Developments in Wishart ensemble and Bayesian application

by

Janet van Niekerk

Supervisor:
Prof. Andriëtte Bekker

Co-supervisor:
Prof. Mohammad Arashi

Submitted in partial fulfillment of the requirements for the degree Doctor of Philosophy

in the

Faculty of Natural and Agricultural Sciences

University of Pretoria

Pretoria

February 2017

UNIVERSITEIT VAN PRETORIA
UNIVERSITY OF PRETORIA
YUNIBESITHI YA PRETORIA

Declaration of Authorship

I, Janet van Niekerk, declare that this thesis titled, "Developments in Wishart ensemble and Bayesian application", which I hereby submit for the degree Doctor of Philosopy at the University of Pretoria, is my own work and has not previously been submitted by me for a degree at this or any other tertiary institution.

Signed: __________________________________

Date: __________________________________

UNIVERSITY OF PRETORIA

Abstract

The increased complexity and dimensionality of data necessitates the development of new models that can adequately model the data. Advances in computational approaches have pathed the way for consideration and implementation of more complicated models, previously avoided due to practical difficulties. New models within the Wishart ensemble are developed and some properties are derived. Algorithms for the practical implementation of these matrix variate models are proposed. Simulation studies and real datasets are used to illustrate the use and improved performance of these new models in Bayesian analysis of the multivariate and univariate normal models.

From this speculative research study the following papers emanated:

1. J. Van Niekerk, A. Bekker, M. Arashi, and J.J.J. Roux (2015). "Subjective Bayesian analysis of the elliptical model". In: *Communications in Statistics - Theory and Methods* 44.17, 3738–3753

2. J. Van Niekerk, A. Bekker, M. Arashi, and D.J. De Waal (2016). "Estimation under the matrix variate elliptical model". In: *South African Statistical Journal* 50.1, 149–171

3. J. Van Niekerk, A. Bekker, and M. Arashi (2016). "A gamma-mixture class of distributions with Bayesian application". In: *Communications in Statistics - Simulation and Computation* (Accepted)

4. M. Arashi, A. Bekker, and J. Van Niekerk (2017). "Weighted-type Wishart distributions with application". In: *Revstat* 15(2), 205–222

5. A. Bekker, J. Van Niekerk, and M. Arashi (2017). "Wishart distributions - Advances in Theory with Bayesian application". In: *Journal of Multivariate Analysis* 155, 272–283

UNIVERSITEIT VAN PRETORIA

Opsomming

Die toenemende komplekse en meerdimensionele aard van data noodsaak die ontwikkeling van nuwe modelle om dit beter te modelleer. Vooruitgang in tegnologie baan die weg vir die implementering van meer komplekse modelle, wat voorheen vermy is weens praktiese oorwegings. Nuwe modelle in die Wishartgroep word ontwikkel en sommige eienskappe word afgelei. Algoritmes vir die praktiese implementering van hierdie matriksveranderlike modelle word voorgestel. Simulasie studies en praktiese datastelle word gebruik om die effektiwiteit van hierdie nuwe modelle in Bayes analise van die meer- en eenveranderlike normaalmodelle, te illustreer.

Van hierdie spekulatiewe navorsingstudie is die volgende artikels gepubliseer:

1. J. Van Niekerk, A. Bekker, M. Arashi, and J.J.J. Roux (2015). "Subjective Bayesian analysis of the elliptical model". In: *Communications in Statistics - Theory and Methods* 44.17, 3738–3753
2. J. Van Niekerk, A. Bekker, M. Arashi, and D.J. De Waal (2016). "Estimation under the matrix variate elliptical model". In: *South African Statistical Journal* 50.1, 149–171
3. J. Van Niekerk, A. Bekker, and M. Arashi (2016). "A gamma-mixture class of distributions with Bayesian application". In: *Communications in Statistics - Simulation and Computation* (Accepted)
4. M. Arashi, A. Bekker, and J. Van Niekerk (2017). "Weighted-type Wishart distributions with application". In: *Revstat* 15(2), 205–222
5. A. Bekker, J. Van Niekerk, and M. Arashi (2017). "Wishart distributions - Advances in Theory with Bayesian application". In: *Journal of Multivariate Analysis* 155, 272–283

Acknowledgements

A PhD is a journey where you evolve as a researcher, but also as a person. So many lessons learnt and so many people to thank.

The definition of a supervisor is to steer and guide the student, but you both did so much more than that.
Prof Arashi, thank you for all the encouragement and always being excited about the research. I have learned immensely from your vast knowledge and mathematical abilities. I admire your humility and the way in which you treat people.
Prof Bekker, branches from the same tree, is what we are. I cannot put into words, how much I appreciate everything you have done for me. Thank you for always being available and supportive. I am in awe of your knowledge and patience.
I will forever treasure the memories of this journey...

I would like to thank the University of Pretoria and the National Research Foundation for the financial support. The support from the Department of Statistics and my colleagues throughout this time is highly appreciated.

To my family, especially my parents, thank you so much for always wanting to help, for your genuine interest and continuous support. From now on, I will try to come and visit without my laptop.

Rian, thank you so much for your loving support. It takes a very strong husband to continuously share his wife with a laptop. You have motivated me and inspired me in times when I felt that I will never be able to finish this, but here we are.
Gion (2 years) and Janro (2 months), I am so proud of you two. You are two beautiful strong boys and I thank you for all the times you so patiently shared me, unknowingly. I hope that through this, you will also be inspired to follow your dreams.

Lastly, thank you for carrying me through this Lord. I am so grateful for all the blessings bestowed on me, although I deserve none.

Vir Rian, Gion en Janro…

Contents

List of Symbols

$\prod_{i=1}^{n}$	Product
$\otimes$	Kronecker product
$\sum_{i=1}^{n}$	Sum
$\int$	Indefinite integral
$\int_a^b$	Definite integral over the interval $[a;b]$
$\int_W$	Integral over the space W
$\binom{n}{i}$	n combination i, $\frac{n!}{i!(n-i)!}$
$n!$	n factorial, $n(n-1)(n-2)...(1)$ and $0! = 1$
i	Imaginary unit, $i^2 = -1$
$\|a\|$	Absolute value of scalar a
$etr(\cdot)$	Exponent of trace, $exp(tr(\cdot))$
$diag(a_1, ...a_m)$	Diagonal matrix with $a_1, ..., a_m$ on the diagonal
$\|\boldsymbol{X}\|$	Determinant of matrix $\boldsymbol{X}$
$\|\|\boldsymbol{X}\|\|$	Norm of $\boldsymbol{X}$
$\boldsymbol{X}^{-1}$	Inverse of $\boldsymbol{X}$, $\boldsymbol{X}\boldsymbol{X}^{-1} = \boldsymbol{X}^{-1}\boldsymbol{X} = \boldsymbol{I}$
$\boldsymbol{X}'$	Transpose of $\boldsymbol{X}$
$\boldsymbol{X}^{\frac{1}{2}}$	Non-negative definite square root of $\boldsymbol{X}$, $\boldsymbol{X}^{\frac{1}{2}}\boldsymbol{X}^{\frac{1}{2}} = \boldsymbol{X}$
$\psi_{\boldsymbol{X}}(\boldsymbol{T})$	Characteristic function of $\boldsymbol{X}$ at $\boldsymbol{T}$
$\mathbb{N}$	The set of natural numbers
$\mathbb{R}$	The set of real numbers
$\mathbb{R}^m$	The set of real vectors of size m
$\mathbb{R}^{m\times p}$	The set of real matrices of dimension $m \times p$
$\mathbb{C}$	The set of complex values
Re	The real part of a number
S_m	The group of positive definite matrices of order m $\boldsymbol{X} \in \mathbb{R}^{m\times m} \| \boldsymbol{X} = \boldsymbol{X}', \|\boldsymbol{X}\| > 0$
O_m	The group of orthogonal matrices of order m $\boldsymbol{X} \in \mathbb{R}^{m\times m} \| \boldsymbol{X}'\boldsymbol{X} = \boldsymbol{X}\boldsymbol{X}' = \boldsymbol{I}_m$
$\mathbf{1}_m$	m-dimensional vector of ones
$\boldsymbol{I}_m$	Identity matrix of size m
$N(\mu, \sigma^2)$	Univariate normal distribution
$N_m(\boldsymbol{\mu}, \boldsymbol{\Sigma})$	Multivariate normal distribution
$N_{m,p}(\boldsymbol{\mu}, \boldsymbol{\Sigma} \otimes \boldsymbol{\Omega})$	Matrix variate normal distribution
$T(\mu, \sigma^2, \nu)$	Univariate t distribution
$T_m(\boldsymbol{\mu}, \boldsymbol{\Sigma}, \nu)$	Multivariate t distribution
$T_{m,p}(\boldsymbol{\mu}, \boldsymbol{\Sigma}, \boldsymbol{\Omega}, \nu)$	Matrix variate t distribution
$E(\mu, \sigma^2, g)$	Univariate elliptical distribution
$E_m(\boldsymbol{\mu}, \boldsymbol{\Sigma}, g)$	Multivariate elliptical distribution

$E_{m,p}(\boldsymbol{\mu},\boldsymbol{\Sigma},\boldsymbol{\Omega},g)$	Matrix variate elliptical distribution
$IG(a,b)$	Inverse gamma distribution
$G(a,b)$	Gamma distribution
$HG(\boldsymbol{a},\boldsymbol{b},c,\varphi,n)$	Hypergamma distribution
$KG(a,b,\omega)$	Kummer gamma distribution
$W_m(\boldsymbol{\Sigma},n)$	Wishart distribution
$W_m^{-1}(\boldsymbol{\Sigma},n)$	Inverse Wishart distribution
$HWG_m(\boldsymbol{\Sigma},\boldsymbol{\Omega},\boldsymbol{a},\boldsymbol{b},n,h)$	Hypergeometric Wishart generator distribution
$IHWG_m(\boldsymbol{\Sigma},\boldsymbol{\Omega},\boldsymbol{a},\boldsymbol{b},n,h)$	Inverse hypergeometric Wishart generator distribution
$HW_m(\boldsymbol{\Sigma},\boldsymbol{a}^*,\boldsymbol{b}^*,c,n)$	Hypergeometric Wishart distribution
$NWG_m(\boldsymbol{\Sigma},\boldsymbol{\Omega},b_1,n,h)$	Noncentral Wishart generator distribution
$WG_m(\boldsymbol{\Sigma},n,h)$	Wishart generator distribution
$KW_m(\boldsymbol{\Sigma},\boldsymbol{\Phi},n)$	Kummer Wishart distirbution
$W_m^{I}(\boldsymbol{\Sigma},n)$	Weighted-type I Wishart distribution
$W_m^{II}(\boldsymbol{\Sigma},n)$	Weighted-type II Wishart distribution
$W_m^{III}(\boldsymbol{\Sigma},n)$	Weighted-type III Wishart distribution

$\widehat{\mu}_{B,M}$	Bayes estimator/estimate of μ with the M prior for σ^2
$\widehat{\mu}_{MLE}$	Maximum likelihood estimator/estimate of μ
$\widehat{\boldsymbol{\mu}}_{B,M}$	Bayes estimator/estimate of $\boldsymbol{\mu}$ with the M prior for $\boldsymbol{\Sigma}$
$\widehat{\boldsymbol{\mu}}_{MLE}$	Maximum likelihood estimator/estimate of $\boldsymbol{\mu}$
$\widehat{\sigma^2}_{B,M}$	Bayes estimator/estimate of σ^2 with the M prior for σ^2
$\widehat{\sigma^2}_{MLE}$	Maximum likelihood estimator/estimate of σ^2
$\widehat{\boldsymbol{\Sigma}}_{B,M}$	Bayes estimator/estimate of $\boldsymbol{\Sigma}$ with the M prior for $\boldsymbol{\Sigma}$
$\widehat{\boldsymbol{\Sigma}}_{MLE}$	Maximum likelihood estimator/estimate of $\boldsymbol{\Sigma}$

Chapter 1

Introduction

"He who loves practice without theory is like the sailor who boards ship without a rudder and compass, and never knows where he may be cast..."

- Leonardo da Vinci (1452-1519)

1.1 Rationale of the study

The modeling of real world phenomena is constantly increasing in complexity and standard statistical distributions cannot model these adequately. The question arises whether we can introduce new models to compete with and enhance the standard approaches available in the literature. Various generalizations and extensions have been proposed to augment standard statistical models in order to solve the modeling challenges of real data.

The high velocity of data collection often results in data being stored in vectors or matrices. Matrix variate data is very common in medical applications; Hung and Wang (2013) developed a matrix variate regression model to analyze the risk of a heart attack based on EEG data, while Furlotte and Eskin (2015) used a matrix variate linear mixed model to estimate trait associations based on genetic screening. Image intensities are stored in matrices, thus many image processing algorithms rely on matrix variate data analysis (see Chakrabarty, Biswas, and Bhattacharya (2015) and Murata (2016)).

The Wishart distribution, introduced by Wishart (1928), has been generalized for various application purposes recently. Tourneret, Ferrari, and Letac (2005) applied the noncentral Wishart distribution to image processing while in the context of graphical models, Roverato (2002) defined the hyper-inverse Wishart and Wang and West (2009) extended the inverse Wishart distribution for using hyper-Markov properties (see Dawid and Lauritzen (1993)), while Bryc (2008) proposed the compound Wishart and q-Wishart in graphical models. Adhikari (2010) generalized the Wishart distribution for probabilistic structural dynamics, and Munilla and Cantet (2012) formulated a special structure for the Wishart distribution to apply in modeling the maternal animal. The above generalizations justify the first research question of this study to develop models within the Wishart ensemble.

From a Bayesian viewpoint, alternative subjective priors have been shown to perform better than standard subjective and objective prior models (see Press (1982),Fang and Li (1999)). The consideration of more general subjective priors leads to challenging posterior expressions and computational studies are thus complicated. As technology has developed, it is now possible to apply more complicated prior structures to data to gain efficiency and higher coverage probabilities. Bekker and Roux (1995) considered the Wishart prior as a competitor for the conjugate inverse Wishart prior for the normal model. Although this lead to complicated posterior expressions, Van Niekerk, Bekker, Arashi, and Roux (2015) demonstrated that the Wishart prior is a suitable competitor to the inverse Wishart prior utilising a numerical study. This is a further stimulus for this study, to develop and apply new models within the Wishart ensemble that might perform better than well-known models as priors that are currently available in Bayesian analysis.

Two general models that are proposed in this study within the Wishart ensemble, is the generator-type Wishart and the weighted-type Wishart models. The generator-type Wishart model emanates from replacing the exponential term in the kernel of the Wishart density function with a Borel measurable function and coupling this with a generalized hypergeometric function of matrix argument (See Chapter 3). The second model under investigation has its origin in coupling the kernel of the density function of a Wishart distribution with a Borel measurable function which is termed the weight function (See Chapter 4).

Special members of these newly developed models in the Wishart ensemble are then used in Bayesian analysis. The Bayes estimators and posterior density functions are derived and computed for simulated and real datasets. Practical implementation of these new models depend on the availability of capable software. Some attention is paid to the development of numerical algorithms to ease the application of these new models. Measures of performance for these matrix variate models are proposed and evaluated. This enables the user to compare two or more models for multivariate or matrix variate data.

This thesis addresses the modeling of matrix variate data with new innovative approaches, theoretically and computationally.

1.2 Contributions and research outline

In Chapter 2, the Wishart and inverse Wishart distributions are briefly reported and studied as priors in Bayesian analysis of the elliptical model. Two special cases of the elliptical model are considered, theoretically and computationally. The hypergeometric Wishart generator distribution is introduced in Chapter 3. Some properties are derived and graphically illustrated and a special case is applied in Bayesian analysis of the normal model. Chapter 4 presents the Weighted-type Wishart distributions with their properties and an application in Bayesian analysis of the normal model.

All the numerical Bayesian applications, throughout the thesis, are culminated in Chapter 5 where the developed numerical algorithms are investigated for convergence and robustness to initial values. The proposed priors are evaluated for simulated and real datasets to illustrate the practical usefulness.

Figure 1.1 visually depicts the main outline of the study.

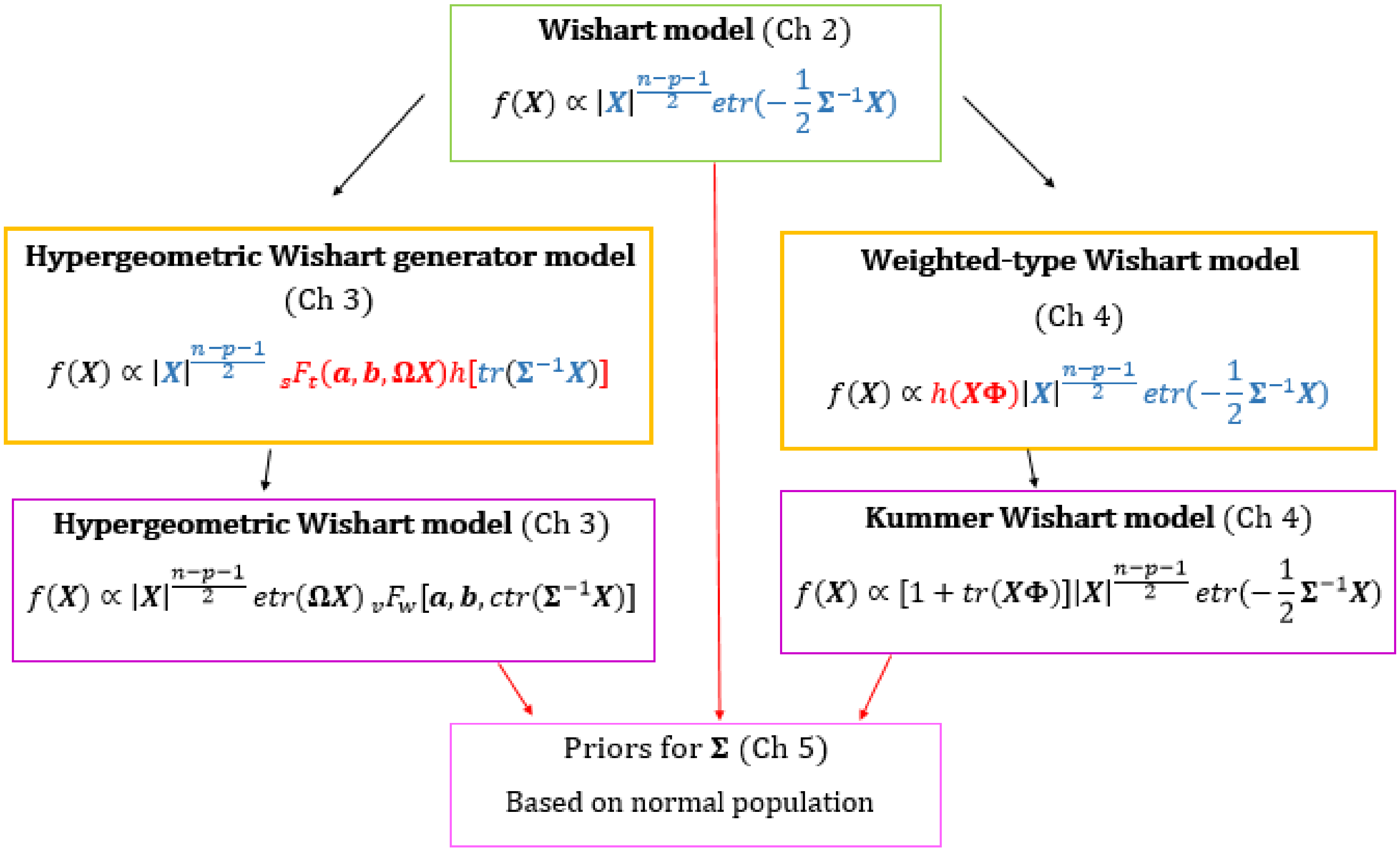


FIGURE 1.1: Main outline of the thesis

Chapter 2

Wishart and inverse Wishart distributions

In this chapter some well-known results of the Wishart and inverse Wishart distributions are reviewed and these distributions are applied as priors for the scale matrix of the matrix variate elliptical model in a Bayesian analysis. Some multivariate and matrix variate measures are defined to study the performance of the different estimators. From this chapter two accredited papers, Van Niekerk, Bekker, Arashi, and Roux (2015) and Van Niekerk, Bekker, Arashi, and De Waal (2016), resulted. The first paper focuses on the vector variate case whereas the second paper focuses on the matrix variate case.

