrm{tr}(\boldsymbol{\Sigma}^{-1}\boldsymbol{X})\} C_\kappa(\boldsymbol{\Omega}\boldsymbol{X}) d\boldsymbol{X} \\
&= \Gamma_m(n/2) \sum_{k=0}^{\infty} \sum_{\kappa} \frac{(a_1)_\kappa \cdots (a_s)_\kappa (n/2)_\kappa \gamma_k(n/2)}{(b_1)_\kappa \cdots (b_t)_\kappa k! \, \Gamma(nm/2+k)} C_\kappa(\boldsymbol{\Omega}\boldsymbol{\Sigma}) \qquad (2)
\end{aligned}
$$

We denote this as $\boldsymbol{X} \sim \mathcal{HWG}_m(\boldsymbol{\Sigma}, \boldsymbol{\Omega}, \boldsymbol{a}, \boldsymbol{b}, n, h)$.

Remark 1. The shape generator in Definition 1 should sometimes admit the Taylor series expansion, but this will be mentioned where needed.

Definition 2. *A random matrix $\boldsymbol{X} = \boldsymbol{Y}^{-1} \in S_m$, with $\boldsymbol{Y} \sim \mathcal{HWG}_m(\boldsymbol{\Sigma}, \boldsymbol{\Omega}, \boldsymbol{a}, \boldsymbol{b}, n, h)$ is said to have the inverse hypergeometric Wishart generator distribution (IH-WGD) with parameters $a_1, \ldots, a_s \in \mathbb{C}$, $b_1, \ldots, b_t \in \mathbb{C}$, $(s \leq t)$, $\boldsymbol{\Omega}, \boldsymbol{\Sigma} \in S_m$, degrees of freedom $n \geq m$ and shape generator $h \neq 1$, if it has the following density function*

$$f(\boldsymbol{X}) = \ell_{n,m} |\boldsymbol{\Sigma}|^{-n/2} |\boldsymbol{X}|^{-n/2-(m+1)/2} \; {}_sF_t\left(\boldsymbol{a}; \boldsymbol{b}; \boldsymbol{\Omega}\boldsymbol{X}^{-1}\right) h\{\mathrm{tr}(\boldsymbol{\Sigma}^{-1}\boldsymbol{X}^{-1})\}, \quad (3)$$

with $\ell_{n,m}^{-1}$ as defined in (2). Eq. (3) follows from (1) by making the transformation $\boldsymbol{Y}^{-1} = \boldsymbol{X}$ with Jacobian $J(\boldsymbol{Y} \to \boldsymbol{X}) = |\boldsymbol{X}|^{-(m+1)}$; see Eq. (1.3.13) in [14]. We denote this by $\boldsymbol{X} \sim \mathcal{IHWG}_m(\boldsymbol{\Sigma}, \boldsymbol{\Omega}, \boldsymbol{a}, \boldsymbol{b}, n, h)$.

3. Properties

In the following section some properties of the hypergeometric Wishart generator distribution (1) are derived.

Theorem 1. *For $\boldsymbol{X} \sim \mathcal{HWG}_m(\boldsymbol{\Sigma}, \boldsymbol{\Omega}, \boldsymbol{a}, \boldsymbol{b}, n, h)$, the rth moment of the determinant of $\boldsymbol{X}$ is equal to $\mathrm{E}(|\boldsymbol{X}|^r) = \ell_{n,m}|\boldsymbol{\Sigma}|^r/\ell_{n+2r,m}$, where $\ell_{n,m}$ and $\ell_{n+2r,m}$ are defined in (2).*

Proof. The result follows immediately from Eq. (1). ∎

Theorem 2. *Suppose that $\boldsymbol{X} \sim \mathcal{HWG}_m(\boldsymbol{\Sigma}, \boldsymbol{\Omega}, \boldsymbol{a}, \boldsymbol{b}, n, h)$. The characteristic function of $\boldsymbol{X}$ is given by*

$$\psi_{\boldsymbol{X}}(\boldsymbol{T}) = \ell_{n,m}|\boldsymbol{\Sigma}|^{-n/2} \sum_{\phi} \frac{\theta_{\phi}^{\kappa,\rho}(a_1)_{\rho} \cdots (a_s)_{\rho} h^{(k)}(0)}{k!r!(b_1)_{\rho} \cdots (b_t)_{\rho}}$$

$$\times \Gamma_m(n/2) \, i^{k+r-mn/2} |\boldsymbol{T}|^{-n/2} C_{\phi}^{\kappa,\rho}\left(\boldsymbol{\Sigma}^{-1}\boldsymbol{T}, \boldsymbol{\Omega}\boldsymbol{T}\right),$$

where

$$\sum_{\phi} = \sum_{k=0}^{\infty} \sum_{\kappa} \sum_{r=0}^{\infty} \sum_{\rho} \sum_{\phi \in \kappa, \rho}$$

with $\ell_{n,m}$ as defined in (2) and $\theta_{\phi}^{\kappa,\tau} = C_{\phi}^{\kappa,\tau}(\boldsymbol{I}_m, \boldsymbol{I}_m)/C_{\phi}(\boldsymbol{I}_m)$.

Proof. From (1) follows

$$\psi_{\boldsymbol{X}}(\boldsymbol{T}) = \ell_{n,m} \int_{S_m} |\boldsymbol{\Sigma}|^{-n/2} |\boldsymbol{X}|^{n/2-(m+1)/2} \,_sF_t\left(\boldsymbol{a}; \boldsymbol{b}; \boldsymbol{\Omega X}\right)$$

$$\times \, h\{\operatorname{tr}\left(\boldsymbol{\Sigma}^{-1}\boldsymbol{X}\right)\}\operatorname{etr}\left(i\boldsymbol{TX}\right)d\boldsymbol{X}. \tag{4}$$

Using the Taylor series expansion of h and Eq. (7) on p. 784 of [13], we have

$$_sF_t\left(\boldsymbol{a}; \boldsymbol{b}; \boldsymbol{\Omega X}\right) h\{\operatorname{tr}\left(\boldsymbol{\Sigma}^{-1}\boldsymbol{X}\right)\}$$

$$= \sum_{k=0}^{\infty}\sum_{\kappa}\sum_{r=0}^{\infty}\sum_{\rho} \frac{(a_1)_\rho \cdots (a_s)_\rho h^{(k)}(0)}{k!r!(b_1)_\rho \cdots (b_t)_\rho} C_\kappa\left(\boldsymbol{\Sigma}^{-1}\boldsymbol{X}\right) C_\rho\left(\boldsymbol{\Omega X}\right),$$

with $h^{(k)}(0)$ the kth derivative of h at the point 0. Hence from (4), using Eq. (2.8) on p. 497 of [7] and Eq. (5.13) on p. 22 of [10], we can write

$$\int_{S_m} |\boldsymbol{X}|^{n/2-(m+1)/2} \,_sF_t\left(\boldsymbol{a}; \boldsymbol{b}; \boldsymbol{\Omega X}\right) h\{\operatorname{tr}\left(\boldsymbol{\Sigma}^{-1}\boldsymbol{X}\right)\}\operatorname{etr}\left(i\boldsymbol{TX}\right)d\boldsymbol{X}$$

$$= \sum_{k=0}^{\infty}\sum_{\kappa}\sum_{r=0}^{\infty}\sum_{\rho} \frac{(a_1)_\rho \cdots (a_s)_\rho h^{(k)}(0)}{k!r!(b_1)_\rho \cdots (b_t)_\rho}$$

$$\times \int_{S_m} |\boldsymbol{X}|^{n/2-(m+1)/2} C_\kappa\left(\boldsymbol{\Sigma}^{-1}\boldsymbol{X}\right) C_\rho\left(\boldsymbol{\Omega X}\right)\operatorname{etr}\left(i\boldsymbol{TX}\right)d\boldsymbol{X},$$

which is the same as

$$\sum_{\phi} \frac{\theta_\phi^{\kappa,\rho}(a_1)_\rho \cdots (a_s)_\rho h^{(k)}(0)}{k!r!(b_1)_\rho \cdots (b_t)_\rho} \int_{S_m} |\boldsymbol{X}|^{n/2-(m+1)/2} C_\phi^{\kappa,\rho}\left(\boldsymbol{\Sigma}^{-1}\boldsymbol{X}, \boldsymbol{\Omega X}\right)\operatorname{etr}\left(i\boldsymbol{TX}\right)d\boldsymbol{X}$$

$$= \sum_{\phi} \frac{\theta_\phi^{\kappa,\rho}(a_1)_\rho \cdots (a_s)_\rho h^{(k)}(0)}{k!r!(b_1)_\rho \cdots (b_t)_\rho} \Gamma_m\left(n/2\right) |i\boldsymbol{T}|^{-n/2} C_\phi^{\kappa,\rho}\left(i\boldsymbol{\Sigma}^{-1}\boldsymbol{T}, i\boldsymbol{\Omega T}\right),$$

and the result follows. ∎

Theorem 3. Let $\boldsymbol{X} \sim \mathcal{HWG}_m(\boldsymbol{\Sigma}, \boldsymbol{\Omega}, \boldsymbol{a}, \boldsymbol{b}, n, h)$, and $\boldsymbol{A} \in S_m$. Then $\boldsymbol{AXA}^\top \sim \mathcal{HWG}_m\left(\boldsymbol{A\Sigma A}^\top, \boldsymbol{A}^{\top-1}\boldsymbol{\Omega A}^{-1}, \boldsymbol{a}, \boldsymbol{b}, n, h\right)$.

Proof. Note that the Jacobian of the transformation $\boldsymbol{Y} = \boldsymbol{AXA}^\top$ is given by $J(\boldsymbol{X} \to \boldsymbol{Y}) = |\boldsymbol{A}|^{-(m+1)}$, and the result follows easily. ∎

Remark 2. From Theorem 3, one can conclude that if $\boldsymbol{X} \sim \mathcal{HWG}_m(\boldsymbol{\Sigma}, \boldsymbol{\Omega}, \boldsymbol{a}, \boldsymbol{b}, n, h)$ with $\boldsymbol{\Sigma} = \boldsymbol{A}^{-1}\boldsymbol{A}^{\top-1}$ and $\boldsymbol{\Omega} = \boldsymbol{A}^\top \boldsymbol{A}$, then $\boldsymbol{AXA}^\top \sim \mathcal{HWG}_m\left(\boldsymbol{I}, \boldsymbol{I}, \boldsymbol{a}, \boldsymbol{b}, n, h\right)$.

Theorem 4. *Let $\boldsymbol{X} \sim \mathcal{HWG}_m(\boldsymbol{\Sigma}, \boldsymbol{\Omega}, \boldsymbol{a}, \boldsymbol{b}, n, h)$. The joint density function of the diagonal matrix of the eigenvalues of $\boldsymbol{X}$, i.e., $\boldsymbol{\Lambda} = \mathrm{diag}(\lambda_1, \ldots, \lambda_m)$, $\lambda_m > \cdots > \lambda_1 > 0$ is given by*

$$g(\boldsymbol{\Lambda}) = \frac{\ell_{n,m} \pi^{m^2/2}}{\Gamma_m(m/2)} \prod_{i<j}^m (\lambda_i - \lambda_j) |\boldsymbol{\Sigma}|^{-n/2} \prod_{i=1}^m \lambda_i^{n/2-(m+1)/2}$$

$$\times \sum_\phi \theta_\phi^{\kappa,\rho} \frac{(a_1)_\rho \cdots (a_s)_\rho h^{(k)}(0) \, C_\phi^{\kappa,\rho}(\boldsymbol{\Sigma}^{-1}, \boldsymbol{\Omega}) C_\phi(\boldsymbol{\Lambda})}{k! r! (b_1)_\rho \cdots (b_t)_\rho C_\phi(\boldsymbol{I}_m)},$$

with $\ell_{n,m}$ as defined in (2).

Proof. From Theorem 3.2.17 on p. 104 of [21] the density function of $\boldsymbol{\Lambda}$ is given by

$$g(\boldsymbol{\Lambda}) = \frac{\pi^{m^2/2}}{\Gamma_m(m/2)} \prod_{i<j}^m (\lambda_i - \lambda_j) \int_{O_m} f(\boldsymbol{H}\boldsymbol{\Lambda}\boldsymbol{H}^\top) d\boldsymbol{H} \tag{5}$$

for any $\boldsymbol{H} \in O_m = \{\boldsymbol{H}_{m \times m} | \boldsymbol{H}^\top \boldsymbol{H} = \boldsymbol{H}\boldsymbol{H}^\top = \boldsymbol{I}_m\}$, where $d\boldsymbol{H}$ is the normalised Haar measure on O_m (see p. 72 of [21]) with O_m the space of orthogonal $m \times m$ matrices. Note that from (1),

$$\int_{O_m} f(\boldsymbol{H}\boldsymbol{\Lambda}\boldsymbol{H}^\top) d\boldsymbol{H} = \ell_{n,m} |\boldsymbol{\Sigma}|^{-n/2} |\boldsymbol{\Lambda}|^{n/2-(m+1)/2}$$

$$\times \int_{O_m} h\{\mathrm{tr}\left(\boldsymbol{\Sigma}^{-1}\boldsymbol{H}\boldsymbol{\Lambda}\boldsymbol{H}^\top\right)\} \, {}_sF_t\left(\boldsymbol{a}; \boldsymbol{b}; \boldsymbol{\Omega}\boldsymbol{H}\boldsymbol{\Lambda}\boldsymbol{H}^\top\right) d\boldsymbol{H}. \tag{6}$$

Since h admits a Taylor expansion and using Eqs. (2.4) and (2.8) on pp. 466–467 of [7] as well as Eq. (7) on p. 784 of [13], we can write

$$\int_{O_m} h\{\mathrm{tr}\left(\boldsymbol{\Sigma}^{-1}\boldsymbol{H}\boldsymbol{\Lambda}\boldsymbol{H}^\top\right)\} \, {}_sF_t\left(\boldsymbol{a}; \boldsymbol{b}; \boldsymbol{\Omega}\boldsymbol{H}\boldsymbol{\Lambda}\boldsymbol{H}^\top\right) d\boldsymbol{H}$$

$$= \sum_{k=0}^\infty \sum_\kappa \sum_{r=0}^\infty \sum_\rho \frac{(a_1)_\rho \cdots (a_s)_\rho h^{(k)}(0)}{k! r! (b_1)_\rho \cdots (b_t)_\rho}$$

$$\times \int_{O_m} C_\kappa\left(\boldsymbol{\Sigma}^{-1}\boldsymbol{H}\boldsymbol{\Lambda}\boldsymbol{H}^\top\right) C_\rho\left(\boldsymbol{\Omega}\boldsymbol{H}\boldsymbol{\Lambda}\boldsymbol{H}^\top\right) d\boldsymbol{H},$$

which reduces to

$$\sum_\phi \frac{\theta_\phi^{\kappa,\rho}(a_1)_\rho \cdots (a_s)_\rho h^{(k)}(0)}{k! r! (b_1)_\rho \cdots (b_t)_\rho} \int_{O_m} C_\phi^{\kappa,\rho}\left(\boldsymbol{\Sigma}^{-1}\boldsymbol{H}\boldsymbol{\Lambda}\boldsymbol{H}^\top, \boldsymbol{\Omega}\boldsymbol{H}\boldsymbol{\Lambda}\boldsymbol{H}^\top\right) d\boldsymbol{H}$$

$$= \sum_\phi \frac{\theta_\phi^{\kappa,\rho}(a_1)_\rho \cdots (a_s)_\rho h^{(k)}(0)}{k! r! (b_1)_\rho \cdots (b_t)_\rho} \frac{C_\phi^{\kappa,\rho}(\boldsymbol{\Sigma}^{-1}, \boldsymbol{\Omega}) C_\phi(\boldsymbol{\Lambda})}{C_\phi(\boldsymbol{I}_m)}. \tag{7}$$
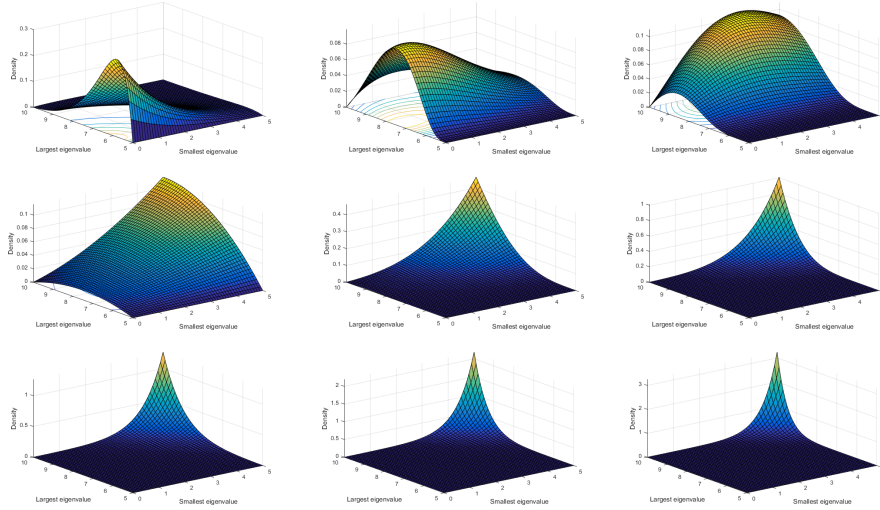
Figure 1: Joint density function of the largest and smallest eigenvalue for $m = 2$ and $n = 5$ (Left), $n = 10$ (Middle) and $n = 15$ (Right) with $c_1 = 1; c_2 = -2$ (Top), $c_2 = -1$ (Middle) and $c_2 = 0$ (Bottom).

The proof is complete from (5), (6) and (7) . ∎

Remark 3. Let $\boldsymbol{\Sigma} = c_1 \boldsymbol{I}_m, \boldsymbol{\Omega} = c_2 \boldsymbol{I}_m$. Then

$$g\left(\boldsymbol{\Lambda}\right) = \frac{\ell_{n,m} \pi^{m^2/2}}{\Gamma_m\left(m/2\right)} \prod_{i<j}^{m} \left(\lambda_i - \lambda_j\right) c_1^{-nm/2}$$

$$\times \prod_{i=1}^{m} \lambda_i^{n/2 - (m+1)/2} h\left(c_1^{-1} \mathrm{tr} \boldsymbol{\Lambda}\right) \, {}_sF_t\left(\boldsymbol{a}; \boldsymbol{b}; c_2 \boldsymbol{\Lambda}\right). \quad (8)$$

Figure 1 illustrates the joint density function of the eigenvalues — see (8) — for $m = 2, \boldsymbol{\Sigma} = c_1 \boldsymbol{I}_m, \boldsymbol{\Omega} = c_2 \boldsymbol{I}_m, s = 0, t = 0$, and $h(\cdot) = \exp(\cdot)$ for specific values of the parameters.