2.1 Introduction

The Wishart distribution for general dimension was first introduced by Wishart (1928) as the generalized product moment distribution of the sum of squares of multivariate normal vectors using a geometrical argument. Previously, Fisher (1915) derived the Wishart distribution with dimension 2. The Wishart distribution can also be viewed as the multivariate extension of the gamma distribution, and for integer degrees of freedom, the chi-square distribution. Since 1928 ample literature have appeared on the Wishart distribution and associated properties, only some are stated in Section 2.3. The Wishart distribution is defined as follows,

Definition 2.1.1. *(Gupta and Nagar 2000, Definition 3.2.1, p.87). A random positive definite matrix $\boldsymbol{X}$ is said to have a Wishart distribution with parameters m,n and $\boldsymbol{\Sigma}$ if its density function is given by*

$$f(\boldsymbol{X}) = \left[2^{\frac{mn}{2}}\Gamma_m\left(\frac{n}{2}\right)\right]^{-1} |\boldsymbol{\Sigma}|^{-\frac{n}{2}} |\boldsymbol{X}|^{\frac{n}{2}-\frac{m+1}{2}} etr\left[-\frac{1}{2}\boldsymbol{\Sigma}^{-1}\boldsymbol{X}\right], \quad \boldsymbol{X},\boldsymbol{\Sigma} \in S_m,$$

where $\Gamma(\cdot)$ is defined in R.8, $\Gamma_m(\cdot)$ is defined in R.9 and $n \geq m$. It is denoted as $\boldsymbol{X} \sim W_m(\boldsymbol{\Sigma}, n)$.

For any inference of the population covariance matrix of a multivariate normal distribution, the Wishart distribution is essential. The inverse Wishart distribution is the matrix variate generalisation of the inverse gamma distribution (see R.40), and defined as follows,

Definition 2.1.2. *(Gupta and Nagar 2000, Definition 3.4.1, p.111). A random positive definite matrix $\boldsymbol{X}$ follows the inverse Wishart distribution with parameter matrix $\boldsymbol{\Sigma}$ and n degrees of freedom, $n > 2m$, if its density function is given by*

$$f(\boldsymbol{X}) = \left[2^{\frac{m(n-m-1)}{2}}\Gamma_m\left(\frac{n-m-1}{2}\right)\right]^{-1} |\boldsymbol{X}|^{-\frac{n}{2}} |\boldsymbol{\Sigma}|^{\frac{n}{2}-\frac{m+1}{2}} etr\left[-\frac{1}{2}\boldsymbol{X}^{-1}\boldsymbol{\Sigma}\right], \quad \boldsymbol{X},\boldsymbol{\Sigma} \in S_m,$$

where $\Gamma(.)$ is defined in R.8, $\Gamma_m(.)$ is defined in R.9 and $n > 2m$. It is denoted as $\boldsymbol{X} \sim W_m^{-1}(\boldsymbol{\Sigma}, n)$.

The inverse Wishart distribution has been shown to be a conjugate prior (R.57) for the Bayesian analysis of the covariance matrix of the multivariate normal model by Press (1982), amongst others. The Wishart distribution has also been considered as a prior for the covariance matrix of the matrix variate normal model by Bekker and Roux (1995). The question arises whether the normal-Wishart prior performs as well as or better than the normal-inverse Wishart prior. To this end, the Bayesian analysis of the normal model under the latter prior should be studied. However, in this chapter a more general framework is considered where the underlying model is matrix variate elliptical rather than matrix variate normal, which includes the latter as a special case (see Sections 2.4 and 2.5).

2.2 Construction methodology

In this section the intuitive understanding of the Wishart distribution is given and the usefulness is illustrated. Assume that a random vector, $\boldsymbol{X} \in \mathbb{R}^m$, from a multivariate normal distribution with mean vector $\boldsymbol{\mu} \in \mathbb{R}^m$ and covariance matrix $\boldsymbol{\Sigma} \in S_m$ is available such that

$$\boldsymbol{X} \sim N_m(\boldsymbol{\mu}, \boldsymbol{\Sigma}), \tag{2.1}$$

then the density function of $\boldsymbol{X}$ is given by R.45,

$$f(\boldsymbol{X}) = (2\pi)^{-\frac{m}{2}} |\boldsymbol{\Sigma}|^{-\frac{1}{2}} etr\left(-\frac{1}{2}\boldsymbol{\Sigma}^{-1}(\boldsymbol{X}-\boldsymbol{\mu})(\boldsymbol{X}-\boldsymbol{\mu})'\right).$$

For further properties of the multivariate normal distribution, the reader is referred to Press (1982), Muirhead (1982) and Anderson (2003). Now if a sample of size p is available from the distribution in (2.1), define the sample matrix

$$\boldsymbol{Z} = \begin{bmatrix} \boldsymbol{X}_1 & \boldsymbol{X}_2 & \ldots & \boldsymbol{X}_{p-1} & \boldsymbol{X}_p \end{bmatrix},$$

then $\boldsymbol{Z} \sim N_{m,p}(\boldsymbol{\mu}\boldsymbol{1}_{1\times p}, \boldsymbol{\Sigma} \otimes \boldsymbol{I}_p)$ and if $\boldsymbol{\mu} = \boldsymbol{0}$, then $\boldsymbol{Z}\boldsymbol{Z}' \sim W_m(\boldsymbol{\Sigma}, p)$ (Muirhead 1982, Definition 3.1.3, p.82) with density function given in Definition 2.1.1. The sample covariance matrix, $\boldsymbol{S}$, is given by

$$\begin{aligned} \boldsymbol{S} &= \sum_{i=1}^{p} (\boldsymbol{X}_i - \overline{\boldsymbol{X}})(\boldsymbol{X}_i - \overline{\boldsymbol{X}})' \\ &= (\boldsymbol{Z} - \overline{\boldsymbol{X}}\boldsymbol{1}_{1\times p})(\boldsymbol{Z} - \overline{\boldsymbol{X}}\boldsymbol{1}_{1\times p})'. \end{aligned}$$

It is well-known that $\boldsymbol{S} \sim W_m(\boldsymbol{\Sigma}, p-1)$ (Muirhead 1982, Corollary 3.2.2, p.86). It is evident that the Wishart distribution is an essential distribution for inference regarding the multivariate normal distribution. In the next section some properties of the Wishart and inverse Wishart distributions, necessary for this study, are briefly reviewed.

2.3 Properties

The following section contains some well-known relevant properties of the Wishart and inverse Wishart distributions which can be found in Muirhead (1982).

Theorem 2.3.1. *(Muirhead 1982, p.87) If $X \sim W_1(\sigma^2, n)$ then $\frac{X}{\sigma^2} \sim \chi^2(n)$.*

Theorem 2.3.2. *(Muirhead 1982, Theorem 3.2.3, p.87) If $\boldsymbol{X} \sim W_m(\boldsymbol{\Sigma}, n)$ then the characteristic function of $\boldsymbol{X}$ (the joint characteristic function of the $\frac{m(m+1)}{2}$ variables $a_{ij}, 1 \le i \le j \le m$) is*

$$\psi_{\boldsymbol{X}}(\boldsymbol{T}) = |\boldsymbol{I}_m - 2i\boldsymbol{T}\boldsymbol{\Sigma}|^{-\frac{n}{2}}.$$

Theorem 2.3.3. *(Muirhead 1982, p.90) The first moment of $\boldsymbol{X} \sim W_m(\boldsymbol{\Sigma}, n)$ is $E(\boldsymbol{X}) = n\boldsymbol{\Sigma}$.*

Theorem 2.3.4. *(Muirhead 1982, Eq.15, p.101) If $\boldsymbol{X} \sim W_m(\boldsymbol{\Sigma}, n)$ then the r^{th} moment of $|\boldsymbol{X}|$ is*

$$E(|\boldsymbol{X}|^r) = \frac{2^{mr}\Gamma_m\left(\frac{n}{2}+r\right)}{\Gamma_m\left(\frac{n}{2}\right)}|\boldsymbol{\Sigma}|^r.$$

Theorem 2.3.5. *(Muirhead 1982, Theorem 3.2.18, p.106) If* $\boldsymbol{X} \sim W_m(\boldsymbol{\Sigma}, n)$ *with* $n \geq m$*, The joint density function of the eigenvalues* $\boldsymbol{\Lambda} = diag(\lambda_1, ..., \lambda_m)$*,* $\lambda_1 \geq ... \geq \lambda_m > 0$ *of* $\boldsymbol{X}$ *is given by*

$$g'(\lambda_1, ..., \lambda_m) = \frac{\pi^{\frac{m^2}{2}} 2^{-\frac{mn}{2}} |\boldsymbol{\Sigma}|^{-\frac{n}{2}}}{\Gamma_m(\frac{m}{2})\Gamma_m(\frac{n}{2})} \prod_{i=1}^{m} \lambda_i^{\frac{n}{2}-\frac{m+1}{2}} \prod_{i<j}^{m} (\lambda_i - \lambda_j) \int_{O_m} etr(-\frac{1}{2}\boldsymbol{\Sigma}^{-1}\boldsymbol{H}\boldsymbol{\Lambda}\boldsymbol{H}')d\boldsymbol{H}.$$

Remark 2.3.1. *If* $\boldsymbol{X} \sim W_m(c_1\boldsymbol{I}_m, n)$ *with* $n \geq m$*, The joint density function of the eigenvalues* $\boldsymbol{\Lambda} = diag(\lambda_1, ..., \lambda_m)$*,* $\lambda_1 \geq ... \geq \lambda_m > 0$ *of* $\boldsymbol{X}$ *is given by*

$$g'(\lambda_1, ..., \lambda_m) = \frac{\pi^{\frac{m^2}{2}}}{(2c_1)^{\frac{nm}{2}}\Gamma_m(\frac{m}{2})\Gamma_m(\frac{n}{2})} \prod_{i=1}^{m} \lambda_i^{\frac{n}{2}-\frac{m+1}{2}} \prod_{i<j}^{m} (\lambda_i - \lambda_j)\, etr\left(-\frac{1}{2c_1}\boldsymbol{\Lambda}\right).$$

Earlier it was mentioned that Bekker and Roux (1995) considered the non-conjugate normal-Wishart prior in the Bayesian analysis of the matrix variate normal model. Now, this prior will be considered for the matrix variate elliptical model and the benefit of the additional complexity is investigated in the forthcoming sections.

2.4 Bayesian application

The Wishart and inverse Wishart distributions will now be applied as subjective priors (see R.56) for the scale matrix of the matrix variate elliptical model. See Figure 2.1 for the outlay of this section.

In this section, the matrix variate elliptical model where $\boldsymbol{X} \sim E_{m,p}(\boldsymbol{\mu}, \boldsymbol{\Sigma}, \boldsymbol{\Omega}, g'')$, (see R.51), is considered as the underlying model, instead of the matrix variate normal model as to derive more general results which can easily be used when a special case of the elliptical model is considered. Two special cases, the matrix variate normal and matrix variate t distributions, will be investigated theoretically and practically as illustrated in Figure 2.1. The aim is to estimate the location and scale matrices, $\boldsymbol{\mu}$ and $\boldsymbol{\Sigma}$, respectively where $\boldsymbol{\Omega}$ is assumed to be a known hyperparameter. The considered prior distributions will thus be for $\boldsymbol{\mu}$ and $\boldsymbol{\Sigma}$.

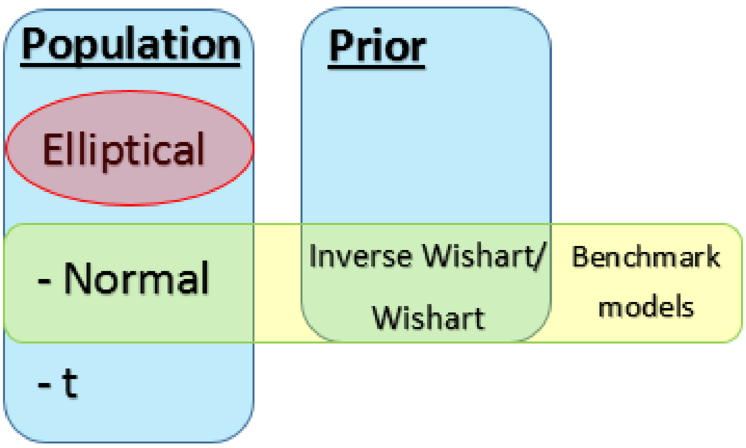


FIGURE 2.1: Outlay of Section 2.4

Although objective Bayesian analysis (see R.55) for the matrix variate elliptical model was considered by Fang and Li (1999), very few results and estimators for the matrix variate elliptical model exist from a Bayesian viewpoint. Prior information will be reflected by using the normal-inverse Wishart and the normal-Wishart prior distributions respectively (see Van Niekerk, Bekker, Arashi, and Roux (2015) for the multivariate elliptical model). In this chapter, the squared error loss function (SEL, see R.60) as well as the loss function defined by Das and Dey (2010) (see R.61) will be used for the Bayesian inference.

2.4.1 Normal-inverse Wishart prior

Let z be a known positive scalar and $\mathbf{\Psi} = z^{-1}\mathbf{\Sigma}$ follow an inverse Wishart distribution with scale matrix $\mathbf{\Phi}$ and n degrees of freedom. For any $z > 0$, a generated variate of $\mathbf{\Psi}$ will produce a generated variate of $\mathbf{\Sigma}$ since $\mathbf{\Sigma} = z\mathbf{\Psi}$. Now assume a normal-inverse Wishart prior for $(\boldsymbol{\mu},\mathbf{\Psi})$ with prior distributions for $\boldsymbol{\mu}$ and $\mathbf{\Psi}$ respectively, $\boldsymbol{\mu}|\mathbf{\Psi} \sim N_{m,p}(\boldsymbol{\theta}, \frac{1}{n_0}\mathbf{\Psi}\otimes\mathbf{\Omega})$ and $\mathbf{\Psi} \sim W_m^{-1}(\mathbf{\Phi}, n)$. Note that n_0 is an *enriching parameter*. From R.49 and Definition 2.1.2 the prior density functions are

$$\pi(\boldsymbol{\mu}|\mathbf{\Psi}) = (2\pi)^{-\frac{mp}{2}}|\frac{1}{n_0}\mathbf{\Psi}|^{-\frac{p}{2}}|\mathbf{\Omega}|^{-\frac{m}{2}}etr\left[-\frac{n_0}{2}\mathbf{\Psi}^{-1}(\boldsymbol{\mu}-\boldsymbol{\theta})\mathbf{\Omega}^{-1}(\boldsymbol{\mu}-\boldsymbol{\theta})'\right], \boldsymbol{\mu} \in \mathbb{R}^{m\times p}, \tag{2.2}$$

and

$$\pi(\mathbf{\Psi}) = \left[\Gamma_m\left(\frac{n-m-1}{2}\right)\right]^{-1}|\frac{1}{2}\mathbf{\Phi}|^{\frac{n-m-1}{2}}|\mathbf{\Psi}|^{-\frac{n}{2}}etr\left[-\frac{1}{2}\mathbf{\Psi}^{-1}\mathbf{\Phi}\right], \mathbf{\Psi},\mathbf{\Phi} \in S_m \text{ and } n > 2m,$$

with the joint prior density function

$$\pi(\boldsymbol{\mu},\mathbf{\Psi}) \propto |\mathbf{\Psi}|^{-\frac{n+p}{2}}etr\left[-\frac{1}{2}\mathbf{\Psi}^{-1}\left(n_0(\boldsymbol{\mu}-\boldsymbol{\theta})\mathbf{\Omega}^{-1}(\boldsymbol{\mu}-\boldsymbol{\theta})'+\mathbf{\Phi}\right)\right], \tag{2.3}$$

and known hyperparameters $\boldsymbol{\theta}, \mathbf{\Omega}$, n_0 and n. Now

$$\begin{aligned} & \pi(\boldsymbol{\mu},\mathbf{\Sigma}|z) \\ \propto\ & \pi(\boldsymbol{\mu},\mathbf{\Psi})|J(\mathbf{\Psi}\to\mathbf{\Sigma})| \\ \propto\ & z^{-\frac{m(m+1-n-p)}{2}}|\mathbf{\Sigma}|^{-\frac{n+p}{2}} \\ \times\ & etr\left[-\frac{z}{2}\mathbf{\Sigma}^{-1}\left(n_0(\boldsymbol{\mu}-\boldsymbol{\theta})\mathbf{\Omega}^{-1}(\boldsymbol{\mu}-\boldsymbol{\theta})'+\mathbf{\Phi}\right)\right], \end{aligned} \tag{2.4}$$

since the Jacobian is $J(\mathbf{\Psi}\to\mathbf{\Sigma}) = z^{\frac{-m(m+1)}{2}}$ (see R.2). Following Arashi, Iranmanesh, and Salarzadeh Jenatabadi (2013) the conjugate prior for $(\boldsymbol{\mu},\mathbf{\Sigma})$ of the matrix variate elliptical model can be obtained as

$$\pi(\boldsymbol{\mu},\mathbf{\Sigma}) \propto \int_0^\infty \pi(\boldsymbol{\mu},\mathbf{\Sigma}|z)w(z)dz. \tag{2.5}$$

This representation of a prior distribution coincides with the representation in R.51. It should be noted that z is not a random variable.

Suppose that a random sample of size n_1, $\boldsymbol{X} = \{\boldsymbol{X}_1, ..., \boldsymbol{X}_{n_1}\}$, is available, then the likelihood function (see R.54) is obtained from R.51 as follows

$$\begin{aligned}
& L(\boldsymbol{\mu},\boldsymbol{\Sigma}|\boldsymbol{X}) \\
= & \prod_{i=1}^{n_1}\int_0^\infty w(z) f_{N_{\boldsymbol{\mu},z^{-1}\boldsymbol{\Sigma}\otimes\boldsymbol{\Omega}}}(\boldsymbol{X}_i|\boldsymbol{\mu},\boldsymbol{\Sigma},\boldsymbol{\Omega})dz \\
\propto & \int_0^\infty w(z)|z^{-1}\boldsymbol{\Sigma}|^{-\frac{n_1 p}{2}}|\boldsymbol{\Omega}|^{-\frac{n_1 m}{2}} etr\left[-\frac{1}{2}\sum_{i=1}^{n_1}(\boldsymbol{X}_i-\boldsymbol{\mu})' z\boldsymbol{\Sigma}^{-1}(\boldsymbol{X}_i-\boldsymbol{\mu})\boldsymbol{\Omega}^{-1}\right] dz \\
= & \int_0^\infty w(z) z^{\frac{n_1 mp}{2}}|\boldsymbol{\Sigma}|^{-\frac{n_1 p}{2}}|\boldsymbol{\Omega}|^{-\frac{n_1 m}{2}} \\
\times & \; etr\left[-\frac{z}{2}\boldsymbol{\Sigma}^{-1}\left[\boldsymbol{V}+n(\overline{\boldsymbol{X}}-\boldsymbol{\mu})\boldsymbol{\Omega}^{-1}(\overline{\boldsymbol{X}}-\boldsymbol{\mu})'\right]\right] dz, \qquad (2.6)
\end{aligned}$$

where

$$\boldsymbol{V} = \sum_{i=1}^{n_1}(\boldsymbol{X}_i-\overline{\boldsymbol{X}})\boldsymbol{\Omega}^{-1}(\boldsymbol{X}_i-\overline{\boldsymbol{X}})'. \qquad (2.7)$$

From (2.4) and (2.6) the joint posterior density function (see R.58) is

$$\begin{aligned}
q(\boldsymbol{\mu},\boldsymbol{\Sigma}|\boldsymbol{X}) \propto & \; \pi(\boldsymbol{\mu},\boldsymbol{\Sigma})L(\boldsymbol{\mu},\boldsymbol{\Sigma}|\boldsymbol{X}) \\
= & \int_0^\infty z^{-\frac{m(m+1-n-p-n_1 p)}{2}} w(z)|\boldsymbol{\Sigma}|^{-\frac{n+n_1 p+p}{2}} \\
& \times etr\left[-\frac{z}{2}\boldsymbol{\Sigma}^{-1}\left(n_0(\boldsymbol{\mu}-\boldsymbol{\theta})\boldsymbol{\Omega}^{-1}(\boldsymbol{\mu}-\boldsymbol{\theta})'+\boldsymbol{\Phi}\right)\right] \\
& \times etr\left[-\frac{z}{2}\boldsymbol{\Sigma}^{-1}\left[\boldsymbol{V}+n_1(\overline{\boldsymbol{X}}-\boldsymbol{\mu})\boldsymbol{\Omega}^{-1}(\overline{\boldsymbol{X}}-\boldsymbol{\mu})'\right]\right] dz. \qquad (2.8)
\end{aligned}$$

Theorem 2.4.1. *The marginal posterior distribution of $\boldsymbol{\mu}$ for the normal model with prior (2.4) is a matrix variate t distribution with parameters $\boldsymbol{b}$,$\boldsymbol{D}$ and $\frac{1}{n_1+n_0}\boldsymbol{\Omega}$ and degrees of freedom $n+n_1 p-2m$ with density function*

$$\begin{aligned}
q(\boldsymbol{\mu}|\boldsymbol{X}) = & \frac{\Gamma_m\left(\frac{n+n_1 p+p-m-1}{2}\right)}{\pi^{\frac{mp}{2}}\Gamma_m\left(\frac{n+n_1 p-m-1}{2}\right)}\left|\frac{1}{n_1+n_0}\boldsymbol{\Omega}\right|^{-\frac{m}{2}}|\boldsymbol{D}|^{-\frac{p}{2}} \\
& \times\left|\boldsymbol{I}_m+\boldsymbol{D}^{-1}(\boldsymbol{\mu}-\boldsymbol{b})\left(\frac{1}{n_1+n_0}\boldsymbol{\Omega}\right)^{-1}(\boldsymbol{\mu}-\boldsymbol{b})'\right|^{-\frac{n+n_1 p+p-m-1}{2}} \qquad (2.9)
\end{aligned}$$

with $\boldsymbol{D} = \frac{n_1 n_0}{n_1+n_0}(\overline{\boldsymbol{X}}-\boldsymbol{\theta})\boldsymbol{\Omega}^{-1}(\overline{\boldsymbol{X}}-\boldsymbol{\theta})'+\boldsymbol{\Phi}+\boldsymbol{V}$ and $\boldsymbol{b} = \frac{1}{n_1+n_0}\left(n_1\overline{\boldsymbol{X}}-n_0\boldsymbol{\theta}\right)$.

Proof. From (2.8) follows that

$$\begin{aligned} q(\boldsymbol{\mu}|\boldsymbol{X}) &\propto \int_0^\infty z^{-\frac{m(m+1-n-p-n_1p)}{2}} w(z) \int_{S_m} |\boldsymbol{\Sigma}|^{-\frac{n+n_1p+p}{2}} \\ &\quad \times etr\left[-\frac{z}{2}\boldsymbol{\Sigma}^{-1}\left(n_0(\boldsymbol{\mu}-\boldsymbol{\theta})\boldsymbol{\Omega}^{-1}(\boldsymbol{\mu}-\boldsymbol{\theta})'+\boldsymbol{\Phi}\right)\right] \\ &\quad \times etr\left[-\frac{z}{2}\boldsymbol{\Sigma}^{-1}\left[\boldsymbol{V}+n_1(\overline{\boldsymbol{X}}-\boldsymbol{\mu})\boldsymbol{\Omega}^{-1}(\overline{\boldsymbol{X}}-\boldsymbol{\mu})'\right]\right] d\boldsymbol{\Sigma} dz. \end{aligned} \tag{2.10}$$

Let

$$\boldsymbol{A} = \boldsymbol{V} + n_1(\overline{\boldsymbol{X}}-\boldsymbol{\mu})\boldsymbol{\Omega}^{-1}(\overline{\boldsymbol{X}}-\boldsymbol{\mu})', \tag{2.11}$$

with $\boldsymbol{V}$ as defined in (2.7) and

$$\boldsymbol{B} = n_0(\boldsymbol{\mu}-\boldsymbol{\theta})\boldsymbol{\Omega}^{-1}(\boldsymbol{\mu}-\boldsymbol{\theta})' + \boldsymbol{\Phi}, \tag{2.12}$$

then from (2.10) follows that

$$q(\boldsymbol{\mu}|\boldsymbol{X}) \propto \int_0^\infty z^{-\frac{m(m+1-n-p-n_1p)}{2}} w(z) \int_{S_m} |\boldsymbol{\Sigma}|^{-\frac{n+n_1p+p}{2}} etr\left[-\frac{z}{2}\boldsymbol{\Sigma}^{-1}(\boldsymbol{B}+\boldsymbol{A})\right] d\boldsymbol{\Sigma} dz.$$

Note that one can identify, from the above, the functional form of the density function of an inverse Wishart distribution with parameters $z(\boldsymbol{A}+\boldsymbol{B})$ and $n+n_1p+p-m-1$, $n+n_1p+p-m-1 > 2m$ (see Definition 2.1.2). The marginal posterior density function of $\boldsymbol{\mu}$ is then given by

$$\begin{aligned} q(\boldsymbol{\mu}|\boldsymbol{X}) &\propto \int_0^\infty z^{-\frac{m(m+1-n-p-n_1p)}{2}} z^{-\frac{m(n+n_1p+p-m-1)}{2}} w(z) |\boldsymbol{B}+\boldsymbol{A}|^{-\frac{n+n_1p+p-m-1}{2}} dz \\ &= |\boldsymbol{B}+\boldsymbol{A}|^{-\frac{n+n_1p+p-m-1}{2}}, \end{aligned} \tag{2.13}$$

since $\int_0^\infty w(z)dz = 1$. Note that

$$\begin{aligned} & n_1(\overline{\boldsymbol{X}}-\boldsymbol{\mu})\boldsymbol{\Omega}^{-1}(\overline{\boldsymbol{X}}-\boldsymbol{\mu})' + n_0(\boldsymbol{\mu}-\boldsymbol{\theta})\boldsymbol{\Omega}^{-1}(\boldsymbol{\mu}-\boldsymbol{\theta})' \\ = \; & n_1\overline{\boldsymbol{X}}\boldsymbol{\Omega}^{-1}\overline{\boldsymbol{X}}' - n_1\overline{\boldsymbol{X}}\boldsymbol{\Omega}^{-1}\boldsymbol{\mu}' - n_1\boldsymbol{\mu}\boldsymbol{\Omega}^{-1}\overline{\boldsymbol{X}}' + n_1\boldsymbol{\mu}\boldsymbol{\Omega}^{-1}\boldsymbol{\mu}' \\ & + n_0\boldsymbol{\mu}\boldsymbol{\Omega}^{-1}\boldsymbol{\mu}' - n_0\boldsymbol{\mu}\boldsymbol{\Omega}^{-1}\boldsymbol{\theta}' - n_0\boldsymbol{\theta}\boldsymbol{\Omega}^{-1}\boldsymbol{\mu}' + n_0\boldsymbol{\theta}\boldsymbol{\Omega}^{-1}\boldsymbol{\theta}' \\ = \; & (n_1+n_0)\boldsymbol{\mu}\boldsymbol{\Omega}^{-1}\boldsymbol{\mu}' - (n_1\overline{\boldsymbol{X}}+n_0\boldsymbol{\theta})\boldsymbol{\Omega}^{-1}\boldsymbol{\mu}' - \boldsymbol{\mu}\boldsymbol{\Omega}^{-1}(n_1\overline{\boldsymbol{X}}'+n_0\boldsymbol{\theta}') \\ & + n_1\overline{\boldsymbol{X}}\boldsymbol{\Omega}^{-1}\overline{\boldsymbol{X}}' + n_0\boldsymbol{\theta}\boldsymbol{\Omega}^{-1}\boldsymbol{\theta}' \\ = \; & (n_1+n_0)\left[\boldsymbol{\mu}\boldsymbol{\Omega}^{-1}\boldsymbol{\mu}' - \frac{1}{n_1+n_0}(n_1\overline{\boldsymbol{X}}+n_0\boldsymbol{\theta})\boldsymbol{\Omega}^{-1}\boldsymbol{\mu}' - \frac{1}{n_1+n_0}\boldsymbol{\mu}\boldsymbol{\Omega}^{-1}(n_1\overline{\boldsymbol{X}}'+n_0\boldsymbol{\theta}')\right] \\ & + n_1\overline{\boldsymbol{X}}\boldsymbol{\Omega}^{-1}\overline{\boldsymbol{X}}' + n_0\boldsymbol{\theta}\boldsymbol{\Omega}^{-1}\boldsymbol{\theta}' \\ = \; & (n_1+n_0)\left[\left(\boldsymbol{\mu} - \frac{1}{n_1+n_0}(n_1\overline{\boldsymbol{X}}+n_0\boldsymbol{\theta})\right)\boldsymbol{\Omega}^{-1}\left(\boldsymbol{\mu} - \frac{1}{n_1+n_0}(n_1\overline{\boldsymbol{X}}+n_0\boldsymbol{\theta})\right)'\right] \\ & - \frac{1}{n_1+n_0}(n_1\overline{\boldsymbol{X}}+n_0\boldsymbol{\theta})\boldsymbol{\Omega}^{-1}\frac{1}{n_1+n_0}(n_1\overline{\boldsymbol{X}}+n_0\boldsymbol{\theta})' + n_1\overline{\boldsymbol{X}}\boldsymbol{\Omega}^{-1}\overline{\boldsymbol{X}}' + n_0\boldsymbol{\theta}\boldsymbol{\Omega}^{-1}\boldsymbol{\theta}' \end{aligned}$$

$$
\begin{aligned}
&= (n_1+n_0)\left[(\boldsymbol{\mu}-\boldsymbol{b})\boldsymbol{\Omega}^{-1}(\boldsymbol{\mu}-\boldsymbol{b})'\right]-\frac{1}{n_1+n_0}\left[n_1^2\overline{\boldsymbol{X}}\boldsymbol{\Omega}^{-1}\overline{\boldsymbol{X}}'+n_0^2\boldsymbol{\theta}\boldsymbol{\Omega}^{-1}\boldsymbol{\theta}'\right]\\
&\quad +n_1\overline{\boldsymbol{X}}\boldsymbol{\Omega}^{-1}\overline{\boldsymbol{X}}'+n_0\boldsymbol{\theta}\boldsymbol{\Omega}^{-1}\boldsymbol{\theta}'-\frac{1}{n_1+n_0}\left[n_1n_0\overline{\boldsymbol{X}}\boldsymbol{\Omega}^{-1}\boldsymbol{\theta}'+n_1n_0\boldsymbol{\theta}\boldsymbol{\Omega}^{-1}\overline{\boldsymbol{X}}'\right]\\
&= (n_1+n_0)\left[(\boldsymbol{\mu}-\boldsymbol{b})\boldsymbol{\Omega}^{-1}(\boldsymbol{\mu}-\boldsymbol{b})'\right]-\frac{1}{n_1+n_0}\left[n_1n_0\overline{\boldsymbol{X}}\boldsymbol{\Omega}^{-1}\boldsymbol{\theta}'+n_1n_0\boldsymbol{\theta}\boldsymbol{\Omega}^{-1}\overline{\boldsymbol{X}}'\right]\\
&\quad -\frac{1}{n_1+n_0}\left[n_1^2\overline{\boldsymbol{X}}\boldsymbol{\Omega}^{-1}\overline{\boldsymbol{X}}'-n_1(n_1+n_0)\overline{\boldsymbol{X}}\boldsymbol{\Omega}^{-1}\overline{\boldsymbol{X}}'-n_0(n_1+n_0)\boldsymbol{\theta}\boldsymbol{\Omega}^{-1}\boldsymbol{\theta}'+n_0^2\boldsymbol{\theta}\boldsymbol{\Omega}^{-1}\boldsymbol{\theta}'\right]\\
&= (n_1+n_0)\left[(\boldsymbol{\mu}-\boldsymbol{b})\boldsymbol{\Omega}^{-1}(\boldsymbol{\mu}-\boldsymbol{b})'\right]-\frac{n_1n_0}{n_1+n_0}\left[\overline{\boldsymbol{X}}\boldsymbol{\Omega}^{-1}\boldsymbol{\theta}'+\boldsymbol{\theta}\boldsymbol{\Omega}^{-1}\overline{\boldsymbol{X}}'\right]\\
&\quad +\frac{1}{n_1+n_0}\left[n_1n_0\overline{\boldsymbol{X}}\boldsymbol{\Omega}^{-1}\overline{\boldsymbol{X}}'+n_1n_0\boldsymbol{\theta}\boldsymbol{\Omega}^{-1}\boldsymbol{\theta}'\right]\\
&= (n_1+n_0)\left[(\boldsymbol{\mu}-\boldsymbol{b})\boldsymbol{\Omega}^{-1}(\boldsymbol{\mu}-\boldsymbol{b})'\right]+\frac{n_1n_0}{n_1+n_0}\left[\overline{\boldsymbol{X}}\boldsymbol{\Omega}^{-1}\overline{\boldsymbol{X}}'-\boldsymbol{\theta}\boldsymbol{\Omega}^{-1}\overline{\boldsymbol{X}}'-\overline{\boldsymbol{X}}\boldsymbol{\Omega}^{-1}\boldsymbol{\theta}'+\boldsymbol{\theta}\boldsymbol{\Omega}^{-1}\boldsymbol{\theta}'\right]\\
&= (n_1+n_0)(\boldsymbol{\mu}-\boldsymbol{b})\boldsymbol{\Omega}^{-1}(\boldsymbol{\mu}-\boldsymbol{b})'+\frac{n_1n_0}{n_1+n_0}(\overline{\boldsymbol{X}}-\boldsymbol{\theta})\boldsymbol{\Omega}^{-1}(\overline{\boldsymbol{X}}-\boldsymbol{\theta})' \qquad (2.14)
\end{aligned}
$$

with

$$
\boldsymbol{b}=\frac{1}{n_1+n_0}\left(n_1\overline{\boldsymbol{X}}-n_0\boldsymbol{\theta}\right). \qquad (2.15)
$$

Hence from (2.11), (2.12) and (2.14) it follows that

$$
\begin{aligned}
\boldsymbol{B}+\boldsymbol{A} &= \frac{n_1n_0}{n_1+n_0}(\overline{\boldsymbol{X}}-\boldsymbol{\theta})\boldsymbol{\Omega}^{-1}(\overline{\boldsymbol{X}}-\boldsymbol{\theta})'+\boldsymbol{\Phi}+\boldsymbol{V}\\
&\quad +(n_1+n_0)(\boldsymbol{\mu}-\boldsymbol{b})\boldsymbol{\Omega}^{-1}(\boldsymbol{\mu}-\boldsymbol{b})'\\
&= \boldsymbol{D}+(n_1+n_0)(\boldsymbol{\mu}-\boldsymbol{b})\boldsymbol{\Omega}^{-1}(\boldsymbol{\mu}-\boldsymbol{b})', \qquad (2.16)
\end{aligned}
$$

with

$$
\boldsymbol{D}=\frac{n_1n_0}{n_1+n_0}(\overline{\boldsymbol{X}}-\boldsymbol{\theta})\boldsymbol{\Omega}^{-1}(\overline{\boldsymbol{X}}-\boldsymbol{\theta})'+\boldsymbol{\Phi}+\boldsymbol{V}. \qquad (2.17)
$$

Therefore from (2.13) and (2.16),

$$
q(\boldsymbol{\mu}|\boldsymbol{X})\propto|\boldsymbol{D}+(n_1+n_0)(\boldsymbol{\mu}-\boldsymbol{b})\boldsymbol{\Omega}^{-1}(\boldsymbol{\mu}-\boldsymbol{b})'|^{-\frac{n+n_1p+p-m-1}{2}}.
$$

From R.50, this can be identified as the functional form of the density function of a matrix variate t distribution with parameters $\boldsymbol{b},\boldsymbol{D}$ and $\frac{1}{n_1+n_0}\boldsymbol{\Omega}$ and degrees of freedom $n+n_1p-2m$. □

Theorem 2.4.2. *The marginal posterior density function of* $\mathbf{\Sigma}$ *for the normal model with prior (2.4) is*

$$\begin{aligned} q(\mathbf{\Sigma}|\mathbf{X}) &= \frac{2^{-\frac{m(n+n_1p-m-1)}{2}}}{\Gamma_m\left(\frac{n+n_1p-m-1}{2}\right)}|\mathbf{D}|^{\frac{n+n_1p-m-1}{2}}|\mathbf{\Sigma}|^{-\frac{n+n_1p}{2}} \\ &\quad \times \sum_{k=0}^{\infty}\frac{\left(-\frac{1}{2}tr\left[\mathbf{\Sigma}^{-1}\mathbf{D}\right]\right)^k}{k!}\varsigma\Big(-\frac{m(m+1-n-n_1p)-2k}{2}\Big), \end{aligned} \tag{2.18}$$

with $\mathbf{D}$ *as defined in* (2.17)*, provided* $\varsigma(-\frac{m(m+1-n-n_1p)-2k}{2})$ *exists and* $n+n_1p-m-1>2m$*, where*

$$\varsigma(\alpha) = \int_0^{\infty} z^{\alpha}w(z)dz. \tag{2.19}$$

Proof. From (2.8) follows that

$$\begin{aligned} q(\mathbf{\Sigma}|\mathbf{X}) &\propto \int_0^{\infty} z^{-\frac{m(m+1-n-p-n_1p)}{2}}w(z)\int_{\mathbb{R}^{m\times p}}|\mathbf{\Sigma}|^{-\frac{n+n_1p+p}{2}} \\ &\quad \times etr\left[-\frac{z}{2}\mathbf{\Sigma}^{-1}\left(n_0(\boldsymbol{\mu}-\boldsymbol{\theta})\mathbf{\Omega}^{-1}(\boldsymbol{\mu}-\boldsymbol{\theta})'+\mathbf{\Phi}\right)\right] \\ &\quad \times etr\left[-\frac{z}{2}\mathbf{\Sigma}^{-1}\left[\mathbf{V}+n_1(\overline{\mathbf{X}}-\boldsymbol{\mu})\mathbf{\Omega}^{-1}(\overline{\mathbf{X}}-\boldsymbol{\mu})'\right]\right]d\boldsymbol{\mu}dz. \end{aligned}$$

From (2.14) and (2.15) follows that

$$\begin{aligned} q(\mathbf{\Sigma}|\mathbf{X}) &= \int_0^{\infty} z^{-\frac{m(m+1-n-p-n_1p)}{2}}w(z)|\mathbf{\Sigma}|^{-\frac{n+n_1p+p}{2}} \\ &\quad \times etr\left[-\frac{z}{2}\mathbf{\Sigma}^{-1}\left(\frac{n_1n_0}{n_1+n_0}(\overline{\mathbf{X}}-\boldsymbol{\theta})\mathbf{\Omega}^{-1}(\overline{\mathbf{X}}-\boldsymbol{\theta})'+\mathbf{\Phi}+\mathbf{V}\right)\right] \\ &\quad \times \int_{\mathbb{R}^{m\times p}} etr\left[-\frac{1}{2}\left(\frac{1}{n_1+n_0}z^{-1}\mathbf{\Sigma}\right)^{-1}(\boldsymbol{\mu}-\mathbf{b})\mathbf{\Omega}^{-1}(\boldsymbol{\mu}-\mathbf{b})'\right]d\boldsymbol{\mu}dz. \end{aligned} \tag{2.20}$$