4. Special cases

This section focuses on special cases of (1) and link them to existing literature. Table 1 lists some members of the hypergeometric Wishart generator distribution; see (1). The hypergeometric Wishart distribution, as given in Table 1, is considered further in Section 5.

Distribution	$f(\boldsymbol{X})$	$s,t,\boldsymbol{\Omega}$	$h(x)$
Matrix-variate t	$\dfrac{\Gamma\left(\frac{nm}{2}+p\right)}{\Gamma_m\left(\frac{n}{2}\right)\Gamma(p)}\lvert\boldsymbol{\Sigma}\rvert^{-\frac{n}{2}}$ $\times\lvert\boldsymbol{X}\rvert^{\frac{n}{2}-\frac{m+1}{2}}\left\{1+\operatorname{tr}\left(\boldsymbol{\Sigma}^{-1}\boldsymbol{X}\right)\right\}^{-\left(\frac{nm}{2}+p\right)}$	$s=t=0, \boldsymbol{\Omega}=\boldsymbol{0}$	$(1+x)^{-\left(\frac{nm}{2}+p\right)}$
Power Wishart	$\dfrac{\Gamma(\frac{nm}{2})ba^{\frac{nm}{2b}}}{\Gamma_m(\frac{n}{2})\Gamma\left(\frac{nm}{2b}\right)}\lvert\boldsymbol{\Sigma}\rvert^{-\frac{n}{2}}$ $\times\lvert\boldsymbol{X}\rvert^{\frac{n}{2}-\frac{m+1}{2}}\exp\left[-a\{\operatorname{tr}(\boldsymbol{\Sigma}^{-1}\boldsymbol{X})\}^{b}\right]$	$s=t=0, \boldsymbol{\Omega}=\boldsymbol{0}$	$\exp\left(-ax^{b}\right)$
Logistic Wishart	$\dfrac{\frac{nm}{2}ab^{\frac{nm}{2}}}{\Gamma_m\left(\frac{n}{2}\right)}\lvert\boldsymbol{\Sigma}\rvert^{-\frac{n}{2}}\lvert\boldsymbol{X}\rvert^{\frac{n}{2}-\frac{m+1}{2}}$ $\times\operatorname{etr}\left(-b\boldsymbol{\Sigma}^{-1}\boldsymbol{X}\right)\left\{1-a\operatorname{etr}\left(-b\boldsymbol{\Sigma}^{-1}\boldsymbol{X}\right)\right\}^{-2}$	$s=t=0, \boldsymbol{\Omega}=-b\boldsymbol{\Sigma}^{-1}$	$\{1-a\exp(-bx)\}^{-2}$
Sin Wishart	$\dfrac{2\Gamma\left(\frac{nm}{2}\right)a^{\frac{nm+2}{4}}\exp\left(\frac{b^2}{4a}\right)}{b\Gamma_m\left(\frac{n}{2}\right)\Gamma\left(\frac{nm+2}{4}\right){}_1F_1\left(1-\frac{nm}{4};\frac{3}{2};\frac{b^2}{4a}\right)}\lvert\boldsymbol{\Sigma}\rvert^{-\frac{n}{2}}$ $\times\lvert\boldsymbol{X}\rvert^{\frac{n}{2}-\frac{m+1}{2}}\exp\{-a\operatorname{tr}\left(\boldsymbol{\Sigma}^{-1}\boldsymbol{X}\right)^{2}\}\sin\{b\operatorname{tr}\left(\boldsymbol{\Sigma}^{-1}\boldsymbol{X}\right)\}$	$s=t=0, \boldsymbol{\Omega}=\boldsymbol{0}$	$\exp\left(-ax^2\right)\sin(bx)$
Logarithmic Wishart	$\dfrac{\Gamma\left(\frac{nm}{2}\right)}{\Gamma_m\left(\frac{n}{2}\right)\Gamma\left(\frac{nm}{2}\right)}\lvert\boldsymbol{\Sigma}\rvert^{-\frac{n}{2}}\lvert\boldsymbol{X}\rvert^{\frac{n}{2}-\frac{m+1}{2}}$ $\times\operatorname{etr}\left(-\boldsymbol{\Sigma}^{-1}\boldsymbol{X}\right)\ln\{\operatorname{tr}\left(\boldsymbol{\Sigma}^{-1}\boldsymbol{X}\right)\}$	$s=t=0, \boldsymbol{\Omega}=-\boldsymbol{\Sigma}^{-1}$	$\ln(x)$
Hypergeometric Wishart	$\dfrac{1}{\Gamma_m(\frac{n}{2}){}_{v+1}F_w\left(\frac{nm}{2},a_1,\ldots,a_v;b_1,\ldots,b_w;c\right)}\lvert\boldsymbol{\Sigma}\rvert^{-\frac{n}{2}}\lvert\boldsymbol{X}\rvert^{\frac{n}{2}-\frac{m+1}{2}}$ $\times{}_vF_w\{a_1,\ldots,a_v;b_1,\ldots,b_w;c\operatorname{tr}(\boldsymbol{\Sigma}^{-1}\boldsymbol{X})\}\operatorname{etr}(-\boldsymbol{\Sigma}^{-1}\boldsymbol{X})$	$s=t=0, \boldsymbol{\Omega}=-\boldsymbol{\Sigma}^{-1}$	${}_vF_w(a_1,\ldots,a_v;b_1,\ldots,b_w;cx)$

Table 1: Some special cases of the hypergeometric Wishart generator distribution (1)

Case 1: Let $s = 0$, $t = 1$ and $b_1 > 0$, then (1) simplifies to

$$f(\boldsymbol{X}) = \ell_{n,m}|\boldsymbol{\Sigma}|^{-n/2}|\boldsymbol{X}|^{n/2-(m+1)/2} \, _0F_1\left(\frac{n}{2}; \boldsymbol{\Omega X}\right) h\{\mathrm{tr}(\boldsymbol{\Sigma}^{-1}\boldsymbol{X})\}. \quad (9)$$

The normalizing constant is given by

$$\ell_{n,m}^{-1} = \Gamma_m\left(n/2\right) \sum_{k=0}^{\infty} \sum_{\kappa} \frac{\gamma_k(n/2)}{k!\Gamma(nm/2+k)} \, C_{\kappa}(\boldsymbol{\Omega}),$$

The density function (9) is termed the non-central Wishart generator distribution (NWGD) and denoted as $\boldsymbol{X} \sim \mathcal{NWG}_m(\boldsymbol{\Sigma}, \boldsymbol{\Omega}, b_1, n, h)$. The density function of $\boldsymbol{Z}_{m \times m} = \boldsymbol{Y}^{\top}\boldsymbol{Y}$, where $\boldsymbol{Y}_{p \times m}$ has a matrix elliptical distribution, i.e., $\boldsymbol{Y} \sim \mathcal{EC}(\boldsymbol{M}, \boldsymbol{\Sigma}, g)$ derived by [27] (see also [9] and [10]) is of the same functional form as in Definition 1 if we take $s = 0$, $t = 1$, $b_1 = n/2$, $\boldsymbol{\Omega} = \boldsymbol{\Sigma}^{-1/2}\boldsymbol{\Upsilon}\boldsymbol{\Sigma}^{-1/2}$ and $h(x) = g^{(2)}(x + \mathrm{tr}\boldsymbol{\Upsilon})$.

Case 2: Note that letting $s = 0, t = 0$ and $\boldsymbol{\Omega} = \boldsymbol{0}$ in Definition 1, the density function (1) simplifies to

$$f(\boldsymbol{X}) = \ell_{n,m}|\boldsymbol{\Sigma}|^{-n/2}|\boldsymbol{X}|^{n/2-(m+1)/2}h\{\mathrm{tr}(\boldsymbol{\Sigma}^{-1}\boldsymbol{X})\} \quad (10)$$

with

$$\ell_{n,m}^{-1} = \frac{\Gamma_m(n/2)\gamma_0(n/2)}{\Gamma(nm/2)}, \quad \gamma_0\left(n/2\right) = \int_{\mathbb{R}^+} y^{nm/2-1}h(y)dy$$

from Theorem 1 on p. 480 of [27], provided that the above integral exists. The density function in (10) is termed the Wishart generator distribution (WGD) and denoted as $\boldsymbol{X} \sim \mathcal{WG}_m(\boldsymbol{\Sigma}, n, h)$. The name of the distribution is motivated by the fact that for $s = 0, t = 0, \boldsymbol{\Omega} = \boldsymbol{0}$ and $h(x) = \exp(-x/2)$ in Definition 1, it yields the Wishart distribution (see [3]) with

$$\ell_{n,m} = \frac{\Gamma(nm/2)}{\Gamma_m(n/2)\gamma_0(n/2)} \quad \text{and} \quad \gamma_0\left(n/2\right) = 2^{nm/2}\Gamma\left(nm/2\right),$$

denoted as $\boldsymbol{X} \sim \mathcal{W}_m(\boldsymbol{\Sigma}, n)$. This specific form (10) is also called the Wishart elliptical distribution by [6].