From R.49 the latter integral can be solved as

$$\begin{aligned} &\int_{\mathbb{R}^{m\times p}} etr\left[-\frac{1}{2}\left(\frac{1}{n_1+n_0}z^{-1}\mathbf{\Sigma}\right)^{-1}(\boldsymbol{\mu}-\mathbf{b})\mathbf{\Omega}^{-1}(\boldsymbol{\mu}-\mathbf{b})'\right]d\boldsymbol{\mu} \\ &= (2\pi)^{\frac{mp}{2}}\left|\frac{1}{n_1+n_0}z^{-1}\mathbf{\Sigma}\right|^{\frac{p}{2}}|\mathbf{\Omega}|^{\frac{m}{2}}. \end{aligned} \tag{2.21}$$

Therefore from (2.17), (2.20), (2.21) and using the Taylor series expansion (see R.7) for $exp(.)$ it follows that

$$
\begin{aligned}
q(\boldsymbol{\Sigma}|\boldsymbol{X}) &\propto \int_0^\infty z^{-\frac{m(m+1-n-n_1p)}{2}} w(z)|\boldsymbol{\Sigma}|^{-\frac{n+n_1p}{2}} \\
&\quad \times etr\left[-\frac{z}{2}\boldsymbol{\Sigma}^{-1}(\frac{n_1n_0}{n_1+n_0}(\overline{\boldsymbol{X}}-\boldsymbol{\theta})\boldsymbol{\Omega}^{-1}(\overline{\boldsymbol{X}}-\boldsymbol{\theta})'+\boldsymbol{\Phi}+\boldsymbol{V})\right]dz \\
&= \int_0^\infty z^{-\frac{m(m+1-n-n_1p)}{2}} w(z)|\boldsymbol{\Sigma}|^{-\frac{n+n_1p}{2}} etr\left[-\frac{z}{2}\boldsymbol{\Sigma}^{-1}\boldsymbol{D}\right]dz \qquad (2.22)\\
&= |\boldsymbol{\Sigma}|^{-\frac{n+n_1p}{2}} \sum_{k=0}^{\infty} \frac{\left(-\frac{1}{2}tr\left[\boldsymbol{\Sigma}^{-1}\boldsymbol{D}\right]\right)^k}{k!} \int_0^\infty z^{-\frac{m(m+1-n-n_1p)-2k}{2}} w(z)dz.
\end{aligned}
$$

Hence from (2.19)

$$
q(\boldsymbol{\Sigma}|\boldsymbol{X}) \propto |\boldsymbol{\Sigma}|^{-\frac{n+n_1p}{2}} \sum_{k=0}^{\infty} \frac{\left(-\frac{1}{2}tr\left[\boldsymbol{\Sigma}^{-1}\boldsymbol{D}\right]\right)^k}{k!} \varsigma\left(-\frac{m(m+1-n-n_1p)-2k}{2}\right), \qquad (2.23)
$$

and the normalizing constant, $c_{\boldsymbol{\Sigma}}$, equals

$$
\left(\int_0^\infty z^{-\frac{m(m+1-n-n_1p)}{2}} w(z) \int_{S_m} |\boldsymbol{\Sigma}|^{-\frac{n+n_1p}{2}} etr\left[-\frac{z}{2}\boldsymbol{\Sigma}^{-1}\boldsymbol{D}\right] d\boldsymbol{\Sigma}dz\right)^{-1}.
$$

From Definition 2.1.2, since $\int_0^\infty w(z)dz = 1$, follows that

$$
\begin{aligned}
c_{\boldsymbol{\Sigma}}^{-1} &= \int_0^\infty z^{-\frac{m(m+1-n-n_1p)}{2}} w(z) \int_{S_m} |\boldsymbol{\Sigma}|^{-\frac{n+n_1p}{2}} etr\left[-\frac{z}{2}\boldsymbol{\Sigma}^{-1}\boldsymbol{D}\right] d\boldsymbol{\Sigma}dz \\
&= \int_0^\infty \left(z^{\frac{m(m+1-n-n_1p)}{2}} 2^{\frac{m(n+n_1p-m-1)}{2}} \Gamma_m\left(\frac{n+n_1p-m-1}{2}\right) |\boldsymbol{D}|^{-\frac{n+n_1p-m-1}{2}}\right) \\
&\quad \times z^{-\frac{m(m+1-n-n_1p)}{2}} w(z)dz \\
&= 2^{\frac{m(n+n_1p-m-1)}{2}} \Gamma_m\left(\frac{n+n_1p-m-1}{2}\right) |\boldsymbol{D}|^{-\frac{n+n_1p-m-1}{2}} \int_0^\infty w(z)dz \\
&= 2^{\frac{m(n+n_1p-m-1)}{2}} \Gamma_m\left(\frac{n+n_1p-m-1}{2}\right) |\boldsymbol{D}|^{-\frac{n+n_1p-m-1}{2}}, \qquad (2.24)
\end{aligned}
$$

for $n+n_1p-m-1 > 2m$. Therefore (2.18) follows. □

Now, it is of importance that the existence of $\varsigma(.)$ in (2.19) is fulfilled as a regularity condition. This is considered below.

Remark 2.4.1. *The convergence of $\varsigma(\alpha)$ is dependent on the specific weight function $w(z)$ chosen (see R.51). Consider the matrix variate t distribution with parameters $\boldsymbol{\mu} \in \mathbb{R}^{m\times p}, \boldsymbol{\Sigma} \in S_m, \boldsymbol{\Omega} \in S_p$ and ν_0 degrees of freedom, as an example. Then from Chu (1973, Table 1, p.648), the associated weight function is the inverse gamma density function (see R.40) such that $w(z) = \frac{(\frac{\nu_0}{2})^{\frac{\nu_0}{2}} z^{\frac{\nu_0}{2}-1}\exp(\frac{-\nu_0 z}{2})}{\Gamma(\frac{\nu_0}{2})}$, hence $\varsigma(\alpha)$ will only exist if $\alpha > -\frac{\nu_0}{2}$.*

Remark 2.4.2. *Under the SEL function, the Bayes estimators (see R.59) of $\boldsymbol{\mu}$ and $\boldsymbol{\Sigma}$, with prior (2.4) are the means of the posterior distributions (PM estimators) of $\boldsymbol{\mu}$ and $\boldsymbol{\Sigma}$, respectively (see R.60). Under the loss function defined by Das and Dey (2010) given in R.61, the Bayes estimators of $\boldsymbol{\mu}$ and $\boldsymbol{\Sigma}$ are the modes of the posterior distributions (MAP estimators) of $\boldsymbol{\mu}$ and $\boldsymbol{\Sigma}$, respectively.*

Theorem 2.4.3. *Under the SEL function, the Bayes estimator of $\boldsymbol{\mu}$ for the normal model with prior (2.4) is*

$$\widehat{\boldsymbol{\mu}}_{B,W^{-1}} = \boldsymbol{b} = \frac{1}{n_1+n_0}\left(n_1\overline{\boldsymbol{X}} - n_0\boldsymbol{\theta}\right). \tag{2.25}$$

Proof. From Theorem 2.4.1 and R.50 the Bayes estimator for $\boldsymbol{\mu}$ is

$$\widehat{\boldsymbol{\mu}}_{B,W^{-1}} = E\left[\boldsymbol{\mu}|\boldsymbol{X}\right] = \boldsymbol{b}.$$

□

Lemma 1. *The r^{th} posterior moment of $|\boldsymbol{\Sigma}|$ for the normal model with prior (2.4) is*

$$m_r = E\left[|\boldsymbol{\Sigma}|^r|\boldsymbol{X};z\right] = \frac{2^{-mr}\varsigma(mr)\Gamma_m\left(\frac{n+n_1p-m-1-2r}{2}\right)}{\Gamma_m\left(\frac{n+n_1p-m-1}{2}\right)}|\boldsymbol{D}|^r, \tag{2.26}$$

with $\boldsymbol{D}$ as defined in (2.17), provided $\varsigma(mr)$ as in (2.19) exists and $n+n_1p-m-1>2m$.

Proof. From (2.22) and (2.24) the rth posterior moment is given by

$$\begin{aligned}
E\left[|\boldsymbol{\Sigma}|^r|\boldsymbol{X};z\right] &= \frac{2^{-\frac{m(n+n_1p-m-1)}{2}}}{\Gamma_m\left(\frac{n+n_1p-m-1}{2}\right)}|\boldsymbol{D}|^{\frac{n+n_1p-m-1}{2}}\int_0^\infty z^{-\frac{m(m+1-n-n_1p)}{2}}w(z) \\
&\quad\times\int_{S_m}|\boldsymbol{\Sigma}|^{-\frac{n+n_1p-2r}{2}}etr\left[-\frac{z}{2}\boldsymbol{\Sigma}^{-1}\boldsymbol{D}\right]d\boldsymbol{\Sigma}dz \\
&= \frac{2^{-\frac{m(n+n_1p-m-1)}{2}}}{\Gamma_m\left(\frac{n+n_1p-m-1}{2}\right)}|\boldsymbol{D}|^{\frac{n+n_1p-m-1}{2}}\int_0^\infty z^{-\frac{m(m+1-n-n_1p)}{2t}}w(z) \\
&\quad\times z^{\frac{m(m+1-n-n_1p+2r)}{2}}\Gamma_m\left(\frac{n+n_1p-m-1-2r}{2}\right) \\
&\quad\times 2^{\frac{m(n+n_1p-m-1-2r)}{2}}|\boldsymbol{D}|^{-\frac{n+n_1p-m-1-2r}{2}}dz \\
&= \frac{2^{-mr}\Gamma_m\left(\frac{n+n_1p-m-1-2r}{2}\right)}{\Gamma_m\left(\frac{n+n_1p-m-1}{2}\right)}|\boldsymbol{D}|^r\int_0^\infty z^{mr}w(z)dz \\
&= \frac{2^{-mr}\varsigma(mr)\Gamma_m\left(\frac{n+n_1p-m-1-2r}{2}\right)}{\Gamma_m\left(\frac{n+n_1p-m-1}{2}\right)}|\boldsymbol{D}|^r,
\end{aligned} \tag{2.27}$$

from Definition 2.1.2 and (2.19) with $n+n_1p-m-1-2r>2m$. □

Theorem 2.4.4. *Under the SEL function, the Bayes estimator of $|\mathbf{\Sigma}|$ for the normal model with prior (2.4) is*

$$\widehat{|\mathbf{\Sigma}|}_{B,W^{-1}} = \frac{2^{-m}\varsigma(m)\Gamma_m\left(\frac{n+n_1p-m-3}{2}\right)}{\Gamma_m\left(\frac{n+n_1p-m-1}{2}\right)}|\boldsymbol{D}|, \tag{2.28}$$

with $\boldsymbol{D}$ as defined in (2.17), provided $\varsigma(m)$ as in (2.19) exists and $n+n_1p-m-1>2m$.

Proof. The result is immediate from (2.26) with $r=1$. □

2.4.2 Normal-Wishart prior

In this section the normal-Wishart prior for the matrix variate elliptical model is considered.

The prior distributions for $\boldsymbol{\mu}$ and $\mathbf{\Psi}=z^{-1}\mathbf{\Sigma}$ respectively, are from R.49 and Definition 2.1.1, $\boldsymbol{\mu}|\mathbf{\Psi}\sim N_{m,p}(\boldsymbol{\theta},\frac{1}{n_0}\mathbf{\Psi}\otimes\mathbf{\Omega})$ and $\mathbf{\Psi}\sim W_m(\mathbf{\Phi},n)$ with density functions (2.2) and

$$\pi(\mathbf{\Psi}) = 2^{-\frac{nm}{2}}\left[\Gamma_m\left(\frac{n}{2}\right)\right]^{-1}|\mathbf{\Phi}|^{-\frac{n}{2}}|\mathbf{\Psi}|^{\frac{n}{2}-\frac{m+1}{2}}etr\left[-\frac{1}{2}\mathbf{\Psi}\mathbf{\Phi}^{-1}\right], \quad \mathbf{\Psi},\mathbf{\Phi}\in S_m \text{ and } n\geq m.$$

The joint prior density function is then

$$\pi(\boldsymbol{\mu},\mathbf{\Psi}) \propto |\mathbf{\Psi}|^{\frac{n-m-1-p}{2}}etr\left[-\frac{n_0}{2}\mathbf{\Psi}^{-1}(\boldsymbol{\mu}-\boldsymbol{\theta})\mathbf{\Omega}^{-1}(\boldsymbol{\mu}-\boldsymbol{\theta})'\right]etr\left[-\frac{1}{2}\mathbf{\Psi}\mathbf{\Phi}^{-1}\right]. \tag{2.29}$$

It then follows that

$$\pi(\boldsymbol{\mu},\mathbf{\Sigma}|z) \propto z^{-\frac{m(n-p)}{2}}|\mathbf{\Sigma}|^{\frac{n-m-1-p}{2}}etr\left[-\frac{1}{2z}\mathbf{\Sigma}\mathbf{\Phi}^{-1}\right]etr\left[-\frac{n_0z}{2}\mathbf{\Sigma}^{-1}(\boldsymbol{\mu}-\boldsymbol{\theta})\mathbf{\Omega}^{-1}(\boldsymbol{\mu}-\boldsymbol{\theta})'\right]. \tag{2.30}$$

From (2.6) and (2.30) the joint posterior density function follows as,

$$\begin{aligned} q(\boldsymbol{\mu},\mathbf{\Sigma}|\boldsymbol{X}) \propto & \int_0^\infty z^{-\frac{m(n-p-n_1p)}{2}}w(z)|\mathbf{\Sigma}|^{\frac{n-m-1-p-n_1p}{2}}etr\left[-\frac{1}{2z}\mathbf{\Sigma}\mathbf{\Phi}^{-1}\right] \\ & \times etr\left[-\frac{n_0z}{2}\mathbf{\Sigma}^{-1}(\boldsymbol{\mu}-\boldsymbol{\theta})\mathbf{\Omega}^{-1}(\boldsymbol{\mu}-\boldsymbol{\theta})'\right] \\ & \times etr\left[-\frac{z}{2}\mathbf{\Sigma}^{-1}\left[\boldsymbol{V}+n(\overline{\boldsymbol{X}}-\boldsymbol{\mu})\mathbf{\Omega}^{-1}(\overline{\boldsymbol{X}}-\boldsymbol{\mu})'\right]\right]dz \end{aligned} \tag{2.31}$$

with $\boldsymbol{V}$ as defined in (2.7).

Theorem 2.4.5. *The marginal posterior density function of $\boldsymbol{\mu}$ for the normal model with prior (2.30) is*

$$q(\boldsymbol{\mu}|\boldsymbol{X}) = \frac{|\boldsymbol{\Omega}|^{-\frac{m}{2}}(n_1+n_0)^{\frac{mp}{2}}}{(2\pi)^{\frac{mp}{2}} B_{\frac{-n+n_1 p}{2}}\left(\frac{1}{4}\boldsymbol{\Phi}^{-1}\boldsymbol{Y}\right)}|2\boldsymbol{\Phi}|^{-\frac{p}{2}} \times B_{\frac{-n+p+n_1 p}{2}}\left(\frac{1}{4}\boldsymbol{\Phi}^{-1}\left(\boldsymbol{Y}+(n_1+n_0)(\boldsymbol{\mu}-\boldsymbol{b})\boldsymbol{\Omega}^{-1}(\boldsymbol{\mu}-\boldsymbol{b})'\right)\right) \tag{2.32}$$

with $n-p-n_1 p > m-1$, $\boldsymbol{Y} = \frac{n_1 n_0}{n_1+n_0}(\overline{\boldsymbol{X}}-\boldsymbol{\theta})\boldsymbol{\Omega}^{-1}(\overline{\boldsymbol{X}}-\boldsymbol{\theta})'+\boldsymbol{V}$, $\boldsymbol{V}$ *as defined in* (2.7) *and* $B_\delta(.)$ *the Bessel function of the second kind with matrix argument (see R.15).*

Proof. The marginal posterior density for $\boldsymbol{\mu}$ follows from (2.31) as

$$\begin{aligned} q(\boldsymbol{\mu}|\boldsymbol{X}) \propto & \int_0^\infty z^{-\frac{m(n-p-n_1 p)}{2}} w(z) \int_{S_m} |\boldsymbol{\Sigma}|^{\frac{n-m-1-p-n_1 p}{2}} etr\left[-\frac{1}{2z}\boldsymbol{\Sigma}\boldsymbol{\Phi}^{-1}\right] \\ & \times etr\left[-\frac{n_0 z}{2}\boldsymbol{\Sigma}^{-1}(\boldsymbol{\mu}-\boldsymbol{\theta})\boldsymbol{\Omega}^{-1}(\boldsymbol{\mu}-\boldsymbol{\theta})'\right] \\ & \times etr\left[-\frac{z}{2}\boldsymbol{\Sigma}^{-1}\left[\boldsymbol{V}+n_1(\overline{\boldsymbol{X}}-\boldsymbol{\mu})\boldsymbol{\Omega}^{-1}(\overline{\boldsymbol{X}}-\boldsymbol{\mu})'\right]\right] d\boldsymbol{\Sigma} dz. \end{aligned} \tag{2.33}$$

Let

$$\boldsymbol{F} = n_0(\boldsymbol{\mu}-\boldsymbol{\theta})\boldsymbol{\Omega}^{-1}(\boldsymbol{\mu}-\boldsymbol{\theta})' \tag{2.34}$$

Applying R.15, it follows from (2.33), with $\boldsymbol{A}$ as defined in (2.11), that

$$\begin{aligned} q(\boldsymbol{\mu}|\boldsymbol{X}) \propto & \int_0^\infty z^{-\frac{m(n-p-n_1 p)}{2}} w(z) \int_{S_m} |\boldsymbol{\Sigma}|^{\frac{n-m-1-p-n_1 p}{2}} \\ & \times etr\left[-\frac{1}{2z}\boldsymbol{\Sigma}\boldsymbol{\Phi}^{-1}\right] etr\left[-\frac{z}{2}\boldsymbol{\Sigma}^{-1}(\boldsymbol{F}+\boldsymbol{A})\right] d\boldsymbol{\Sigma} dz \\ = & \int_0^\infty z^{-\frac{m(n-p-n_1 p)}{2}} w(z) B_{\frac{-n+p+n_1 p}{2}}\left(\frac{1}{4}\boldsymbol{\Phi}^{-1}(\boldsymbol{F}+\boldsymbol{A})\right)|2z\boldsymbol{\Phi}|^{\frac{n-p-n_1 p}{2}} dz \\ \propto & \int_0^\infty z^{-\frac{m(n-p-n_1 p)}{2}} z^{\frac{m(n-p-n_1 p)}{2}} w(z) B_{\frac{-n+p+n_1 p}{2}}\left(\frac{1}{4}\boldsymbol{\Phi}^{-1}(\boldsymbol{F}+\boldsymbol{A})\right) dz \\ = & B_{\frac{-n+p+n_1 p}{2}}\left(\frac{1}{4}\boldsymbol{\Phi}^{-1}(\boldsymbol{F}+\boldsymbol{A})\right), \end{aligned} \tag{2.35}$$

since $\int_0^\infty w(z)dz = 1$ with $n-p-n_1 p > m-1$. From (2.11), (2.14) and (2.34) follows that

$$\begin{aligned} \boldsymbol{F}+\boldsymbol{A} &= (n_1+n_0)(\boldsymbol{\mu}-\boldsymbol{b})\boldsymbol{\Omega}^{-1}(\boldsymbol{\mu}-\boldsymbol{b})' + \frac{n_1 n_0}{n_1+n_0}(\overline{\boldsymbol{X}}-\boldsymbol{\theta})\boldsymbol{\Omega}^{-1}(\overline{\boldsymbol{X}}-\boldsymbol{\theta})'+\boldsymbol{V} \\ &= \boldsymbol{Y}+(n_1+n_0)(\boldsymbol{\mu}-\boldsymbol{b})\boldsymbol{\Omega}^{-1}(\boldsymbol{\mu}-\boldsymbol{b})', \end{aligned} \tag{2.36}$$

where

$$\boldsymbol{Y} = \frac{n_1 n_0}{n_1+n_0}(\overline{\boldsymbol{X}}-\boldsymbol{\theta})\boldsymbol{\Omega}^{-1}(\overline{\boldsymbol{X}}-\boldsymbol{\theta})'+\boldsymbol{V}, \tag{2.37}$$

with $\boldsymbol{b}$ as defined in (2.15). Therefore from (2.35) and (2.36) follows

$$q(\boldsymbol{\mu}|\boldsymbol{X}) = c_{\boldsymbol{\mu}} B_{\frac{-n+p+n_1p}{2}}\left(\frac{1}{4}\boldsymbol{\Phi}^{-1}\left(\boldsymbol{Y} + (n_1+n_0)(\boldsymbol{\mu}-\boldsymbol{b})\boldsymbol{\Omega}^{-1}(\boldsymbol{\mu}-\boldsymbol{b})'\right)\right).$$