Some properties of the Wishart generator distribution (10) are given below. Suppose that $\boldsymbol{X} \sim \mathcal{WG}_m(\boldsymbol{\Sigma}, n, h)$.

1. The rth moment of the determinant of $\boldsymbol{X}$ is equal to

$$
\mathrm{E}\left(|\boldsymbol{X}|^{\mathrm{r}}\right) \;=\; \frac{\Gamma\left(mn/2\right)\Gamma_m\left(r+n/2\right)\gamma_0\left(r+n/2\right)}{\Gamma\left(mr+mn/2\right)\Gamma_m\left(n/2\right)\gamma_0\left(n/2\right)}\,|\boldsymbol{\Sigma}|^r.
$$

2. The characteristic function of $\boldsymbol{X}$ is given by

$$
\psi_{\boldsymbol{X}}(\boldsymbol{T}) = \frac{\Gamma\left(nm/2\right)}{\gamma_0\left(n/2\right)} \sum_{k=0}^{\infty}\sum_{\kappa} \frac{\gamma_k\left(n/2\right)(n/2)_\kappa}{k!\Gamma\left(mn/2+k\right)}\, C_\kappa\left(i\boldsymbol{T}\boldsymbol{\Sigma}\right).
$$

3. The joint density function of the eigenvalues $\boldsymbol{\Lambda} = \operatorname{diag}(\lambda_1,\ldots,\lambda_m)$, $\lambda_m > \cdots > \lambda_1 > 0$, of $\boldsymbol{X}$ is given by

$$
\begin{aligned}
g\left(\boldsymbol{\Lambda}\right) \;=\; & \frac{\pi^{m^2/2}\Gamma(nm/2)}{\Gamma_m\left(m/2\right)\Gamma_m(n/2)\gamma_0(n/2)} \prod_{i<j}^{m}\left(\lambda_i-\lambda_j\right)|\boldsymbol{\Sigma}|^{-n/2} \\
& \times \prod_{i=1}^{m}\lambda_i^{n/2-(m+1)/2}\sum_{k=0}^{\infty}\sum_{\kappa} \frac{h^{(k)}\left(0\right)C_\kappa\left(\boldsymbol{\Sigma}^{-1}\right)C_\kappa\left(\boldsymbol{\Lambda}\right)}{k!C_\kappa\left(\boldsymbol{I}_m\right)}.
\end{aligned}
$$

4. The distribution of statistics is important in hypothesis testing in order to compute p-values and power. Therefore, we give the distributions of statistics of two independent components of matrix variates where one has the Wishart generator distribution. Let $\boldsymbol{\Sigma} = \alpha\boldsymbol{\Psi}$ and $\boldsymbol{Y} \sim \mathcal{W}_m(\beta\boldsymbol{\Psi}, p)$, where α and β are enriching parameters. Then

    (i) The random variable $\boldsymbol{B}_1 = \boldsymbol{X}^{-1/2}\boldsymbol{Y}\boldsymbol{X}^{-1/2}$ has the following density function

$$
\begin{aligned}
g(\boldsymbol{B}_1) = & \frac{\Gamma_m\{(n+p)/2\}}{\Gamma_m\left(p/2\right)\Gamma_m(n/2)}\left(\frac{\alpha}{\beta}\right)^{pm/2}|\boldsymbol{\Psi}|^{-(n+p)}|\boldsymbol{B}_1|^{p/2-(m+1)/2} \\
& \times \mathrm{E}_{\boldsymbol{D}}\left\{\operatorname{etr}\left(-\frac{1}{2\beta}\,\boldsymbol{\Psi}^{-1}\boldsymbol{D}^{1/2}\boldsymbol{B}_1\boldsymbol{D}^{1/2}\right)\right\}
\end{aligned}
$$

with $\boldsymbol{D} \sim \mathcal{WG}_m(\alpha\boldsymbol{\Psi}, n+p, h)$ as in (10). $\mathrm{E}_{\boldsymbol{D}}(\cdot)$ denotes the expected value with respect to the distribution of $\boldsymbol{D}$. Alternatively

$$
\begin{aligned}
g(\boldsymbol{B}_1) = & \frac{\Gamma_m\left\{(n+p)/2\right\}}{2^{(n+p)m/2}\Gamma_m\left(p/2\right)\Gamma_m(n/2)}\left(\frac{\alpha}{\beta}\right)^{pm/2}|\boldsymbol{B}_1|^{p/2-(m+1)/2} \\
& \times \sum_{k=0}^{\infty}\sum_{\kappa}\frac{1}{k!}\left(-\frac{\alpha}{2\beta}\right)^k \frac{\left(\frac{n+p}{2}\right)_\kappa \gamma_\kappa\left(\frac{n+p}{2}\right)}{\Gamma\left\{(mn+mp)/2+k\right\}}C_\kappa(\boldsymbol{B}_1)
\end{aligned}
$$

(ii) The random variable $\boldsymbol{B}_2 = (\boldsymbol{X} + \boldsymbol{Y})^{-1/2}\boldsymbol{X}(\boldsymbol{X} + \boldsymbol{Y})^{-1/2}$ has the following density function

$$g(\boldsymbol{B}_2) = \frac{\Gamma_m\{(n + p)/2\}}{\Gamma_m\,(p/2)\,\Gamma_m(n/2)} \left(\frac{\alpha}{\beta}\right)^{pm/2} |\boldsymbol{B}_2|^{-p/2-(m+1)/2}$$
$$\times |\boldsymbol{I}_m - \boldsymbol{B}_2|^{p/2-(m+1)/2}\mathrm{E}_{\boldsymbol{B}}\left[\mathrm{etr}\left\{-\frac{1}{2\beta}\,\boldsymbol{B}(\boldsymbol{I}_\mathrm{m} - \boldsymbol{B}_2)\right\}\right]$$

with $\boldsymbol{B} \sim WG_m(\alpha\boldsymbol{B}_2^{-1}, n + p, h)$ as in (10) and $\boldsymbol{\Psi} = \boldsymbol{I}$.

5. Application

To emphasize the contribution of model (1), a special case of the hypergeometric Wishart generator distribution, namely the hypergeometric Wishart distribution (see Table 1, [28]), is applied as a subjective prior for the covariance matrix of the matrix-variate normal model.

The conjugate inverse Wishart prior for the covariance matrix has been considered and applied in the literature by many authors; see, e.g., [23]. The use of the Wishart distribution as a prior was investigated in [4, 29, 30]. The good performance of the Wishart prior paves the way for the speculative research of other unconventional prior distributions for the covariance matrix of the matrix variate normal model.

Consider the matrix variate normal model, $\boldsymbol{X}_{p\times m} \sim \mathcal{N}_{p\times m}(\boldsymbol{\mu}_{p\times m}, \boldsymbol{\Phi}_{p\times p} \otimes \boldsymbol{\Sigma}_{m\times m})$ with density function given by

$$f(\boldsymbol{X}) = (2\pi)^{-mp/2}\,|\boldsymbol{\Phi}|^{-m/2}|\boldsymbol{\Sigma}|^{-p/2}\mathrm{etr}\left\{-\frac{1}{2}\,(\boldsymbol{X} - \boldsymbol{\mu})^\top\boldsymbol{\Phi}^{-1}\,(\boldsymbol{X} - \boldsymbol{\mu})\,\boldsymbol{\Sigma}^{-1}\right\}$$

with $\boldsymbol{\Phi}$ a known $p \times p$ matrix and $\boldsymbol{\mu}$ and $\boldsymbol{\Sigma}$ unknown. Now consider an objective prior for $\boldsymbol{\mu}$, and the hypergeometric Wishart distribution (Definition 1 with $s = 0, t = 0$ and $h(x) = {}_vF_w(a_1, \ldots, a_v; b_1, \ldots, b_w; cx)$ as in Table 1) with parameters $\boldsymbol{\Theta}, a_1, \ldots, a_v, b_1, \ldots, b_w, c$ and n as the subjective prior for $\boldsymbol{\Sigma}$, denoted as $\boldsymbol{\Sigma} \sim \mathcal{HW}_m(\boldsymbol{\Theta}, a_1, \ldots, a_v, b_1, \ldots, b_w, c, n)$, such that the joint prior density

11

function is

$$\pi\left(\boldsymbol{\mu}, \boldsymbol{\Sigma}\right) = \frac{|\boldsymbol{\Theta}|^{-n/2}|\boldsymbol{\Sigma}|^{n/2-(m+1)/2}}{\Gamma_m\left(n/2\right) {}_{v+1}F_w\left(nm/2, a_1, \ldots, a_v; b_1, \ldots, b_w; c\right)}$$

$$\times {}_vF_w\{a_1, \ldots, a_v, b_1, \ldots, b_w; c \operatorname{tr}\left(\boldsymbol{\Theta}^{-1}\boldsymbol{\Sigma}\right)\}\operatorname{etr}\left(-\boldsymbol{\Theta}^{-1}\boldsymbol{\Sigma}\right). \quad (11)$$

For $v = w = 0$ and $c = 1/2$ in (11), the prior density function of $\boldsymbol{\Sigma}$ is the density of a Wishart random matrix with parameter $\boldsymbol{\Theta}$ and n degrees of freedom.