The normalizing constant, $c_{\boldsymbol{\mu}}$, is calculated by using the integral presentation of the Bessel function as follows (see R.15)

$$\begin{aligned}
c_{\boldsymbol{\mu}}^{-1} &= \int_{\mathbb{R}^{m\times p}} B_{\frac{-n+p+n_1p}{2}}\left(\frac{1}{4}\boldsymbol{\Phi}^{-1}\left(\boldsymbol{Y} + (n_1+n_0)(\boldsymbol{\mu}-\boldsymbol{b})\boldsymbol{\Omega}^{-1}(\boldsymbol{\mu}-\boldsymbol{b})'\right)\right) d\boldsymbol{\mu} \\
&= \int_{\mathbb{R}^{m\times p}}\int_{S_m} etr\left(-\frac{1}{2}\boldsymbol{U}^{-1}\left(\boldsymbol{Y} + (n_1+n_0)(\boldsymbol{\mu}-\boldsymbol{b})\boldsymbol{\Omega}^{-1}(\boldsymbol{\mu}-\boldsymbol{b})'\right)\right) \\
&\quad \times etr\left(-\frac{1}{2}\boldsymbol{\Phi}\boldsymbol{U}\right)|\boldsymbol{U}|^{-\frac{-n+p+n_1p}{2}-\frac{m+1}{2}}\left|\frac{1}{2}\boldsymbol{\Phi}^{-1}\right|^{-\frac{-n+p+n_1p}{2}} d\boldsymbol{U}d\boldsymbol{\mu} \\
&= |2\boldsymbol{\Phi}|^{\frac{-n+p+n_1p}{2}}\int_{S_m} etr\left(-\frac{1}{2}\boldsymbol{\Phi}\boldsymbol{U}\right) etr\left(-\frac{1}{2}\boldsymbol{U}^{-1}\boldsymbol{Y}\right)|\boldsymbol{U}|^{-\frac{-n+p+n_1p}{2}-\frac{m+1}{2}} \\
&\quad \times \int_{\mathbb{R}^{m\times p}} etr\left(-\frac{n_1+n_0}{2}\boldsymbol{U}^{-1}(\boldsymbol{\mu}-\boldsymbol{b})\boldsymbol{\Omega}^{-1}(\boldsymbol{\mu}-\boldsymbol{b})'\right) d\boldsymbol{\mu}d\boldsymbol{U}.
\end{aligned}$$

Note that from R.49,

$$\int_{\mathbb{R}^{m\times p}} etr\left(-\frac{n_1+n_0}{2}\boldsymbol{U}^{-1}(\boldsymbol{\mu}-\boldsymbol{b})\boldsymbol{\Omega}^{-1}(\boldsymbol{\mu}-\boldsymbol{b})'\right) d\boldsymbol{\mu} = (2\pi)^{\frac{mp}{2}}\left|\frac{1}{n_1+n_0}\boldsymbol{U}\right|^{\frac{p}{2}}|\boldsymbol{\Omega}|^{\frac{m}{2}}.$$

Using R.15 it follows that $c_{\boldsymbol{\mu}}^{-1}$ equals

$$\begin{aligned}
& (2\pi)^{\frac{mp}{2}}|\boldsymbol{\Omega}|^{\frac{m}{2}}(n_1+n_0)^{-\frac{mp}{2}}|2\boldsymbol{\Phi}|^{\frac{-n+p+n_1p}{2}} \\
& \times \int_{S_m} etr\left(-\frac{1}{2}\boldsymbol{\Phi}^{-1}\boldsymbol{U}\right) etr\left(-\frac{1}{2}\boldsymbol{U}^{-1}\boldsymbol{Y}\right)|\boldsymbol{U}|^{-\frac{-n+n_1p}{2}-\frac{m+1}{2}} d\boldsymbol{U} \\
= \; & (2\pi)^{\frac{mp}{2}}|\boldsymbol{\Omega}|^{\frac{m}{2}}(n_1+n_0)^{-\frac{mp}{2}}|2\boldsymbol{\Phi}|^{\frac{p}{2}} B_{\frac{-n+n_1p}{2}}\left(\frac{1}{4}\boldsymbol{\Phi}^{-1}\boldsymbol{Y}\right),
\end{aligned}$$

with $n - n_1 p > m-1$. □

Remark 2.4.3. *The marginal posterior distribution of the location matrix in matrix elliptical models is robust with respect to departures from normality, under the non-conjugate normal-Wishart prior since the marginal posterior distribution of the location matrix (Theorem 2.4.5) is independent of $w(z)$.*

Theorem 2.4.6. *The marginal posterior density function of $\mathbf{\Sigma}$ for the normal model with prior (2.30) is*

$$\begin{aligned} q(\mathbf{\Sigma}|\mathbf{X}) &= \frac{2^{\frac{m(-n+n_1p)}{2}}}{B_{\frac{-n+n_1p}{2}}\left(\frac{1}{4}\mathbf{\Phi}^{-1}\mathbf{Y}\right)}|\mathbf{\Phi}|^{\frac{-n+n_1p}{2}}|\mathbf{\Sigma}|^{\frac{n-m-1-n_1p}{2}} \\ &\times\sum_{k=0}^{\infty}\sum_{l=0}^{\infty}\frac{(-2)^{-k-l}\left(tr\left[\mathbf{\Sigma\Phi}^{-1}\right]\right)^k\left(tr\left[\mathbf{\Sigma}^{-1}\mathbf{Y}\right]\right)^l}{k!l!} \\ &\times\varsigma\left(-\frac{m(n-n_1p)}{2}-k+l\right), \end{aligned} \tag{2.38}$$

with $\mathbf{V}$ as define din (2.7), $\mathbf{Y}$ as defined in (2.37) and $\mathbf{\Sigma}\in S_m$, provided $\varsigma\left(-\frac{m(n-n_1p)}{2}-k+l\right)$ as in (2.19) exists and $n-n_1p>m-1$.

Proof. From (2.31) and (2.14), and applying R.49 follows

$$\begin{aligned} q(\mathbf{\Sigma}|\mathbf{X}) &\propto \int_0^\infty z^{-\frac{m(n-p-n_1p)}{2}}w(z)|\mathbf{\Sigma}|^{\frac{n-m-1-p-n_1p}{2}}etr\left[-\frac{1}{2z}\mathbf{\Sigma\Phi}^{-1}\right] \\ &\times\int_{\mathbb{R}^{m\times p}}etr\left[-\frac{n_0z}{2}\mathbf{\Sigma}^{-1}(\boldsymbol{\mu}-\boldsymbol{\theta})\mathbf{\Omega}^{-1}(\boldsymbol{\mu}-\boldsymbol{\theta})'\right] \\ &\times etr\left[-\frac{z}{2}\mathbf{\Sigma}^{-1}\left[\mathbf{V}+n_1(\bar{\mathbf{X}}-\boldsymbol{\mu})\mathbf{\Omega}^{-1}(\bar{\mathbf{X}}-\boldsymbol{\mu})'\right]\right]d\boldsymbol{\mu}dz \\ &= \int_0^\infty z^{-\frac{m(n-p-n_1p)}{2}}w(z)|\mathbf{\Sigma}|^{\frac{n-m-1-p-n_1p-}{2}}etr\left[-\frac{1}{2z}\mathbf{\Sigma\Phi}^{-1}\right] \\ &\times etr\left[-\frac{z}{2}\mathbf{\Sigma}^{-1}\left(\frac{n_1n_0}{n_1+n_0}(\bar{\mathbf{X}}-\boldsymbol{\theta})\mathbf{\Omega}^{-1}(\bar{\mathbf{X}}-\boldsymbol{\theta})'+\mathbf{V}\right)\right] \\ &\times\int_{\mathbb{R}^{m\times p}}etr\left[\left(\frac{1}{n_1+n_0}z^{-1}\mathbf{\Sigma}\right)^{-1}(\boldsymbol{\mu}-\mathbf{b})\mathbf{\Omega}^{-1}(\boldsymbol{\mu}-\mathbf{b})'\right]d\boldsymbol{\mu}dz \\ &\propto \int_0^\infty z^{-\frac{m(n-p-n_1p)}{2}}w(z)|\mathbf{\Sigma}|^{\frac{n-m-1-p-n_1p}{2}}etr\left[-\frac{1}{2z}\mathbf{\Sigma\Phi}^{-1}\right] \\ &\times etr\left[-\frac{z}{2}\mathbf{\Sigma}^{-1}\mathbf{Y}\right]|\frac{1}{n_1+n_0}z^{-1}\mathbf{\Sigma}|^{\frac{p}{2}}dz \\ &\propto \int_0^\infty z^{-\frac{m(n-n_1p)}{2}}w(z)|\mathbf{\Sigma}|^{\frac{n-m-1-n_1p}{2}}etr\left[-\frac{1}{2z}\mathbf{\Sigma\Phi}^{-1}\right]etr\left[-\frac{z}{2}\mathbf{\Sigma}^{-1}\mathbf{Y}\right]dz \qquad (2.39) \\ &= |\mathbf{\Sigma}|^{\frac{n-m-1-n_1p}{2}}\sum_{k=0}^{\infty}\sum_{l=0}^{\infty}\frac{(-2)^{-k-l}\left(tr\left[\mathbf{\Sigma\Phi}^{-1}\right]\right)^k\left(tr\left[\mathbf{\Sigma}^{-1}\mathbf{Y}\right]\right)^l}{k!l!} \\ &\times\int_0^\infty z^{-\frac{m(n-n_1p)}{2}-k+l}w(z)dz \\ &= |\mathbf{\Sigma}|^{\frac{n-m-1-n_1p}{2}}\sum_{k=0}^{\infty}\sum_{l=0}^{\infty}\frac{(-2)^{-k-l}\left(tr\left[\mathbf{\Sigma\Phi}^{-1}\right]\right)^k\left(tr\left[\mathbf{\Sigma}^{-1}\mathbf{Y}\right]\right)^l}{k!l!}\varsigma\left(-\frac{m(n-n_1p)}{2}-k+l\right), \end{aligned}$$

from the Taylor series expansion for $exp(.)$ (see R.7) and (2.19). From (2.39), using R.15 follows that

$$\begin{aligned} c_{\boldsymbol{\Sigma}}^{-1} &= \int_0^\infty z^{-\frac{m(n-n_1p)}{2}} w(z) \int_{S_m} |\boldsymbol{\Sigma}|^{\frac{n-m-1-n_1p}{2}} etr\left[-\frac{1}{2z}\boldsymbol{\Sigma}\boldsymbol{\Phi}^{-1}\right] etr\left[-\frac{z}{2}\boldsymbol{\Sigma}^{-1}\boldsymbol{Y}\right] d\boldsymbol{\Sigma} dz \\ &= \int_0^\infty z^{-\frac{m(n-n_1p)}{2}} w(z) |2z\boldsymbol{\Phi}|^{\frac{n-n_1p}{2}} B_{\frac{-n+n_1p}{2}}\left(\frac{1}{4}\boldsymbol{\Phi}^{-1}\boldsymbol{Y}\right) dz \\ &= 2^{\frac{m(n-n_1p)}{2}} |\boldsymbol{\Phi}|^{\frac{n-n_1p}{2}} B_{\frac{-n+n_1p}{2}}\left(\frac{1}{4}\boldsymbol{\Phi}^{-1}\boldsymbol{Y}\right)\left(\int_0^\infty w(z)dz\right) \\ &= 2^{\frac{m(n-n_1p)}{2}} |\boldsymbol{\Phi}|^{\frac{n-n_1p}{2}} B_{\frac{-n+n_1p}{2}}\left(\frac{1}{4}\boldsymbol{\Phi}^{-1}\boldsymbol{Y}\right) \end{aligned} \tag{2.40}$$

with $n-n_1p > m-1$, since $\int_0^\infty w(z)dz = 1$. □

Theorem 2.4.7. *Under the SEL function, the Bayes estimator of $\boldsymbol{\mu}$ for the normal model with prior (2.30) is*

$$\widehat{\boldsymbol{\mu}}_{B,W} = \boldsymbol{b} = \frac{1}{n_1+n_0}\left(n_1\overline{\boldsymbol{X}} - n_0\boldsymbol{\theta}\right), \tag{2.41}$$

with $n-p-n_1p > m-1$, which is the same as under prior (2.4). See Remark 2.4.3.

Proof. Under the SEL function the Bayes estimator of $\boldsymbol{\mu}$ is the posterior mean, i.e. $\widehat{\boldsymbol{\mu}}_{B,W} = E\left[\boldsymbol{\mu}|\boldsymbol{X}\right]$. Note that the expected value of $(\boldsymbol{\mu}-\boldsymbol{b})$ is

$$\begin{aligned} E\left[\boldsymbol{\mu}-\boldsymbol{b}|\boldsymbol{X}\right] &= \frac{|\boldsymbol{\Omega}|^{-\frac{m}{2}}(n_1+n_0)^{\frac{mp}{2}}}{(2\pi)^{\frac{mp}{2}} B_{\frac{-n+n_1p}{2}}\left(\frac{1}{4}\boldsymbol{\Phi}^{-1}\boldsymbol{Y}\right)} |2\boldsymbol{\Phi}|^{-\frac{p}{2}} \int_{\mathbb{R}^{m\times p}} (\boldsymbol{\mu}-\boldsymbol{b}) \\ &\quad \times B_{\frac{-n+p+n_1p}{2}}\left(\frac{1}{4}\boldsymbol{\Phi}^{-1}\left(\boldsymbol{Y}+(n_1+n_0)(\boldsymbol{\mu}-\boldsymbol{b})\boldsymbol{\Omega}^{-1}(\boldsymbol{\mu}-\boldsymbol{b})'\right)\right) d\boldsymbol{\mu}, \end{aligned}$$

with $\boldsymbol{Y}$ as in (2.37). This is an integral of an odd function and hence $E\left[\boldsymbol{\mu}-\boldsymbol{b}|\boldsymbol{X}\right] = \boldsymbol{0}$. Therefore the Bayes estimator of $\boldsymbol{\mu}$, is $\widehat{\boldsymbol{\mu}}_{B,W} = \boldsymbol{b} = \frac{1}{n_1+n_0}\left(n_1\overline{\boldsymbol{X}} - n_0\boldsymbol{\theta}\right)$. □

Lemma 2. *Under the SEL function, the Bayes estimator of $|\boldsymbol{\Sigma}|^r$ for the normal model with prior (2.30) is*

$$m_r = E\left[|\boldsymbol{\Sigma}|^r|\overline{\boldsymbol{X}},\boldsymbol{V}\right] = \frac{2^{mr}\varsigma(mr) B_{\frac{-n+n_1p-2r}{2}}\left(\frac{1}{4}\boldsymbol{\Phi}^{-1}\boldsymbol{Y}\right)}{B_{\frac{-n+n_1p}{2}}\left(\frac{1}{4}\boldsymbol{\Phi}^{-1}\boldsymbol{Y}\right)} |\boldsymbol{\Phi}|^r, \tag{2.42}$$

with $\boldsymbol{Y}$ as defined in (2.37), provided $\varsigma(mr)$ as in (2.19) exists and $n-n_1p > m-1$.

Proof. From (2.39) and (2.40), the rth posterior moment of $|\mathbf{\Sigma}|$ is given by

$$\begin{aligned}
& E\left[|\mathbf{\Sigma}|^r|\boldsymbol{X};z\right] \\
= & \frac{2^{\frac{m(-n+n_1p)}{2}}}{B_{\frac{-n+n_1p}{2}}\left(\frac{1}{4}\mathbf{\Phi}^{-1}\boldsymbol{Y}\right)}|\mathbf{\Phi}|^{\frac{-n+n_1p}{2}}\int_0^\infty z^{-\frac{m(n-n_1p)}{2}}w(z) \\
& \times\int_{S_m}|\mathbf{\Sigma}|^{\frac{n-m-1-n_1p+2r}{2}}etr\left[-\frac{1}{2z}\mathbf{\Sigma}\mathbf{\Phi}^{-1}\right]etr\left[-\frac{z}{2}\mathbf{\Sigma}^{-1}\boldsymbol{Y}\right]d\mathbf{\Sigma}dz \\
= & \frac{2^{\frac{m(-n+n_1p)}{2}}}{B_{\frac{-n+n_1p}{2}}\left(\frac{1}{4}\mathbf{\Phi}^{-1}\boldsymbol{Y}\right)}|\mathbf{\Phi}|^{\frac{-n+n_1p}{2}}\int_0^\infty z^{-\frac{m(n-n_1p)}{2}}w(z) \\
& \times|2z\mathbf{\Phi}|^{\frac{n-n_1p+2r}{2}}B_{\frac{-n+n_1p-2r}{2}}\left(\frac{1}{4}\mathbf{\Phi}^{-1}\boldsymbol{Y}\right)dz \\
= & \frac{2^{mr}B_{\frac{-n+n_1p-2r}{2}}\left(\frac{1}{4}\mathbf{\Phi}^{-1}\boldsymbol{Y}\right)}{B_{\frac{-n+n_1p}{2}}\left(\frac{1}{4}\mathbf{\Phi}^{-1}\boldsymbol{Y}\right)}|\mathbf{\Phi}|^r\int_0^\infty z^{mr}w(z)dz,
\end{aligned} \tag{2.43}$$

from R.15. Using (2.19), the proof is complete. □

Theorem 2.4.8. *Under the SEL function, the Bayes estimator of $|\mathbf{\Sigma}|$ for the normal model with prior (2.30) is*

$$\widehat{|\mathbf{\Sigma}|}_{B,W} = \frac{2^m\varsigma(m)B_{\frac{-n+n_1p-2}{2}}\left(\frac{1}{4}\mathbf{\Phi}^{-1}\boldsymbol{Y}\right)}{B_{\frac{-n+n_1p}{2}}\left(\frac{1}{4}\mathbf{\Phi}^{-1}\boldsymbol{Y}\right)}|\mathbf{\Phi}|, \tag{2.44}$$

with $\boldsymbol{Y}$ as defined in (2.37), provided $\varsigma(m)$ as in (2.19) exists and $n-n_1p>m-1$.

Proof. The result is immediate from (2.42) with $r=1$. □

2.4.3 Particular subfamilies

In this section the newly developed results will be applied to the matrix variate normal distribution and the matrix variate t distribution as special cases of the matrix variate elliptical model.

Remark 2.4.4. *(See Remark 2.4.3).The marginal posterior distribution of $\boldsymbol{\mu}$ for all matrix variate elliptical distributions and a normal-inverse Wishart prior is given in (2.9) and for the normal-Wishart prior in (2.32). The Bayes estimator of $\boldsymbol{\mu}$ for both prior structures considered, is from equations (2.25) and (2.41),*

$$\widehat{\boldsymbol{\mu}}_B = \boldsymbol{b} = \frac{1}{n_1+n_0}\left(n_1\overline{\boldsymbol{X}} - n_0\boldsymbol{\theta}\right)$$

2.4.3.1. Matrix variate normal distribution

Let $\boldsymbol{X}$ follow a matrix variate normal distribution (R.49) with parameters $\boldsymbol{\mu}\in\mathbb{R}^{m\times p}$ and $\mathbf{\Sigma}\otimes\mathbf{\Omega}$ with $\mathbf{\Sigma}\in S_m,\mathbf{\Omega}\in S_p$. Then from (Chu 1973, Table 1, p.648), the associated weight function is:

$$w(z) = \delta(z-1), \tag{2.45}$$

where $\delta(\cdot)$ is the Dirac delta function (see R.13).

- First consider the normal-inverse Wishart case. Note that from (2.8) and (2.45) follows that

$$q(\boldsymbol{\mu}|\boldsymbol{X},\boldsymbol{\Sigma}) \quad \propto \quad etr\left[-\frac{1}{2}\left(\frac{1}{n_1+n_0}\boldsymbol{\Sigma}\right)^{-1}(\boldsymbol{\mu}-\boldsymbol{b})\boldsymbol{\Omega}^{-1}(\boldsymbol{\mu}-\boldsymbol{b})'\right],$$

hence $\boldsymbol{\mu}|\boldsymbol{X},\boldsymbol{\Sigma} \sim N_{m,p}(\boldsymbol{b},\frac{1}{n_1+n_0}\boldsymbol{\Sigma}\otimes\boldsymbol{\Omega})$, and

$$q(\boldsymbol{\Sigma}|\boldsymbol{X},\boldsymbol{\mu}) \quad \propto \quad |\boldsymbol{\Sigma}|^{-\frac{n+n_1p+p}{2}} etr\left[-\frac{1}{2}\boldsymbol{\Sigma}^{-1}(\boldsymbol{A}+\boldsymbol{B})\right],$$

hence $\boldsymbol{\Sigma}|\boldsymbol{X},\boldsymbol{\mu} \sim W^{-1}(\boldsymbol{A}+\boldsymbol{B},n+n_1p+p)$, with $\boldsymbol{A}$ and $\boldsymbol{B}$ defined in (2.11) and (2.12), respectively. The marginal posterior density function of $\boldsymbol{\Sigma}$ for the normal-inverse Wishart prior is obtained by using (2.22) and (2.45),

$$\begin{aligned} q(\boldsymbol{\Sigma}|\boldsymbol{X}) \quad &= \quad \frac{2^{-\frac{m(n+n_1p-m-1)}{2}}}{\Gamma_m\left(\frac{n+n_1p-m-1}{2}\right)}|\boldsymbol{D}|^{\frac{n+n_1p-m-1}{2}} \\ &\quad \times|\boldsymbol{\Sigma}|^{-\frac{n+n_1p}{2}}\int_0^\infty z^{-\frac{m(m+1-n-n_1p)}{2}}\delta(z-1)etr\left[-\frac{z}{2}\boldsymbol{\Sigma}^{-1}\boldsymbol{D}\right]dz, \end{aligned}$$

with $\boldsymbol{D}$ as defined in (2.17). Note that $\int_0^\infty f(x)\delta(x)dx = f(0)$ from R.13 with $x=z-1$ and $f(x)=(x+1)^{-\frac{m(m+1-n-n_1p)}{2}}etr\left[-\frac{x+1}{2}\boldsymbol{\Sigma}^{-1}\boldsymbol{D}\right]$. Therefore $f(0)=etr\left[-\frac{1}{2}\boldsymbol{\Sigma}^{-1}\boldsymbol{D}\right]$, and

$$\begin{aligned} q(\boldsymbol{\Sigma}|\boldsymbol{X}) \quad &= \quad \frac{2^{-\frac{m(n+n_1p-m-1)}{2}}}{\Gamma_m\left(\frac{n+n_1p-m-1}{2}\right)}|\boldsymbol{D}|^{\frac{n+n_1p-m-1}{2}}|\boldsymbol{\Sigma}|^{-\frac{n+n_1p}{2}}f(0) \\ &= \quad \frac{2^{-\frac{m(n+n_1p-m-1)}{2}}}{\Gamma_m\left(\frac{n+n_1p-m-1}{2}\right)}|\boldsymbol{D}|^{\frac{n+n_1p-m-1}{2}}|\boldsymbol{\Sigma}|^{-\frac{n+n_1p}{2}}etr\left[-\frac{1}{2}\boldsymbol{\Sigma}^{-1}\boldsymbol{D}\right]. \end{aligned}$$

Therefore $\boldsymbol{\Sigma}|\boldsymbol{X} \sim W_m^{-1}(\boldsymbol{D},n+n_1p)$ (see Definition 2.1.2). From (2.24), (2.28) and (2.45) the Bayes estimator of $|\boldsymbol{\Sigma}|$ is

$$|\widehat{\boldsymbol{\Sigma}}|_{B,W^{-1}} = \frac{2^{-m}\Gamma_m\left(\frac{n+n_1p-m-3}{2}\right)}{\Gamma_m\left(\frac{n+n_1p-m-1}{2}\right)}|\boldsymbol{D}|\int_0^\infty z^m w(z)dz = \frac{2^{-m}\Gamma_m\left(\frac{n+n_1p-m-3}{2}\right)}{\Gamma_m\left(\frac{n+n_1p-m-1}{2}\right)}|\boldsymbol{D}|. \quad (2.46)$$

- Secondly, the normal-Wishart prior is considered. Note that from (2.31) and (2.45) follows that

$$q(\boldsymbol{\mu}|\boldsymbol{X},\boldsymbol{\Sigma}) \quad \propto \quad etr\left[-\frac{1}{2}\left(\frac{1}{n_1+n_0}\boldsymbol{\Sigma}\right)^{-1}(\boldsymbol{\mu}-\boldsymbol{b})\boldsymbol{\Omega}^{-1}(\boldsymbol{\mu}-\boldsymbol{b})'\right],$$

hence $\boldsymbol{\mu}|\boldsymbol{X},\boldsymbol{\Sigma} \sim N_{m,p}(\boldsymbol{b}, \frac{1}{n_1+n_0}\boldsymbol{\Sigma}\otimes\boldsymbol{\Omega})$, and

$$q(\boldsymbol{\Sigma}|\boldsymbol{X},\boldsymbol{\mu}) \propto |\boldsymbol{\Sigma}|^{\frac{n-m-1-p-n_1p}{2}} etr\left[-\frac{1}{2}\boldsymbol{\Sigma}\boldsymbol{\Phi}^{-1}\right] etr\left[-\frac{1}{2}\boldsymbol{\Sigma}^{-1}(\boldsymbol{A}+\boldsymbol{F})\right], \quad (2.47)$$

with $\boldsymbol{A}$ and $\boldsymbol{F}$ defined in (2.11) and (2.34), respectively. The marginal posterior distribution of $\boldsymbol{\Sigma}$ is obtained by using (2.39) and (2.45),

$$\begin{aligned} q(\boldsymbol{\Sigma}|\boldsymbol{X}) &= \frac{2^{\frac{m(-n+n_1p)}{2}}|\boldsymbol{\Phi}|^{\frac{-n+n_1p}{2}}}{B_{\frac{-n+n_1p}{2}}\left(\frac{1}{4}\boldsymbol{\Phi}^{-1}\boldsymbol{Y}\right)}\int_0^\infty z^{-\frac{m(n-n_1p)}{2}}\delta(z-1)|\boldsymbol{\Sigma}|^{\frac{n-m-1-n_1p}{2}} \\ &\quad\times etr\left[-\frac{1}{2z}\boldsymbol{\Sigma}\boldsymbol{\Phi}^{-1}\right] etr\left[-\frac{z}{2}\boldsymbol{\Sigma}^{-1}\boldsymbol{Y}\right]dz \\ &= \frac{2^{\frac{m(-n+n_1p)}{2}}|\boldsymbol{\Phi}|^{\frac{-n+n_1p}{2}}}{B_{\frac{-n+n_1p}{2}}\left(\frac{1}{4}\boldsymbol{\Phi}^{-1}\boldsymbol{Y}\right)}|\boldsymbol{\Sigma}|^{\frac{n-m-1-n_1p}{2}} etr\left[-\frac{1}{2}\boldsymbol{\Sigma}\boldsymbol{\Phi}^{-1}\right] etr\left[-\frac{1}{2}\boldsymbol{\Sigma}^{-1}\boldsymbol{Y}\right], \end{aligned} \quad (2.48)$$

(see Bekker and Roux (1995)) from R.13 with $\boldsymbol{Y}$ as defined in (2.37). From (2.40), (2.44) and (2.45) the Bayes estimator of $|\boldsymbol{\Sigma}|$ is

$$\widehat{|\boldsymbol{\Sigma}|}_{B,W} = \frac{2^m B_{\frac{-n+n_1p-2}{2}}\left(\frac{1}{4}\boldsymbol{\Phi}^{-1}\boldsymbol{Y}\right)}{B_{\frac{-n+n_1p}{2}}\left(\frac{1}{4}\boldsymbol{\Phi}^{-1}\boldsymbol{Y}\right)}|\boldsymbol{\Phi}|. \quad (2.49)$$

2.4.3.2. Matrix variate t-distribution

Let $\boldsymbol{X}$ follow a matrix variate t distribution with parameters $\boldsymbol{\mu}\in\mathbb{R}^{m\times p}$, $\boldsymbol{\Sigma}\in S_m, \boldsymbol{\Omega}\in S_p$ and ν_0 degrees of freedom. Then from (Chu 1973, Table 1, p.648), the associated weight function is

$$w(z) = \frac{(\frac{\nu_0}{2})^{\frac{\nu_0}{2}} z^{\frac{\nu_0}{2}-1}\exp(\frac{-\nu_0 z}{2})}{\Gamma(\frac{\nu_0}{2})}. \quad (2.50)$$

- As before consider the normal-inverse Wishart prior. Note that from (2.8) and (2.50) follows that

$$q(\boldsymbol{\mu}|\boldsymbol{X},\boldsymbol{\Sigma}) \propto \left(\nu_0 + tr\left[\boldsymbol{\Sigma}^{-1}(\boldsymbol{A}+\boldsymbol{B})\right]\right)^{-\frac{m(m-n-n_1)}{2}-\frac{\nu_0}{2}}, \quad (2.51)$$

and

$$q(\boldsymbol{\Sigma}|\boldsymbol{X},\boldsymbol{\mu}) \propto |\boldsymbol{\Sigma}|^{\frac{1}{2}(n+n_1+1)}\left(\nu_0 + tr\left(\boldsymbol{\Sigma}^{-1}(\boldsymbol{A}+\boldsymbol{B}+\boldsymbol{\Phi})\right)\right)^{-\frac{m(m-n-n_1p-p+1)}{2}-\frac{\nu_0}{2}}, \quad (2.52)$$

with $\boldsymbol{A}$ and $\boldsymbol{B}$ defined in (2.11) and (2.12), respectively. The marginal posterior distribution of $\boldsymbol{\Sigma}$ using (2.22), (2.24) and (2.50) is,

$$\begin{aligned} q(\boldsymbol{\Sigma}|\boldsymbol{X}) &= \frac{2^{-\frac{m(n+n_1p-m-1)}{2}}\left(\frac{\nu_0}{2}\right)^{\frac{\nu_0}{2}}}{\Gamma_m\left(\frac{n+n_1p-m-1}{2}\right)\Gamma\left(\frac{\nu_0}{2}\right)}|\boldsymbol{D}|^{\frac{n+n_1p-m-1}{2}}|\boldsymbol{\Sigma}|^{-\frac{n+n_1p}{2}} \\ &\times \int_0^\infty z^{-\frac{m(m+1-n-n_1p-\nu_0+2)}{2}}\exp\left[-\frac{z}{2}\left(tr\left(\boldsymbol{\Sigma}^{-1}\boldsymbol{D}\right)+\nu_0\right)\right]dz, \end{aligned}$$

with $\boldsymbol{D}$ as defined in (2.17). Note that from R.41,

$$\begin{aligned} q(\boldsymbol{\Sigma}|\boldsymbol{X}) &= \frac{2^{-\frac{m(n+n_1p-m-1)}{2}}\left(\frac{\nu_0}{2}\right)^{\frac{\nu_0}{2}}\Gamma\left(1+\frac{m(n+n_1p+\nu_0-m-1)}{2}\right)}{\Gamma_m\left(\frac{n+n_1p-m-1}{2}\right)\Gamma\left(\frac{\nu_0}{2}\right)} \\ &\times \left(\frac{tr\left(\boldsymbol{\Sigma}^{-1}\boldsymbol{D}\right)+\nu_0}{2}\right)^{-1-\frac{m(n+n_1p+\nu_0-m-1)}{2}}|\boldsymbol{D}|^{\frac{n+n_1p-m-1}{2}}|\boldsymbol{\Sigma}|^{-\frac{n+n_1p}{2}}. \end{aligned} \tag{2.53}$$

From (2.50) and (2.28) the Bayes estimator of $|\boldsymbol{\Sigma}|$ is

$$|\widehat{\boldsymbol{\Sigma}}|_{B,W^{-1}} = \frac{2^{-m}\Gamma_m\left(\frac{n+n_1p-m-3}{2}\right)}{\Gamma_m\left(\frac{n+n_1p-m-1}{2}\right)}|\boldsymbol{W}|\int_0^\infty z^m\frac{\left(\frac{\nu_0}{2}\right)^{\frac{\nu_0}{2}}z^{\frac{\nu_0}{2}-1}\exp\left(-\frac{\nu_0 z}{2}\right)}{\Gamma\left(\frac{\nu_0}{2}\right)}dz. \tag{2.54}$$

Hence from R.41,

$$|\widehat{\boldsymbol{\Sigma}}|_{B,W^{-1}} = \frac{2^{-m-\frac{\nu_0}{2}}\nu_0^{-m}\Gamma_m\left(\frac{n+n_1p-m-3}{2}\right)\Gamma\left(\frac{\nu_0}{2}+m\right)}{\Gamma_m\left(\frac{n+n_1p-m-1}{2}\right)\Gamma\left(\frac{\nu_0}{2}\right)}|\boldsymbol{D}|. \tag{2.55}$$

- Secondly, the normal-Wishart prior is the focus. Note that from (2.31) and (2.50) follows that

$$q(\boldsymbol{\mu}|\boldsymbol{X},\boldsymbol{\Sigma}) \propto \left(\nu_0+tr\left[\boldsymbol{\Sigma}^{-1}\left(\boldsymbol{A}+\boldsymbol{F}\right)\right]\right)^{-\frac{m(m-n-n_1p-p+1)}{2}-\frac{\nu_0}{2}}, \tag{2.56}$$

and

$$q(\boldsymbol{\Sigma}|\boldsymbol{X},\boldsymbol{\mu}) \propto |\boldsymbol{\Sigma}|^{-\frac{n+n_1p+1}{2}}\left(\nu_0+tr\left[\boldsymbol{\Sigma}^{-1}\left(\boldsymbol{A}+\boldsymbol{F}\right)\right]+tr\left(\boldsymbol{\Sigma}\boldsymbol{\Phi}\right)\right)^{-\frac{m(m-n-n_1p-p+1)}{2}-\frac{\nu_0}{2}}, \tag{2.57}$$

with $\boldsymbol{A}$ and $\boldsymbol{F}$ defined in (2.11) and (2.34), respectively. The marginal posterior distribution of $\boldsymbol{\Sigma}$ using (2.39), (2.40) and (2.50) is,

$$\begin{aligned} q(\boldsymbol{\Sigma}|\boldsymbol{X}) &= \frac{2^{\frac{m(-n+n_1p)}{2}}\left(\frac{\nu_0}{2}\right)^{\frac{\nu_0}{2}}}{B_{\frac{-n+n_1p}{2}}\left(\frac{1}{4}\boldsymbol{\Phi}^{-1}\boldsymbol{Y}\right)\Gamma(\frac{\nu_0}{2})}|\boldsymbol{\Phi}|^{\frac{-n+n_1p}{2}}|\boldsymbol{\Sigma}|^{\frac{n-m-1-n_1p}{2}} \\ &\quad\times\int_0^\infty z^{-\frac{m(n-n_1p-\nu_0)+2}{2}}etr\left[-\frac{1}{2z}\boldsymbol{\Sigma}\boldsymbol{\Phi}^{-1}\right]\exp\left[-\frac{z}{2}\left(tr\left[\boldsymbol{\Sigma}^{-1}\boldsymbol{Y}\right]+\nu_0\right)\right]dz, \end{aligned}$$

with $\boldsymbol{Y}$ as defined in (2.37). Note that

$$\begin{aligned} &\int_0^\infty z^{-\frac{m(n-n_1p-\nu_0)+1}{2}}etr\left[-\frac{1}{2z}\boldsymbol{\Sigma}\boldsymbol{\Phi}^{-1}\right]\exp\left[-\frac{z}{2}\left(tr\left[\boldsymbol{\Sigma}^{-1}\boldsymbol{Y}\right]+2\nu_0\right)\right]dz \\ &= 2\left[\frac{tr\left(\boldsymbol{\Sigma}\boldsymbol{\Phi}^{-1}\right)}{(tr\left[\boldsymbol{\Sigma}^{-1}\boldsymbol{Y}\right]+2\nu_0)}\right]^{-\frac{m(n-n_1p-\nu_0)-1}{4}} \\ &\quad\times K_{\frac{m(-n+n_1p+\nu_0)+1}{2}}\left(\sqrt{tr\left(\boldsymbol{\Sigma}\boldsymbol{\Phi}^{-1}\right)\left(tr\left[\boldsymbol{\Sigma}^{-1}\boldsymbol{Y}\right]+2\nu_0\right)}\right), \end{aligned}$$

where $K_\nu(.)$ is the Bessel function of the third kind (see R.14). Hence

$$\begin{aligned} q(\boldsymbol{\Sigma}|\boldsymbol{X}) &= \frac{2^{\frac{m(-n+n_1p)}{2}+1}\left(\frac{\nu_0}{2}\right)^{\frac{\nu_0}{2}}|\boldsymbol{\Sigma}|^{\frac{n-m-1-n_1p}{2}}|\boldsymbol{\Phi}|^{\frac{-n+n_1p}{2}}}{B_{\frac{-n+n_1p}{2}}\left(\frac{1}{4}\boldsymbol{\Phi}^{-1}\boldsymbol{Y}\right)\Gamma(\frac{\nu_0}{2})} \\ &\quad\times\left[\frac{tr\left(\boldsymbol{\Sigma}\boldsymbol{\Phi}^{-1}\right)}{(tr\left[\boldsymbol{\Sigma}^{-1}\boldsymbol{Y}\right]+2\nu_0)}\right]^{-\frac{m(n-n_1p-\nu_0)-1}{4}} \\ &\quad\times K_{-\frac{m(n-n_1p-\nu_0)-1}{2}}\left(\sqrt{tr\left(\boldsymbol{\Sigma}\boldsymbol{\Phi}^{-1}\right)\left(tr\left[\boldsymbol{\Sigma}^{-1}\boldsymbol{Y}\right]+2\nu_0\right)}\right). \end{aligned} \tag{2.58}$$

From (2.44) and (2.50) the Bayes estimator of $|\boldsymbol{\Sigma}|$ is

$$\begin{aligned} |\widehat{\boldsymbol{\Sigma}}|_{B,W} &= \frac{2^m B_{\frac{-n+n_1p-2}{2}}\left(\frac{1}{4}\boldsymbol{\Phi}^{-1}\boldsymbol{Y}\right)}{B_{\frac{-n+n_1p}{2}}\left(\frac{1}{4}\boldsymbol{\Phi}^{-1}\boldsymbol{Y}\right)}|\boldsymbol{\Phi}|\int_0^\infty z^m\frac{\left(\frac{\nu_0}{2}\right)^{\frac{\nu_0}{2}}z^{\frac{\nu_0}{2}-1}\exp(-\nu_0 z)}{\Gamma(\frac{\nu_0}{2})}dz \\ &= \frac{2^{m-\frac{\nu_0}{2}}\Gamma(m+\frac{\nu_0}{2})B_{\frac{-n+n_1p-2}{2}}\left(\frac{1}{4}\boldsymbol{\Phi}^{-1}\boldsymbol{Y}\right)}{\nu_0^m\Gamma(\frac{\nu_0}{2})B_{\frac{-n+n_1p}{2}}\left(\frac{1}{4}\boldsymbol{\Phi}^{-1}\boldsymbol{Y}\right)}|\boldsymbol{\Phi}|, \end{aligned}$$

using R.41.