5.1. Posterior density functions and Bayes estimators

Based on a single observation, $\boldsymbol{X}$, the joint posterior density function is

$$q\left(\boldsymbol{\mu}, \boldsymbol{\Sigma}\right) \quad \propto \quad |\boldsymbol{\Sigma}|^{(n-p)/2-(m+1)/2}\operatorname{etr}\left\{-\frac{1}{2}\left(\boldsymbol{X} - \boldsymbol{\mu}\right)^{\top}\boldsymbol{\Phi}^{-1}\left(\boldsymbol{X} - \boldsymbol{\mu}\right)\boldsymbol{\Sigma}^{-1}\right\}$$

$$\times {}_vF_w\{a_1, \ldots, a_v, b_1, \ldots, b_w; c \operatorname{tr}\left(\boldsymbol{\Theta}^{-1}\boldsymbol{\Sigma}\right)\}\operatorname{etr}\left(-\boldsymbol{\Theta}^{-1}\boldsymbol{\Sigma}\right).$$

The marginal posterior density function of $\boldsymbol{\mu}$ is given by

$$q(\boldsymbol{\mu}|\boldsymbol{x}) \quad \propto \quad \int_{S_m} |\boldsymbol{\Sigma}|^{(n-p)/2-(m+1)/2}\operatorname{etr}\left\{-\frac{1}{2}\left(\boldsymbol{x} - \boldsymbol{\mu}\right)^{\top}\boldsymbol{\Phi}^{-1}\left(\boldsymbol{x} - \boldsymbol{\mu}\right)\boldsymbol{\Sigma}^{-1} - \boldsymbol{\Theta}^{-1}\boldsymbol{\Sigma}\right\}$$

$$\times {}_vF_w\{a_1, \ldots, a_v, b_1, \ldots, b_w; c \operatorname{tr}\left(\boldsymbol{\Theta}^{-1}\boldsymbol{\Sigma}\right)\}d\boldsymbol{\Sigma}$$

$$\propto \quad \mathrm{E}_{\boldsymbol{\Sigma}^*}\left[\operatorname{etr}\left\{-\frac{1}{2}\left(\boldsymbol{x} - \boldsymbol{\mu}\right)^{\top}\boldsymbol{\Phi}^{-1}\left(\boldsymbol{x} - \boldsymbol{\mu}\right)\boldsymbol{\Sigma}^{*-1}\right\}\right.$$

$$\left. \times {}_vF_w\{a_1, \ldots, a_v, b_1, \ldots, b_w; c \operatorname{tr}\left(\boldsymbol{\Theta}^{-1}\boldsymbol{\Sigma}^*\right)\}\right],$$

where $\boldsymbol{\Sigma}^* \sim \mathcal{W}_m\left(\boldsymbol{\Theta}/2, n - p\right).$ Hence

$$q\left(\boldsymbol{\mu}|\boldsymbol{x}\right) = C_{\boldsymbol{\mu}}\mathrm{E}_{\boldsymbol{\Sigma}^*}\left[\operatorname{etr}\left\{-\frac{1}{2}\left(\boldsymbol{x} - \boldsymbol{\mu}\right)^{\top}\boldsymbol{\Phi}^{-1}\left(\boldsymbol{x} - \boldsymbol{\mu}\right)\boldsymbol{\Sigma}^{*-1}\right\}\right.$$

$$\left. \times {}_v F_w\{a_1, \ldots, a_v, b_1, \ldots, b_w; c \operatorname{tr}\left(\boldsymbol{\Theta}^{-1}\boldsymbol{\Sigma}^*\right)\}\right],$$

where

$$C_{\boldsymbol{\mu}}^{-1} = \int_{\boldsymbol{\mu}} \mathrm{E}_{\boldsymbol{\Sigma}^*}\left[\operatorname{etr}\left\{-\frac{1}{2}\left(\boldsymbol{x} - \boldsymbol{\mu}\right)^{\top}\boldsymbol{\Phi}^{-1}\left(\boldsymbol{x} - \boldsymbol{\mu}\right)\boldsymbol{\Sigma}^{*-1}\right\}\right.$$

$$\left. \times {}_v F_w\{a_1, \ldots, a_v, b_1, \ldots, b_w; c \operatorname{tr}\left(\boldsymbol{\Theta}^{-1}\boldsymbol{\Sigma}^*\right)\}\right]d\boldsymbol{\mu}$$

can be rewritten as

$$C_{\boldsymbol{\mu}}^{-1} = \int_{S_m} \frac{1}{\Gamma_m\{(n-p)/2\}2^{m(n-p)/2}} |\boldsymbol{\Sigma}|^{(n-p)/2-(m+1)/2} |\frac{1}{2}\boldsymbol{\Theta}|^{-(n-p)/2} \mathrm{etr}\left(-\boldsymbol{\Theta}^{-1}\boldsymbol{\Sigma}\right)$$

$$\times \int_{\boldsymbol{\mu}} \mathrm{etr}\left\{-\frac{1}{2}\left(\boldsymbol{x}-\boldsymbol{\mu}\right)^{\top}\boldsymbol{\Phi}^{-1}(\boldsymbol{x}-\boldsymbol{\mu})\boldsymbol{\Sigma}^{-1}\right\}d\boldsymbol{\mu}$$

$$\times_v F_w\{a_1,\ldots,a_v,b_1,\ldots,b_w; c\,\mathrm{tr}\left(\boldsymbol{\Theta}^{-1}\boldsymbol{\Sigma}\right)\}d\boldsymbol{\Sigma}$$

and hence

$$C_{\boldsymbol{\mu}}^{-1} = \frac{(2\pi)^{mp/2} |\boldsymbol{\Phi}|^{m/2}\Gamma_m(n/2)|\boldsymbol{\Theta}|^{p/2}}{\Gamma_m\{(n-p)/2\}} \, \mathrm{E}_{\boldsymbol{\Sigma}^1}[_v\mathrm{F}_w\{a_1,\ldots,a_v,b_1,\ldots,b_w; c\,\mathrm{tr}\left(\boldsymbol{\Theta}^{-1}\boldsymbol{\Sigma}^*\right)\}],$$

where $\boldsymbol{\Sigma}^1 \sim \mathcal{W}_m\left(\boldsymbol{\Theta}/2, n\right)$. Therefore,

$$q\left(\boldsymbol{\mu}|\boldsymbol{x}\right) = \frac{\Gamma_m\{(n-p)/2\}\,(2\pi)^{-mp/2}\,|\boldsymbol{\Phi}|^{-m/2}|\boldsymbol{\Theta}|^{-p/2}}{\Gamma_m(n/2)\mathrm{E}_{\boldsymbol{\Sigma}^1}[_v\mathrm{F}_w\{a_1,\ldots,a_v,b_1,\ldots,b_w; c\,\mathrm{tr}\left(\boldsymbol{\Theta}^{-1}\boldsymbol{\Sigma}^*\right)\}]}$$

$$\times \mathrm{E}_{\boldsymbol{\Sigma}^*}\left[\mathrm{etr}\left\{-\frac{1}{2}\left(\boldsymbol{x}-\boldsymbol{\mu}\right)^{\top}\boldsymbol{\Phi}^{-1}\left(\boldsymbol{x}-\boldsymbol{\mu}\right)\boldsymbol{\Sigma}^{*-1}\right\}\right.$$

$$\left.\times_v F_w\{a_1,\ldots,a_v,b_1,\ldots,b_w; c\,\mathrm{tr}\left(\boldsymbol{\Theta}^{-1}\boldsymbol{\Sigma}^*\right)\}\right],$$

where $\boldsymbol{\Sigma}^* \sim \mathcal{W}_m\left(\boldsymbol{\Theta}/2, n-p\right)$ and $\boldsymbol{\Sigma}^1 \sim \mathcal{W}_m\left(\boldsymbol{\Theta}/2, n\right)$.