2.5 Numerical studies

In this section, the proposed results are utilised in such a way as to justify the use of the normal-Wishart prior instead of the normal-inverse Wishart prior for some special cases. Gibbs sampling (see R.62) is used to simulate the posterior samples since the exact expressions for the density functions in (2.48), (2.53) and (2.58) cannot be practically implemented. Simulation studies and application to real datasets are done for the univariate and multivariate cases. For the univariate case, software like OpenBUGS can be used for the simulation of posterior samples with conjugate and non-conjugate priors, but in the multivariate case the software can only accommodate conjugate priors which this study is not restricted to. Algorithms for the simulation of multivariate posterior samples are developed and inspected for convergence. Convergence criteria for multivariate random variates are defined and implemented.

2.5.1 Univariate case

2.5.1.1. Simulation study

OpenBUGS (see code in Appendix B.1) or similar software can be used to apply the results for the univariate case.

2.5.1.1.1 Normal population model

A univariate ($m = 1, p = 1$) normal dataset of size $n_1 = 10$ was simulated with mean $\mu = 0$, and variance $\sigma^2 = 1$. The two priors under consideration are:

- Normal-inverse gamma (see R.39 and R.40) $\mu|\sigma \sim N(0.5, \frac{\sigma^2}{2})$ and $\sigma^2 \sim IG(4,3)$
- Normal-gamma (see R.39 and R.41) $\mu|\sigma \sim N(0.5, \frac{\sigma^2}{2})$ and $\sigma^2 \sim G(4,4)$.

A Gibbs sampler in OpenBUGS is applied that converged graphically (see Gelman and Rubin (1992)) as illustrated in Figures 2.2 and 2.3 and the Bayes estimates and credible intervals are as follows:

Parameter	Estimate	95% credible interval	Length of the credible interval
μ	0.2853	(-0.2547,0.824)	1.0787
σ^2	1.4	(0.6473,2.457)	1.8907

TABLE 2.1: Results for $\sigma^2 \sim IG(4,3)$

Parameter	Estimate	95% credible interval	Length of the credible interval
μ	0.2835	(-0.13,0.6985)	0.8285
σ^2	1.211	(0.5602,2.126)	1.5658

TABLE 2.2: Results for $\sigma^2 \sim G(4,4)$

It is quite clear that the point estimates obtained with the normal-gamma prior is closer to the target values ($\mu = 0, \sigma^2 = 1$) and the 95% credible intervals are narrower.

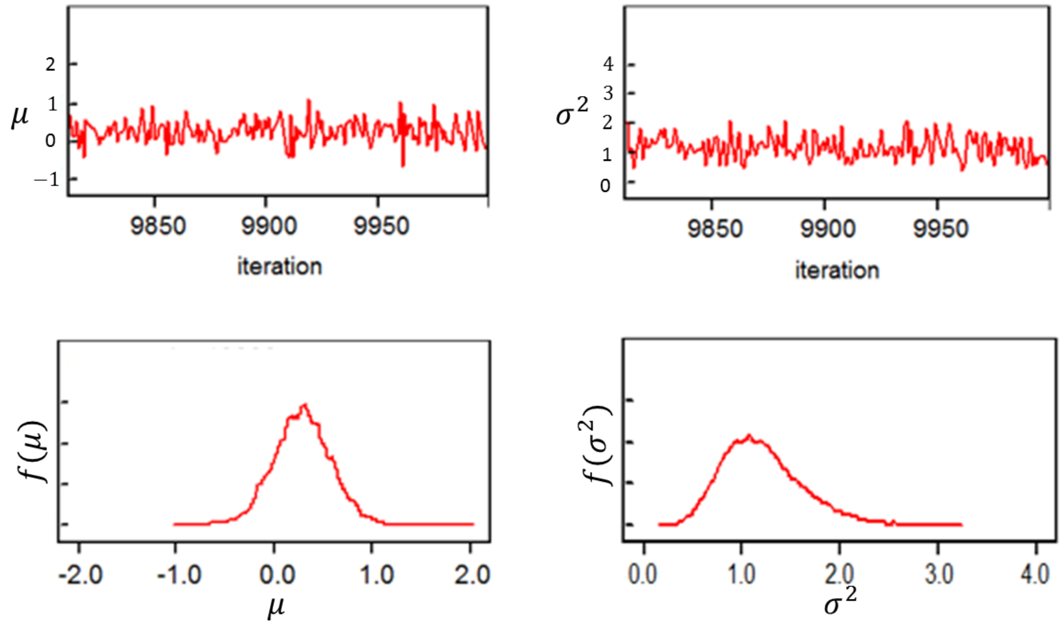


FIGURE 2.2: Chains (top) and posterior densities (bottom) obtained from Gibbs sampling for μ (left) and σ^2 (right) under the normal-inverse gamma prior

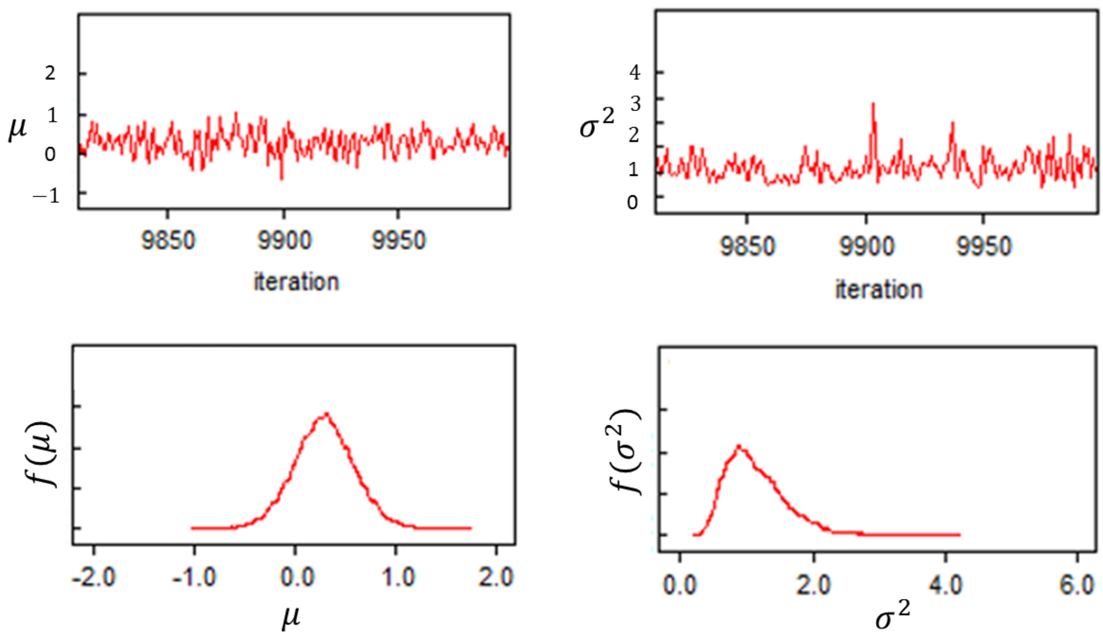


FIGURE 2.3: Chains (top) and posterior densities (bottom) obtained from Gibbs sampling for μ (left) and σ^2 (right) under the normal-gamma prior

2.5.1.1.2 Student's t population model

A univariate ($m = 1, p = 1$) Student's t dataset of size $n_1 = 20$ was simulated with location parameter $\mu = 10$, scale parameter $\sigma^2 = 5$ and degrees of freedom $\nu = 5$. The prior for the degrees of freedom is taken to be a uniform(1,10) distribution. The two priors are:

- Normal-inverse gamma $\mu|\sigma \sim N(12, \frac{\sigma^2}{2})$ and $\sigma^2 \sim IG(4,3)$
- Normal-gamma $\mu|\sigma \sim N(12, \frac{\sigma^2}{2})$ and $\sigma^2 \sim G(4,4)$

A Gibbs sampler that converged graphically as illustrated in Figures 2.4 and 2.5 and the results are as follows:

Parameter	Estimate	95% credible interval	Length of the credible interval
μ	11.13	(10.14,12.11)	1.97
σ^2	5.718	(2.758,11.21)	8.452
ν	5.003	(1.355,9.717)	8.362

TABLE 2.3: Results for $\sigma^2 \sim IG(4,4)$

Parameter	Estimate	95% credible interval	Length of the credible interval
μ	11.12	(10.22,12.02)	1.8
σ^2	5.029	(2.84,8.824)	5.984
ν	5.428	(3.094,9.685)	6.591

TABLE 2.4: Results for $\sigma^2 \sim G(4,3)$

Notice the similar performance as with the normal simulated dataset between the two priors where the credible intervals for the normal-gamma prior is shorter than for the normal-inverse gamma prior.

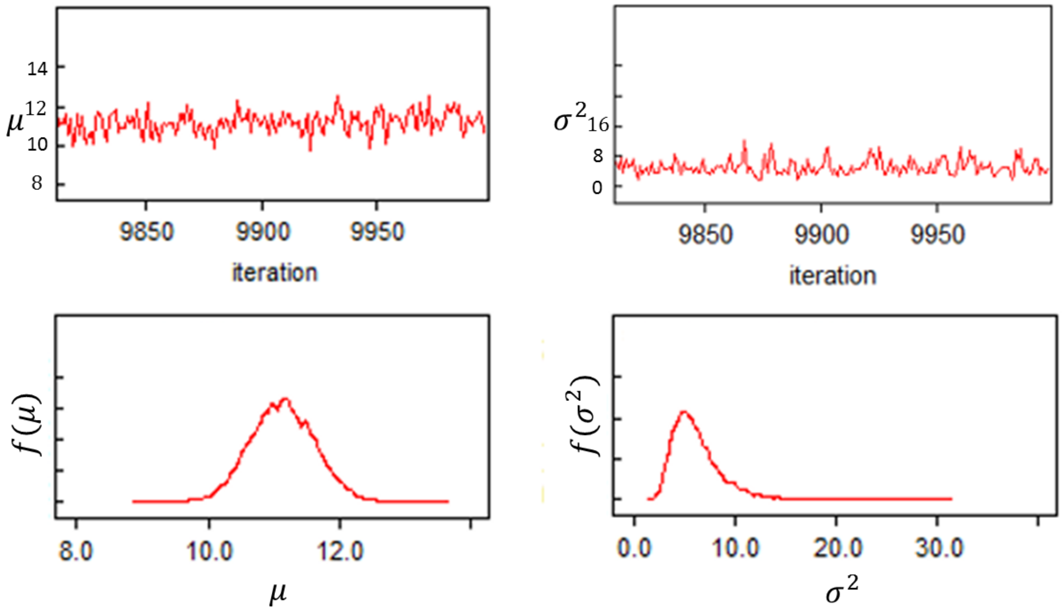


FIGURE 2.4: Chains (top) and posterior densities (bottom) obtained from Gibbs sampling for μ (left) and σ^2 (right) under the normal-inverse gamma prior

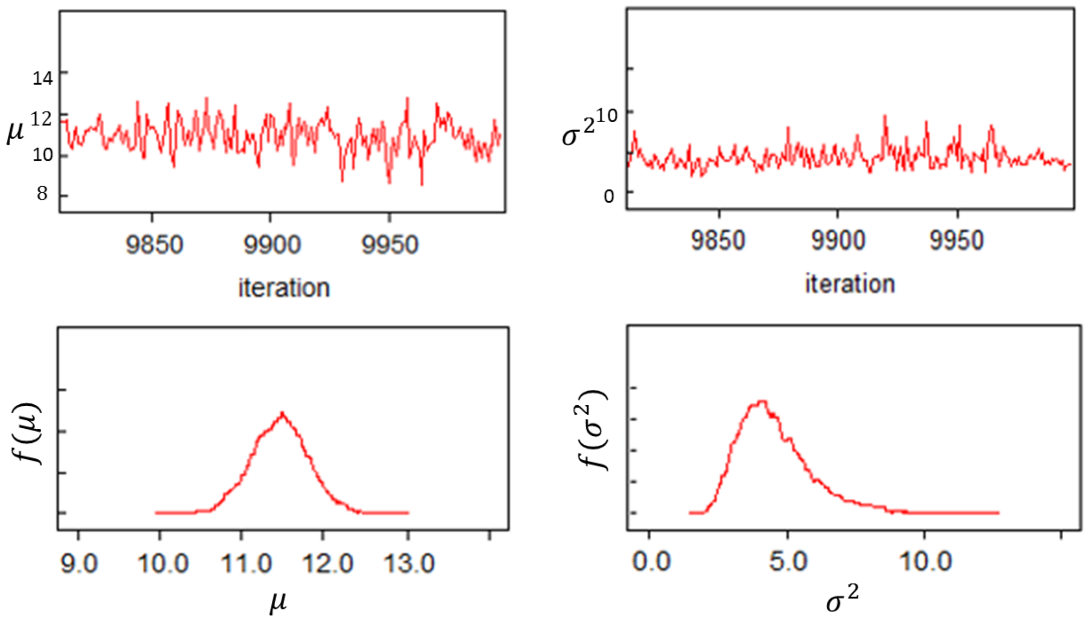


FIGURE 2.5: Chains (top) and posterior densities (bottom) obtained from Gibbs sampling for μ (left) and σ^2 (right) under the normal-gamma prior

2.5.1.2. Kanamycin levels dataset

The dataset presented in Miller Jr (1980), which consists of simultaneous pairs of measurements of serum kanamycin level in blood samples drawn from twenty babies is used. One of the measurements were obtained by a heelstick method (X) and the other by using an umbilical catheter (Y). A simple linear regression model is fit to the data to test whether there is a systematic difference in the two methods. Hence the model is

$$Y_i = \beta_0 + \beta_1 X_i + \varepsilon_i$$

with the assumption $\varepsilon_i \sim t(0, \sigma_\varepsilon^2, v_\varepsilon)$ (see R.42). The prior distributions for β_0 and β_1 are taken to be standard normal distributions (similar to Bolfarine and Arellano-Valle (2005)) and from the distribution on the error it follows that $Y_i \sim t(\beta_0 + \beta_1 X_i, \sigma_\varepsilon^2, v_\varepsilon)$. The prior distribution for the error variance is then either an inverse gamma or gamma distribution and these results are then compared. Firstly the inverse-gamma prior, $\sigma_\varepsilon^2 \sim IG(\alpha_1, \beta_1)$ and secondly the gamma prior, $\sigma_\varepsilon^2 \sim G(\alpha_2, \beta_2)$ are assumed. The hyperparameters are chosen as $\alpha_1 = 4$ and $\beta_1 = 3$ (Bolfarine and Arellano-Valle 2005) and accordingly $\alpha_2 = 4$ and $\beta_2 = 3$. The error degrees of freedom, v_ε, is assumed to be 7 as in Bolfarine and Arellano-Valle (2005). A Gibbs algorithm (see code in Appendix B.2) is used and the convergence is verified graphically. The posterior results are summarized in Tables 2.5 and 2.6.

Parameter	Estimate	95% credible interval	Length of credible interval
β_0	0.7059	(-1.237,2.641)	3.878
β_1	1.761	(1.488,2.04)	0.552
σ_ε^2	22.67	(10.47,43.1)	32.63

TABLE 2.5: Results for $\sigma_\varepsilon^2 \sim IG(4,3)$

Parameter	Estimate	95% credible interval	Length of credible interval
β_0	0.701	(-1.232,2.628)	3.86
β_1	1.762	(1.487,2.04)	0.553
σ_ε^2	23.03	(10.74,43.7)	32.96

TABLE 2.6: Results for $\sigma_\varepsilon^2 \sim G(4,3)$

The length of the intervals are very close for both priors. The interval for β_0 is shorter for the gamma prior than the inverse gamma prior. It is shown in this dataset that the inverse gamma prior is not superior to the gamma prior.
From the simulation and real dataset it follows that a gamma prior might be considered in applications.

2.5.2 Multivariate case

As previously noted, the focus is on the matrix variate normal and t distributions as specific cases of the underlying matrix variate elliptical model, $\boldsymbol{X} \sim E_{m,p}(\boldsymbol{\mu}, \boldsymbol{\Sigma}, \boldsymbol{\Omega}, g'')$ (see R.51). In this section a simulation study is done to obtain the Bayes estimators of $\boldsymbol{\Sigma}$ and to compare these estimators with the maximum likelihood estimators and the Bayes estimators under the inverse Wishart prior. The Frobenius norm (R.17) is used as a comparative measure of the bias of the different estimators. Some other interesting results of the posterior distributions of $\boldsymbol{\Sigma}$ are displayed and discussed.

2.5.2.1. Algorithms

In this section four algorithms are provided for the simulation of matrix variate posterior samples. Gibbs sampling (see R.62) with a Metropolis-Hastings algorithm (see R.63) is used to simulate and the convergency is evaluated.

2.5.2.1.1. Matrix variate normal model

The posterior distributions of $\boldsymbol{\mu}$ and $\boldsymbol{\Sigma}$ with the normal-inverse Wishart prior is then simulated as follows:

Algorithm 1.

1. Initialize $\boldsymbol{\mu}_0$ and $\boldsymbol{\Sigma}_0$
2. Repeat the following steps for $t = 1, ..., 100\,000$ times:
   (a) Generate $\boldsymbol{\mu}_t \sim N_{m,p}(\boldsymbol{\theta}, \frac{1}{n_1+n_0}\boldsymbol{\Sigma}_{t-1} \otimes \boldsymbol{\Omega})$
   (b) Calculate $\boldsymbol{A}_t = \boldsymbol{V} + n_1(\overline{\boldsymbol{X}} - \boldsymbol{\mu}_t)(\overline{\boldsymbol{X}} - \boldsymbol{\mu}_t)'$ and $\boldsymbol{B}_t = n_0(\boldsymbol{\mu}_t - \boldsymbol{\theta})\boldsymbol{\Omega}^{-1}(\boldsymbol{\mu}_t - \boldsymbol{\theta})' + \boldsymbol{\Phi}$
   (c) Generate the random matrix $\boldsymbol{\Sigma}_t \sim W_m^{-1}(\boldsymbol{A}_t + \boldsymbol{B}_t, p + n_1 p + n)$
3. Burn-in the first couple of observations, i.e. the posterior observations are $\boldsymbol{\mu}_{1000}, ..., \boldsymbol{\mu}_{100\,000}$ and $\boldsymbol{\Sigma}_{1000}, ..., \boldsymbol{\Sigma}_{100\,000}$.

The posterior distributions of $\boldsymbol{\mu}$ and $\boldsymbol{\Sigma}$ with the normal-Wishart prior is then simulated by using Gibbs sampling (see R.62) with a Metropolis-Hastings algorithm (see R.63) as follows:

Algorithm 2.