The marginal posterior density function of $\boldsymbol{\Sigma}$ is given by

$$q\left(\boldsymbol{\Sigma}|\boldsymbol{x}\right) = C_{\boldsymbol{\Sigma}}|\boldsymbol{\Sigma}|^{(n-p)/2-(m+1)/2} {}_v F_w\{a_1,\ldots,a_v,b_1,\ldots,b_w; c\,\mathrm{tr}\left(\boldsymbol{\Theta}^{-1}\boldsymbol{\Sigma}\right)\}$$

$$\times \mathrm{etr}\left(-\boldsymbol{\Theta}^{-1}\boldsymbol{\Sigma}\right) \int_{\boldsymbol{\mu}} \mathrm{etr}\left\{-\frac{1}{2}\left(\boldsymbol{x}-\boldsymbol{\mu}\right)^{\top}\boldsymbol{\Phi}^{-1}\left(\boldsymbol{x}-\boldsymbol{\mu}\right)\boldsymbol{\Sigma}^{-1}\right\}d\boldsymbol{\mu}$$

$$\propto |\boldsymbol{\Sigma}|^{n/2-(m+1)/2} {}_v F_w\{a_1,\ldots,a_v,b_1,\ldots,b_w; c\,\mathrm{tr}\left(\boldsymbol{\Theta}^{-1}\boldsymbol{\Sigma}\right)\}$$

$$\times \mathrm{etr}\left(-\boldsymbol{\Theta}^{-1}\boldsymbol{\Sigma}\right)$$

with

$$C_{\boldsymbol{\Sigma}}^{-1} = 2^{nm/2}\Gamma_m\left(\frac{n}{2}\right)\left|\frac{\boldsymbol{\Theta}}{2}\right|^{n/2}\mathrm{E}_{\boldsymbol{\Sigma}_1}[_v\mathrm{F}_w\{a_1,\ldots,a_v,b_1,\ldots,b_w; c\,\mathrm{tr}\left(\boldsymbol{\Theta}^{-1}\boldsymbol{\Sigma}_1\right)\}],$$

where $\boldsymbol{\Sigma}_1 \sim \mathcal{W}_m\left(\boldsymbol{\Theta}/2, n\right)$. Hence

$$q\left(\boldsymbol{\Sigma}|\boldsymbol{x}\right) = \frac{|\boldsymbol{\Theta}|^{-n/2}|\boldsymbol{\Sigma}|^{n/2-(m+1)2}}{\Gamma_m\left(n/2\right)\mathrm{E}_{\boldsymbol{\Sigma}_1}[\,_v\mathrm{F}_w\{a_1,\ldots,a_v,b_1,\ldots,b_w; c\,\mathrm{tr}\left(\boldsymbol{\Theta}^{-1}\boldsymbol{\Sigma}_1\right)\}]}$$

$$\times_v F_w\{a_1,\ldots,a_v,b_1,\ldots,b_w; c\,\mathrm{tr}\left(\boldsymbol{\Theta}^{-1}\boldsymbol{\Sigma}\right)\}\mathrm{etr}\left(-\boldsymbol{\Theta}^{-1}\boldsymbol{\Sigma}\right).$$

Remark 4. Note that

$$q\left(\boldsymbol{\mu}|\boldsymbol{\Sigma}\right) \propto \operatorname{etr}\left\{-\frac{1}{2}\left(\boldsymbol{\mu}-\boldsymbol{x}\right)^{\top}\boldsymbol{\Phi}^{-1}\left(\boldsymbol{\mu}-\boldsymbol{x}\right)\boldsymbol{\Sigma}^{-1}\right\},$$

i.e., $\boldsymbol{\mu}|\boldsymbol{\Sigma},\boldsymbol{x} \sim \mathcal{N}(\boldsymbol{x},\boldsymbol{\Phi}_{p\times p}\otimes\boldsymbol{\Sigma}_{m\times m})$ and

$$q\left(\boldsymbol{\Sigma}|\boldsymbol{\mu}\right) \propto |\boldsymbol{\Sigma}|^{(n-p)/2-(m+1)/2}\operatorname{etr}\left\{-\frac{1}{2}\left(\boldsymbol{X}-\boldsymbol{\mu}\right)^{\top}\boldsymbol{\Phi}^{-1}\left(\boldsymbol{X}-\boldsymbol{\mu}\right)\boldsymbol{\Sigma}^{-1}-\boldsymbol{\Theta}^{-1}\boldsymbol{\Sigma}\right\}$$
$$\times\ {}_{v}F_{w}\{a_{1},\ldots,a_{v},b_{1},\ldots,b_{w};c\ \operatorname{tr}\left(\boldsymbol{\Theta}^{-1}\boldsymbol{\Sigma}\right)\}.$$

Under the squared error loss function, the Bayes estimator is the posterior mean value. Hence the Bayes estimator of $\boldsymbol{\mu}$ is

$$\widehat{\boldsymbol{\mu}} = \operatorname{E}(\boldsymbol{\mu}|\boldsymbol{X}) \tag{12}$$

since

$$\operatorname{E}(\boldsymbol{\mu}-\boldsymbol{X}|\boldsymbol{X}) = \int_{\boldsymbol{\mu}}\left(\boldsymbol{\mu}-\boldsymbol{X}\right)\operatorname{q}\left(\boldsymbol{\mu}|\boldsymbol{X}\right)\operatorname{d}\boldsymbol{\mu}$$
$$= \frac{\Gamma_{m}\{(n-p)/2\}\left(2\pi\right)^{-mp/2}|\boldsymbol{\Phi}|^{-m/2}|\boldsymbol{\Theta}|^{-p/2}}{\Gamma_{m}(n/2)\operatorname{E}_{\boldsymbol{\Sigma}^{1}}\left[\ {}_{v}F_{w}\{a_{1},\ldots,a_{v},b_{1},\ldots,b_{w};c\ \operatorname{tr}\left(\boldsymbol{\Theta}^{-1}\boldsymbol{\Sigma}^{*}\right)\}\right]}$$
$$\times\operatorname{E}_{\boldsymbol{\Sigma}^{*}}\left[\ {}_{v}F_{w}\{a_{1},\ldots,a_{v},b_{1},\ldots,b_{w};c\ \operatorname{tr}\left(\boldsymbol{\Theta}^{-1}\boldsymbol{\Sigma}^{*}\right)\}\right]$$
$$\times\int_{\boldsymbol{\mu}}\left(\boldsymbol{\mu}-\boldsymbol{X}\right)\operatorname{etr}\left\{-\frac{1}{2}\left(\boldsymbol{X}-\boldsymbol{\mu}\right)^{\top}\boldsymbol{\Phi}^{-1}\left(\boldsymbol{X}-\boldsymbol{\mu}\right)\boldsymbol{\Sigma}^{*-1}\right\}d\boldsymbol{\mu}.$$

It is quite clear that the integrand is an odd function and (12) follows. As for the Bayes estimator of $|\boldsymbol{\Sigma}|^{r}$, it is given by

$$\operatorname{E}(|\boldsymbol{\Sigma}|^{\mathrm{r}}|\boldsymbol{X}) \quad = \quad \frac{{}_{v+1}F_{w}\{(n/2+r)\,m,a_{1},\ldots,a_{v},b_{1},\ldots,b_{w};c\}\Gamma_{m}(n/2+r)}{\Gamma_{m}\left(n/2\right)\operatorname{E}_{\boldsymbol{\Sigma}_{1}}\left[{}_{v}F_{w}\{a_{1},\ldots,a_{v},b_{1},\ldots,b_{w};c\ \operatorname{tr}\left(\boldsymbol{\Theta}^{-1}\boldsymbol{\Sigma}_{1}\right)\}\right]}$$

from Eq. 7.522(5), on p. 814 of [12], where $\boldsymbol{\Sigma}_{1} \sim \mathcal{W}_{m}\left(\boldsymbol{\Theta}/2,n\right).$

5.2. Simulation study

In this section the results of the preceding subsection will be applied to a simulated dataset to illustrate the advantage of the hypergeometric Wishart distribution as a prior for the covariance matrix of the matrix variate normal model.

5.2.1. Set-up/Preliminaries

A sample of size n is simulated from a multivariate normal distribution with a dimensionality of m, zero mean vector and identity covariance matrix, i.e., $\boldsymbol{X}_{1\times m} \sim \mathcal{N}_m(\boldsymbol{0}_{1\times m}, \boldsymbol{I}_{m\times m})$. The assumed priors are the inverse Wishart, Wishart and hypergeometric Wishart distributions, i.e., $\boldsymbol{\Sigma} \sim \mathcal{IW}_m(\boldsymbol{\Theta}_1, n_1)$, $\boldsymbol{\Sigma} \sim \mathcal{W}_m(\boldsymbol{\Theta}_1, n_2)$ and $\boldsymbol{\Sigma} \sim \mathcal{HW}_m(\boldsymbol{\Theta}_1, a_1, b_1, c, n_3)$, respectively, with $\boldsymbol{\Theta}_1 = 4 \times \boldsymbol{I}_m$, $m = 3, n_1 = 9.5$, $n_2 = 3, a_1 = 1, b_1 = 2, c = 1, n = 20, n_3 = 3$.

Posterior samples of size 10,000 are simulated using a Gibbs sampling scheme with an additional Metropolis–Hastings algorithm [15], adapted from [29, 30]. The convergence of the Markov chain of matrices is evaluated graphically using three measures, the determinant, trace and largest eigenvalue of the simulated matrix. Figure 2 illustrates these three measures for the chain. It is clear from Figure 2 that the algorithm converges.
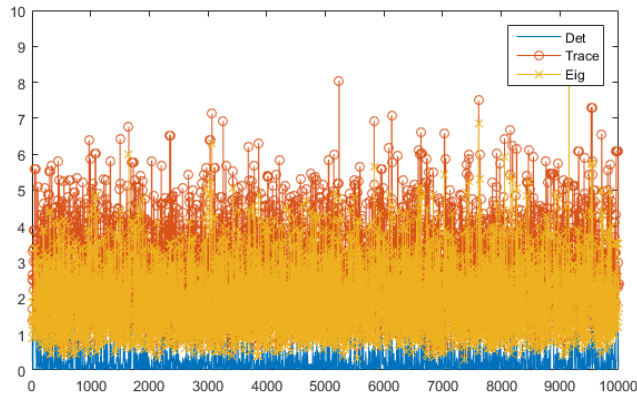

Figure 2: Determinant, trace and largest eigenvalue of matrices simulated from the Gibbs sampler

5.2.2. Results

The estimates calculated for $\mathbf{\Sigma}$ under the three different priors as well as the MLE are

$$\widehat{\mathbf{\Sigma}}_{MLE} = \begin{bmatrix} 3.089 & 1.852 & -0.669 \\ 1.852 & 4.158 & 2.612 \\ -0.669 & 2.612 & 7.191 \end{bmatrix}, \quad \widehat{\mathbf{\Sigma}}_{\mathcal{IW}} = \begin{bmatrix} 15.542 & -6.885 & -0.345 \\ -6.885 & 58.719 & -18.033 \\ -0.345 & -18.033 & 26.675 \end{bmatrix},$$

$$\widehat{\mathbf{\Sigma}}_{\mathcal{W}} = \begin{bmatrix} 1.226 & -0.211 & 0.208 \\ -0.211 & 1.228 & 0.0998 \\ 0.208 & 0.0998 & 0.735 \end{bmatrix}, \quad \widehat{\mathbf{\Sigma}}_{\mathcal{HW}} = \begin{bmatrix} 0.877 & 0.042 & -0.05 \\ 0.042 & 0.963 & 0.005 \\ -0.05 & 0.005 & 1.0026 \end{bmatrix}.$$

The above estimates are obtained for one posterior sample. The Frobenius norm [11] of the errors, defined as

$$||\widehat{\mathbf{\Sigma}} - \mathbf{\Sigma}||_F = \sqrt{\text{tr}\{(\widehat{\mathbf{\Sigma}} - \mathbf{\Sigma})^\top(\widehat{\mathbf{\Sigma}} - \mathbf{\Sigma})\}},$$

is calculated for each estimate and given in Table 2.

Frobenius norm	Value
$||\widehat{\mathbf{\Sigma}}_{MLE} - \mathbf{\Sigma}||_F$	0.5241
$||\widehat{\mathbf{\Sigma}}_{\mathcal{IW}} - \mathbf{\Sigma}||_F$	66.547
$||\widehat{\mathbf{\Sigma}}_{\mathcal{W}} - \mathbf{\Sigma}||_F$	0.4479
$||\widehat{\mathbf{\Sigma}}_{\mathcal{HW}} - \mathbf{\Sigma}||_F$	0.1151

Table 2: Error norms calculated from the simulated sample

The hypergeometric Wishart prior results in the smallest Frobenius norm of the error. For further investigation, this sampling scheme is repeated 100 times to obtain 100 estimates under each prior as well as the MLE for each of the 100 simulated samples. The Frobenius norm of the error for each estimate (excluding the inverse-Wishart due to poor performance) and every repetition is calculated and illustrated in Figure 3.

It is clear that the hypergeometric Wishart prior results in the lowest Frobenius norm and therefore produces less error than the other three estimators.
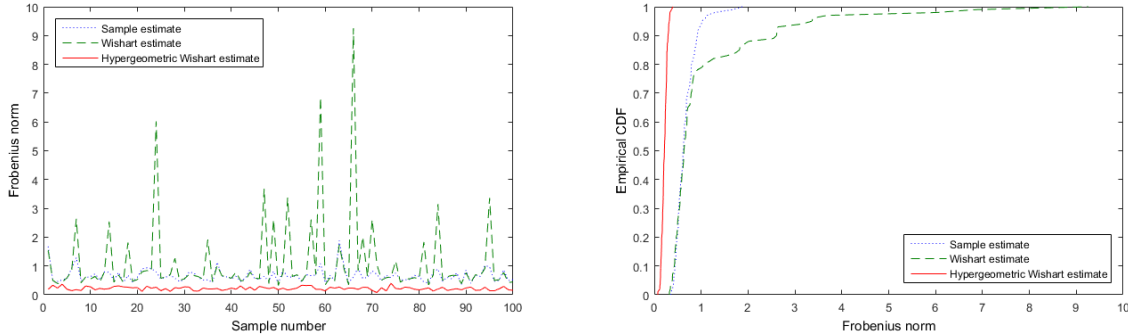
Figure 3: Frobenius norm of the estimation errors (left) and The empirical cumulative distribution function (ECDF) of the Frobenius norm of the estimation errors (right) for 100 repititions

Additionally, the empirical cumulative distribution function (ECDF) of each set of Frobenius norms calculated for each estimator is obtained and displayed in Figure 3. The ECDF which is the leftmost in the figure is regarded as the best since for a specific value of the error norm, a higher proportion of estimates from that particular prior results in less error. From Figure 3 we conclude that the hypergeometric Wishart prior results in an estimate of $\boldsymbol{\Sigma}$ with the least error and preference should be given to this prior. To validate the graphical interpretation, a two-sample Kolmogorov–Smirnov test is performed, pairwise, on the four different ECDF's and the p-value for each pair is calculated as given in Table 3.

Pairwise comparison	p-value
$\mathcal{W}$ and $\mathcal{HW}$	< 0.001
MLE and $\mathcal{HW}$	< 0.001

Table 3: p-values of the Kolmogorov–Smirnov two-sample test based on samples of the Frobenius norms

From Table 3 it is clear that the ECDF of the errors under the hypergeometric Wishart prior is significantly different from the other priors. Therefore, the assertion can be made that the hypergeometric Wishart prior structure produces

17

an estimate with frequentist superiority.

5.3. Abalone dataset

The age of abalone is determined by cutting the shell through the cone, staining it, and counting the number of rings through a microscope. This is a very time-consuming task. Other measurements which can be obtained easily can give an indication of the number of rings. The abalone dataset is from [18] and contains eight variables and has a sample size of 4177. We will use the six continuous variables: length-longest shell measurement, diameter-perpendicular to length, height, whole weight-weight of whole abalone, shucked weight-weight of the meat, viscera weight-gut weight after bleeding and shell weight-weight after being dried. Only a subsample of size 20 will be used due to the scarcity of abalone and since the focus of the study is estimation for small sample sizes. Thus we have a multivariate sample of dimension 6 and sample size 20. The subsample is tested for multivariate normality using Royston's test [25]. Hence, $\boldsymbol{X}_{1\times6} \sim \mathcal{N}(\boldsymbol{\mu}_{1\times6}, \boldsymbol{\Sigma}_{6\times6})$.

The assumed priors are $\boldsymbol{\Sigma} \sim \mathcal{IW}_m(\boldsymbol{\Theta}_{IW}, n_{IW}), \boldsymbol{\Sigma} \sim \mathcal{W}_m(\boldsymbol{\Theta}_W, n_W)$ and $\boldsymbol{\Sigma} \sim \mathcal{HW}_m(\boldsymbol{\Theta}_{HW}, a_1, b_1, c, n_{HW})$, respectively. The hyperparametrs are chosen as $\boldsymbol{\Theta}_{IW} = 4\boldsymbol{S}_1, n_{IW} = 9, \boldsymbol{\Theta}_W = 0.25\boldsymbol{S}_1, n_W = 3, \boldsymbol{\Theta}_{HW} = 0.5\boldsymbol{S}_1, a_1 = 1, b_1 = 2, c = 0.9, n_{HW} = 3$, with $\boldsymbol{S}_1$ equal to the sample covariance matrix of the available data excluding the 20 observations in the chosen subsample.

The Bayes estimate calculated for $\boldsymbol{\Sigma}$ under these priors using the Gibbs sampler, as well as the MLE are

$$\widehat{\boldsymbol{\Sigma}}_{MLE} = \begin{bmatrix} .014 & .012 & .004 & .054 & .015 & .012 \\ .012 & .01 & .003 & .045 & .013 & .01 \\ .004 & .003 & .002 & .017 & .005 & .004 \\ .054 & .045 & .017 & .24 & .065 & .052 \\ .015 & .013 & .005 & .065 & .019 & .014 \\ .012 & .01 & .004 & .052 & .014 & .012 \end{bmatrix},$$

$$\widehat{\boldsymbol{\Sigma}}_{\mathcal{IW}} = \begin{bmatrix} .006 & .005 & .002 & .026 & .001 & .006 \\ .005 & .004 & .002 & .022 & .006 & .005 \\ .002 & .002 & .001 & .008 & .002 & .002 \\ .026 & .022 & .008 & .120 & .033 & .027 \\ .007 & .006 & .002 & .033 & .009 & .007 \\ .006 & .005 & .002 & .027 & .007 & .006 \end{bmatrix},$$

$$\widehat{\boldsymbol{\Sigma}}_{\mathcal{W}} = \begin{bmatrix} .009 & .008 & .003 & .038 & .01 & .008 \\ .008 & .006 & .002 & .032 & .009 & .007 \\ .003 & .002 & .001 & .011 & .003 & .003 \\ .038 & .032 & .011 & .186 & .047 & .038 \\ .01 & .009 & .003 & .047 & .013 & .01 \\ .008 & .007 & .003 & .038 & .01 & .009 \end{bmatrix},$$

and

$$\widehat{\boldsymbol{\Sigma}}_{\mathcal{HW}} = \begin{bmatrix} .011 & .009 & .003 & .045 & .012 & .01 \\ .009 & .008 & .003 & .038 & .01 & .008 \\ .003 & .003 & .001 & .013 & .004 & .003 \\ .045 & .038 & .013 & .209 & .055 & .045 \\ .012 & .01 & .004 & .055 & .015 & .012 \\ .01 & .008 & .003 & .045 & .012 & .011 \end{bmatrix}.$$

Based on these estimates of $\boldsymbol{\Sigma}$, a multivariate normal dataset of size 20 is simulated for each estimate, i.e., $\boldsymbol{Y}_{i,1\times 6} \sim \mathcal{N}_6(\overline{\boldsymbol{X}}, \widehat{\boldsymbol{\Sigma}}_Q), Q = (MLE, \mathcal{IW}, \mathcal{W}, \mathcal{HW})$, and visually illustrated as an image where each observation forms a row and each variable a column of pixels. An entry is used as the intensity for a pixel. This allows for visual inspection of the plausibility of the estimates since the true value of $\boldsymbol{\Sigma}$ is unknown.

Figure 4 displays the images based on the real dataset and the simulated samples using the calculated estimates. The image closest to the image based on the real dataset, is indicative of the most plausible estimate of $\boldsymbol{\Sigma}$.

It is clear from Figure 4, that the sample based on the estimate of $\boldsymbol{\Sigma}$ under the hypergeometric Wishart prior is closest to the real dataset. Additionally,
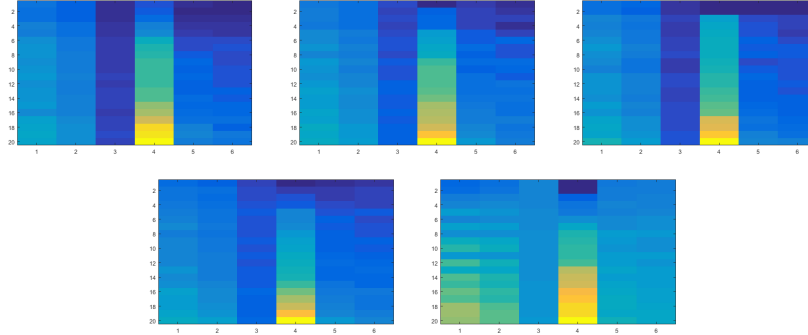
Figure 4: Images based on the real dataset (top left) and simulated samples - based on MLE (top center), $\mathcal{HW}$ (top right), $\mathcal{W}$ (bottom left), $\mathcal{IW}$ (bottom right)

the Mahalanobis distance [20] was calculated for each observation between the real dataset and each of the simulated datasets. The mean, per sample, of these distances is given in Table 4.

Estimate used	MLE	$\mathcal{IW}$	$\mathcal{W}$	$\mathcal{HW}$
Mean Mahalanobis distance	17.3247	29.0114	7.0470	5.4523

Table 4: Mean Mahalanobis distance between the real dataset and each simulated sample

Table 4 supports the findings of Figure 4 since the distance between the real dataset and the simulated sample based on $\widehat{\mathbf{\Sigma}}_{HW}$ is the smallest. Thus, the estimate under the hypergeometric Wishart prior, $\widehat{\mathbf{\Sigma}}_{HW}$, fits the data best. The estimate under the Wishart prior is also plausible from Figure 4 and Table 4. These results support the findings in Section 5.2.2.

6. Conclusion

In this paper we defined the hypergeometric Wishart generator distribution by combining a Borel measurable function of a trace operator over matrix space with the kernel of the generalized matrix variate hypergeometric distribution. Several statistical properties of this newly defined distribution were studied. The Wishart generator distribution is a member of this class and might be impor-

tant for a number of practical signal processing applications including synthetic aperture radar (SAR), multi-antenna wireless communication and direct imaging of extra-solar planets. The Bayesian analysis of the matrix variate normal model demonstrated the novelty of this new model by considering, specifically, the hypergeometric Wishart distribution, as a prior for the covariance matrix.

[1] A. Abul-Magd, G. Akemann, P. Vivo, Superstatistical generalizations of Wishart–Laguerre ensembles of random matrices, Journal of Physics A: Mathematical and Theoretical 42 (2009) 175–207.

[2] S. Adhikari, Generalized Wishart distribution for probabilistic structural dynamics, Computational Mechanics 45 (2010) 495–511.

[3] T.W. Anderson, An Introduction to Multivariate Statistical Analysis, Wiley, New York, 2003.

[4] A. Bekker, J. Roux, Bayesian multivariate normal analysis with a Wishart prior, Communications in Statistics — Theory and Methods 24 (1995) 2485–2497.

[5] W. Bryc, Compound real Wishart and q-Wishart matrices, International Mathematics Research Notices 2008 (2008) rnn079.

[6] F. Caro-Lopera, G. González-Farías, N. Balakrishnan, On generalized Wishart distributions, I: Likelihood ratio test for homogeneity of covariance matrices, Sankhyā: The Indian Journal of Statistics, Series A 76 (2014) 179–194.

[7] A.W. Davis, Invariant polynomials with two matrix arguments extending the zonal polynomials: Applications to multivariate distribution theory, Annals of the Institute of Statistical Mathematics 31 (1979) 465 – 485.

[8] A.P. Dawid, S. Lauritzen, Hyper Markov laws in the statistical analysis of decomposable graphical models, The Annals of Statistics 21 (1993) 1272–1317.

[9] J. Díaz-García, R. Gutiérrez-Jáimez, On Wishart distribution: Some extensions, Linear Algebra and its Applications 435 (2011) 1296–1310.

[10] J. Díaz-García, R.Jáimez, Wishart and pseudo-Wishart distributions under elliptical laws and related distributions in the shape theory context, Journal of Statistical Planning and Inference 136 (2006) 4176–4193.

[11] G. Golub, C.V. Loan, Matrix computations, Johns Hopkins University, Press, Baltimore, MD (1996) pp. 374–426.

[12] I. Gradshteyn, I. Ryzhik, Table of integrals, series and products, Academic Press, 2007.

[13] K. Gross, D. Richards, Special functions of matrix argument. I. Algebraic induction, zonal polynomials, and hypergeometric functions, Transactions of the American Mathematical Society 301 (1987) 781–811.

[14] A. Gupta, D. Nagar, Matrix variate distributions, Chapman & Hall, London, 2000.

[15] W. Hastings, Monte Carlo sampling methods using Markov chains and their applications, Biometrika 57 (1970) 97–109.

[16] S. Kumar, G. Pivaro, G. Fraidenraich, C. Dias, On the exact and approximate eigenvalue distribution for sum of Wishart matrices, arXiv preprint arXiv:1504.00222.

[17] G. Letac, H. Massam, The normal quasi-Wishart distribution, Contemporary Mathematics 287 (2001) 231–240.

[18] M. Lichman, (uci) machine learning repository (2013).
URL http://archive.ics.uci.edu/ml

[19] F. Lopez-Martinez, E. Martos-Naya, J. Paris, A. Goldsmith, Eigenvalue dynamics of a central Wishart matrix with application to MIMO systems, IEEE Transactions on Information Theory 61 (2015) 2693–2707.

[20] P. Mahalanobis, On the generalized distance in statistics, Proceedings of the National Institute of Sciences (Calcutta) 2 (1936) 49–55.

[21] R. Muirhead, Aspects of Multivariate Statistical Theory, Wiley, Hoboken, NJ, 1982.

[22] S. Munilla, R. Cantet, Bayesian conjugate analysis using a generalized inverted Wishart distribution accounts for differential uncertainty among the genetic parameters — An application to the maternal animal model, Journal of Animal Breeding and Genetics 129 (2012) 173–187.

[23] S. Press, Applied Multivariate Analysis: Including Bayesian and Frequentist Methods of Inference, Melbourne, FL: Krieger Publishing Company.

[24] A. Roverato, Hyper inverse Wishart distribution for non-decomposable graphs and its application to Bayesian inference for gaussian graphical models, Scandinavian Journal of Statistics 29 (2002) 391–411.

[25] J. Royston, Some techniques for assessing multivariate normality based on the Shapiro–Wilk w, Applied Statistics (1983) 121–133.

[26] B. Sutradhar, M. Ali, A generalization of the Wishart distribution for the elliptical model and its moments for the multivariate t model, Journal of Multivariate Analysis 29 (1989) 155–162.

[27] C. Teng, H. Fang, W. Deng, The generalized noncentral Wishart distribution, Journal of Mathematical Research and Exposition, 9, 1989, 479–488.

[28] G. Van der Merwe, J. Roux, On generalized matrix-variate hypergeometric distribution, South African Statistical Journal 8 (1974) 49–58.

[29] J. Van Niekerk, A. Bekker, M. Arashi, D. De Waal, Estimation under the matrix-variate elliptical model, South African Statistical Journal 50 (2016) 149–171.

[30] J. Van Niekerk, A. Bekker, M. Arashi, J. Roux, Subjective Bayesian analysis of the elliptical model, Communications in Statistics — Theory and Methods 44 (2015) 3738–3753.

[31] H. Wang, M. West, Bayesian analysis of matrix normal graphical models, Biometrika 96 (2009) 821–834.

[32] C. Wong, T. Wang, Laplace–Wishart distributions and Cochran theorems, Sankhyā, Series A (1995) 342–359.