1. Initialize $\boldsymbol{\mu}_0$ and $\boldsymbol{\Sigma}_0$
2. Repeat the following steps for $t = 1, ..., 100\,000$ times:
   (a) Generate $\boldsymbol{\mu}_t \sim N_{m,p}(\boldsymbol{\theta}, \frac{1}{n_1+n_0}\boldsymbol{\Sigma}_{t-1} \otimes \boldsymbol{\Omega})$
   (b) Calculate $\boldsymbol{A}_t = \boldsymbol{V} + n_1(\overline{\boldsymbol{X}} - \boldsymbol{\mu}_t)(\overline{\boldsymbol{X}} - \boldsymbol{\mu}_t)'$ and $\boldsymbol{B}_t = n_0(\boldsymbol{\mu}_t - \boldsymbol{\theta})\boldsymbol{\Omega}^{-1}(\boldsymbol{\mu}_t - \boldsymbol{\theta})' + \boldsymbol{\Phi}$
   (c) Metropolis-Hastings algorithm:
      i. Generate the random matrices $\boldsymbol{\Sigma}_1^* \sim W_m(\boldsymbol{\Phi}_1, m_1)$ and $\boldsymbol{\Sigma}_2^* \sim W_m^{-1}(\boldsymbol{\Phi}_2, m_1^*)$ such that $E[\boldsymbol{\Sigma}_1^*] = cE[\boldsymbol{\Sigma}_2^*]$.
      ii. Calculate $\boldsymbol{\Sigma}^* = w\boldsymbol{\Sigma}_1^* + (1-w)\boldsymbol{\Sigma}_2^*$ for some $0 < w < 1$
      iii. If $min\left(\frac{f^*[\boldsymbol{\Sigma}^*|\boldsymbol{\mu}_t]}{f^*[\boldsymbol{\Sigma}_{t-1}|\boldsymbol{\mu}_t]}, 1\right) > u$ where u is a random uniform$(0,1)$ variate, then $\boldsymbol{\Sigma}_t = \boldsymbol{\Sigma}^*$ else $\boldsymbol{\Sigma}_t = \boldsymbol{\Sigma}_{t-1}$, with $f^*[\boldsymbol{\Sigma}|\boldsymbol{\mu}_t] \propto |\boldsymbol{\Sigma}|^{\frac{n-m-1-p-n_1 p}{2}} etr\left[-\frac{1}{2}\boldsymbol{\Sigma}\boldsymbol{\Phi}^{-1}\right] etr\left[-\frac{1}{2}\boldsymbol{\Sigma}^{-1}(\boldsymbol{A}_t + \boldsymbol{B}_t)\right]$ from (2.47).

3. Burn-in the first couple of observations, i.e. the posterior observations are $\boldsymbol{\mu}_{1000},...,\boldsymbol{\mu}_{100\,000}$ and $\boldsymbol{\Sigma}_{1000},...,\boldsymbol{\Sigma}_{100\,000}$.

Remark 2.5.1. *The value of c in Algorithm 2 will determine the efficiency of the algorithm, if it is chosen to be close to 1, the algorithm will be more efficient.*

2.5.2.1.2. Matrix variate t model

Similar to Algorithms 1 and 2, Algorithms 3 and 4 are proposed for a matrix variate t population.
The posterior distributions of $\boldsymbol{\mu}$ and $\boldsymbol{\Sigma}$ with the normal-inverse Wishart prior is simulated as follows:

Algorithm 3.

1. Initialize $\boldsymbol{\mu}_0$ and $\boldsymbol{\Sigma}_0$
2. Repeat the following steps for $t = 1,...,100\,000$ times:
   (a) Metropolis-Hastings algorithm for $\boldsymbol{\mu}$:
      i. Generate $\boldsymbol{\mu}^* \sim N_{m,p}(\boldsymbol{b}, \frac{1}{n_1+n_0}\boldsymbol{\Sigma}_{t-1} \otimes \boldsymbol{\Omega})$
      ii. Calculate $\boldsymbol{B}_{t-1} = n_0(\boldsymbol{\mu}_{t-1} - \boldsymbol{\theta})\boldsymbol{\Omega}^{-1}(\boldsymbol{\mu}_{t-1} - \boldsymbol{\theta})' + \boldsymbol{\Phi}, \boldsymbol{A}_{t-1} = \boldsymbol{V} + n_1(\overline{\boldsymbol{X}} - \boldsymbol{\mu}_{t-1})(\overline{\boldsymbol{X}} - \boldsymbol{\mu}_{t-1})'$, $\boldsymbol{B}_* = n_0(\boldsymbol{\mu}_t^* - \boldsymbol{\theta})\boldsymbol{\Omega}^{-1}(\boldsymbol{\mu}_t^* - \boldsymbol{\theta})' + \boldsymbol{\Phi}, \boldsymbol{A}_* = \boldsymbol{V} + n_1(\overline{\boldsymbol{X}} - \boldsymbol{\mu}^*)(\overline{\boldsymbol{X}} - \boldsymbol{\mu}^*)'$
      iii. If $min\left(\frac{f_1[\boldsymbol{\mu}^*|\boldsymbol{\Sigma}_t]}{f_1[\boldsymbol{\mu}_{t-1}|\boldsymbol{\Sigma}_t]}, 1\right) > u$ where u is a random uniform$(0,1)$ variate, then $\boldsymbol{\mu}_t = \boldsymbol{\mu}^*$ else $\boldsymbol{\mu}_t = \boldsymbol{\mu}_{t-1}$, with
      $f_1[\boldsymbol{\mu}_{t-1}|\boldsymbol{\Sigma}_t] \propto \left(\nu_0 + tr\left[\boldsymbol{\Sigma}^{-1}(\boldsymbol{A}_{t-1} + \boldsymbol{B}_{t-1})\right]\right)^{-\frac{m(m-n-n_1)}{2} - \frac{\nu_0}{2}}$ and
      $f_1[\boldsymbol{\mu}^*|\boldsymbol{\Sigma}_t] \propto \left(\nu_0 + tr\left[\boldsymbol{\Sigma}^{-1}(\boldsymbol{A}_* + \boldsymbol{B}_*)\right]\right)^{-\frac{m(m-n-p+1-n_1p)}{2} - \frac{\nu_0}{2}}$ from (2.51).
   (b) Calculate $\boldsymbol{A}_t = \boldsymbol{V} + n_1(\overline{\boldsymbol{X}} - \boldsymbol{\mu}_t)(\overline{\boldsymbol{X}} - \boldsymbol{\mu}_t)'$ and $\boldsymbol{B}_t = n_0(\boldsymbol{\mu}_t - \boldsymbol{\theta})\boldsymbol{\Omega}^{-1}(\boldsymbol{\mu}_t - \boldsymbol{\theta})' + \boldsymbol{\Phi}$
   (c) Metropolis-Hastings algorithm for $\boldsymbol{\Sigma}$:
      i. Generate the random matrices $\boldsymbol{\Sigma}_1^* \sim W_m(\boldsymbol{\Phi}_1, m_2)$ and $\boldsymbol{\Sigma}_2^* \sim W_m^{-1}(\boldsymbol{\Phi}_2, m_2^*)$ such that $E[\boldsymbol{\Sigma}_1^*] = cE[\boldsymbol{\Sigma}_2^*]$.
      ii. Calculate $\boldsymbol{\Sigma}^* = w\boldsymbol{\Sigma}_1^* + (1-w)\boldsymbol{\Sigma}_2^*$ for some $0 < w < 1$
      iii. If $min\left(\frac{f_2[\boldsymbol{\Sigma}^*|\boldsymbol{\mu}_t]}{f_2[\boldsymbol{\Sigma}_{t-1}|\boldsymbol{\mu}_t]}, 1\right) > u$ where u is a random uniform$(0,1)$ variate, then $\boldsymbol{\Sigma}_t = \boldsymbol{\Sigma}^*$ else $\boldsymbol{\Sigma}_t = \boldsymbol{\Sigma}_{t-1}$, with
      $f_2[\boldsymbol{\Sigma_{t-1}}|\boldsymbol{\mu}_t] \propto |\boldsymbol{\Sigma}_{t-1}|^{\frac{1}{2}(n+n_1+1)}\left(\nu_0 + tr\left(\boldsymbol{\Sigma}_{t-1}^{-1}(\boldsymbol{A}_t + \boldsymbol{B}_t + \boldsymbol{\Phi})\right)\right)^{-\frac{m(m-n-n_1p-p+1)}{2} - \frac{\nu_0}{2}}$ from (2.52).
3. Burn-in the first couple of observations, i.e. the posterior observations are $\boldsymbol{\mu}_{1000},...,\boldsymbol{\mu}_{100\,000}$ and $\boldsymbol{\Sigma}_{1000},...,\boldsymbol{\Sigma}_{100\,000}$.

The posterior distributions of $\boldsymbol{\mu}$ and $\boldsymbol{\Sigma}$ with the normal-Wishart prior is then simulated by the following algorithm:

Algorithm 4.

1. Initialize $\boldsymbol{\mu}_0$ and $\boldsymbol{\Sigma}_0$

2. Repeat the following steps for $t = 1, ..., 100\,000$ times:

   (a) Metropolis-Hastings algorithm for $\boldsymbol{\mu}$:

   i. Generate $\boldsymbol{\mu}^* \sim N_{m,p}(\boldsymbol{b}, \frac{1}{n_1+n_0}\boldsymbol{\Sigma}_{t-1} \otimes \boldsymbol{\Omega})$

   ii. Calculate $\boldsymbol{F}_{t-1} = n_0(\boldsymbol{\mu}_{t-1} - \boldsymbol{\theta})(\boldsymbol{\mu}_{t-1} - \boldsymbol{\theta})'$, $\boldsymbol{A}_{t-1} = \boldsymbol{V} + n_1(\overline{\boldsymbol{X}} - \boldsymbol{\mu}_{t-1})(\overline{\boldsymbol{X}} - \boldsymbol{\mu}_{t-1})'$, $\boldsymbol{F}_* = n_0(\boldsymbol{\mu}^* - \boldsymbol{\theta})(\boldsymbol{\mu}^* - \boldsymbol{\theta})'$, $\boldsymbol{A}_* = \boldsymbol{V} + n_1(\overline{\boldsymbol{X}} - \boldsymbol{\mu}^*)(\overline{\boldsymbol{X}} - \boldsymbol{\mu}^*)'$

   iii. If $min\left(\frac{f_3[\boldsymbol{\mu}^*|\boldsymbol{\Sigma}_t]}{f_3[\boldsymbol{\mu}_{t-1}|\boldsymbol{\Sigma}_t]}, 1\right) > u$ where u is a random uniform$(0,1)$ variate, then $\boldsymbol{\mu}_t = \boldsymbol{\mu}^*$ else $\boldsymbol{\mu}_t = \boldsymbol{\mu}_{t-1}$, with
   $f_3[\boldsymbol{\mu}_{t-1}|\boldsymbol{\Sigma}_t] \propto \left(v_0 + tr\left[\boldsymbol{\Sigma}^{-1}\left(\boldsymbol{F}_{t-1} + \boldsymbol{A}_{t-1}\right)\right]\right)^{-\frac{m(m-n-n_1p-p+1)}{2} - \frac{v_0}{2}}$
   and $f_3[\boldsymbol{\mu}^*|\boldsymbol{\Sigma}_t] \propto \left(v_0 + tr\left[\boldsymbol{\Sigma}^{-1}\left(\boldsymbol{F}_* + \boldsymbol{A}_*\right)\right]\right)^{-\frac{m(m-n-n_1p-p+1)}{2} - \frac{v_0}{2}}$ from (2.56).

   (b) Calculate $\boldsymbol{F}_t = n_0(\boldsymbol{\mu}_t - \boldsymbol{\theta})(\boldsymbol{\mu}_t - \boldsymbol{\theta})'$ and $\boldsymbol{A}_t = \boldsymbol{V} + n(\overline{\boldsymbol{X}} - \boldsymbol{\mu}_t)(\overline{\boldsymbol{X}} - \boldsymbol{\mu}_t)'$

   (c) Metropolis-Hastings algorithm for $\boldsymbol{\Sigma}$:

   i. Generate the random matrices $\boldsymbol{\Sigma}_1^* \sim W_m(\boldsymbol{\Phi}_1, m_3)$ and $\boldsymbol{\Sigma}_2^* \sim W_m^{-1}(\boldsymbol{\Phi}_2, m_3^*)$ such that $E[\boldsymbol{\Sigma}_1^*] = cE[\boldsymbol{\Sigma}_2^*]$.

   ii. Calculate $\boldsymbol{\Sigma}^* = w\boldsymbol{\Sigma}_1^* + (1-w)\boldsymbol{\Sigma}_2^*$ for some $0 < w < 1$

   iii. If $min\left(\frac{f_4[\boldsymbol{\Sigma}^*|\boldsymbol{\mu}_t]}{f_4[\boldsymbol{\Sigma}_{t-1}|\boldsymbol{\mu}_t]}, 1\right) > u$ where u is a random uniform$(0,1)$ variate, then $\boldsymbol{\Sigma}_t = \boldsymbol{\Sigma}^*$ else $\boldsymbol{\Sigma}_t = \boldsymbol{\Sigma}_{t-1}$, with
   $f_4[\boldsymbol{\Sigma}_{t-1}|\boldsymbol{\mu}_t] \propto |\boldsymbol{\Sigma}_{t-1}|^{-\frac{n+n_1p+1}{2}}\left(v_0 + tr\left[\boldsymbol{\Sigma}_{t-1}^{-1}\left(\boldsymbol{F}_t + \boldsymbol{A}_t\right)\right] + tr\left(\boldsymbol{\Sigma}_{t-1}\boldsymbol{\Phi}\right)\right)^{-\frac{m(m-n-n_1p-p+1)}{2} - \frac{v_0}{2}}$
   from (2.57).

3. Burn-in the first couple of observations, i.e. the posterior observations are $\boldsymbol{\mu}_{1000}, ..., \boldsymbol{\mu}_{100\,000}$ and $\boldsymbol{\Sigma}_{1000}, ..., \boldsymbol{\Sigma}_{100\,000}$.

Similar as before, the value of c in Algorithms 3 and 4 will determine the efficiency of the algorithm.

Remark 2.5.2. *The convergence of Algorithms 2 and 4 is established graphically using the determinant, trace and largest eigenvalue of the simulated matrix and for Algorithm 3 the mean and median of the vector entries are used as depicted in Figures 2.6 and 2.9. There are no formal tests currently available in literature, to the knowledge of the author, to test for convergence of vectors or matrices and hence the proposed graphical measures will be used.*

2.5.2.2. Simulation study

2.5.2.2.1. Multivariate normal model

The sample of size n_1 is simulated from a multivariate normal $(p=1)$ distribution with a dimensionality of m and a zero mean and identity covariance matrix, i.e. $\boldsymbol{X} \sim N_m(\mathbf{0}, \boldsymbol{I})$ (see R.45). The priors are $\boldsymbol{\mu}|\boldsymbol{\Sigma} \sim N_m(\boldsymbol{\theta}, \frac{1}{n_0}\boldsymbol{\Sigma})$ and $\boldsymbol{\Sigma} \sim W_m^{-1}(\boldsymbol{\Phi}, n_{W^{-1}})$ and $\boldsymbol{\Sigma} \sim W_m(\boldsymbol{\Phi}, n_W)$, respectively, with $\boldsymbol{\theta} = 0.5 \times \mathbf{1}_m, \boldsymbol{\Phi} = 4\boldsymbol{I}_m, \boldsymbol{\Phi}_1 = \boldsymbol{\Phi}_2 = \boldsymbol{\Phi}$, $n_1 = 5$, $n_0 = 1$, $m = 3$, $n_W = m_1 = m_1^* = 3$, $n_{W^{-1}} = 7$, $w = 0.5$. The initial values are chosen as $\boldsymbol{\mu}_0 = [0.5 \ \ 0.5 \ \ 0.5]$ and $\boldsymbol{\Sigma}_0 = 1.1\boldsymbol{I}_m$.

Note that the hyperparameters are assumed to be known and the degrees of freedom of the priors are chosen such that the priors have the same first moment. Convergence of the chains originating from Algorithms 1 and 2 (see code in Appendix B.3) is illustrated graphically according to Remark 2.5.2.

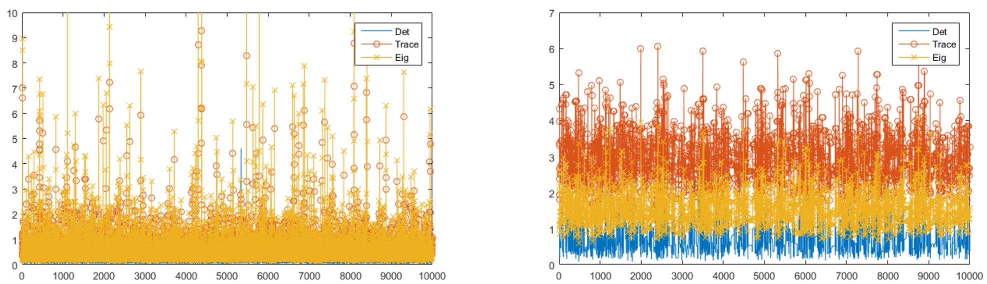


FIGURE 2.6: Convergence measures for the posterior samples of $\boldsymbol{\Sigma}$ simulated from Algorithms 1 (left) and 2 (right)

Remark 2.5.3. *In this chapter two loss functions are used as previously mentioned. The Bayes estimates under the SEL loss function are denoted with a SEL subscript and the Bayes estimates under the loss function defined by Das and Dey (2010) (see R.61) are denoted with a MAP subscript.*

The Bayes estimates under the SEL (PM estimates) of $\boldsymbol{\mu}$ and $\boldsymbol{\Sigma}$, respectively, for the two priors are:

$$\widehat{\boldsymbol{\mu}}_{SEL,W^{-1}} = \begin{bmatrix} -0.42 & -0.33 & -0.26 \end{bmatrix}, \ \widehat{\boldsymbol{\mu}}_{SEL,W} = \begin{bmatrix} -0.14 & -0.033 & -0.26 \end{bmatrix}$$

$$\widehat{\boldsymbol{\Sigma}}_{SEL,W^{-1}} = \begin{bmatrix} 4.2867 & 0.2438 & -0.734 \\ 0.2438 & 1.4037 & 0.9441 \\ -0.734 & 0.9441 & 2.9966 \end{bmatrix}, \ \widehat{\boldsymbol{\Sigma}}_{SEL,W} = \begin{bmatrix} 1.3172 & 0.1827 & -0.3022 \\ 0.1827 & 1.0294 & 0.253 \\ -0.3022 & 0.253 & 1.119 \end{bmatrix}$$

The MAP estimates (see Remark 2.4.2) of $\boldsymbol{\mu}$ and $\boldsymbol{\Sigma}$, respectively, for the two priors are:

$$\widehat{\boldsymbol{\mu}}_{MAP,W^{-1}} = \begin{bmatrix} 0.01 & 0.03 & 0.23 \end{bmatrix}, \ \widehat{\boldsymbol{\mu}}_{MAP,W} = \begin{bmatrix} 0.05 & 0.19 & 0.35 \end{bmatrix}$$

$$\widehat{\boldsymbol{\Sigma}}_{MAP,W^{-1}} = \begin{bmatrix} 1.8314 & 0.3602 & -0.0611 \\ 0.3602 & 0.8022 & 0.0481 \\ -0.0611 & 0.0481 & 0.9872 \end{bmatrix}, \ \widehat{\boldsymbol{\Sigma}}_{MAP,W} = \begin{bmatrix} 0.8379 & 0.1172 & -0.227 \\ 0.1172 & 0.74 & -0.599 \\ -0.227 & -0.599 & 1.1984 \end{bmatrix}$$

The maximum likelihood estimates (MLE's), which are the sample estimates (see Muirhead (1982, Theorem 3.1.5, p.83)) are given by:

$$\widehat{\boldsymbol{\mu}}_{MLE} = \bar{\boldsymbol{x}} = \begin{bmatrix} -0.24 & -0.01 & 0.23 \end{bmatrix}, \widehat{\boldsymbol{\Sigma}}_{MLE} = \begin{bmatrix} 1.5028 & 0.1259 & -0.2181 \\ 0.1259 & 1.0318 & 0.1593 \\ -0.2181 & 0.1593 & 1.0603 \end{bmatrix}$$

Remark 2.5.4. *The Frobenius norm (see R.17) is used as a measure of closeness of the various estimates to the true parameter value and is defined for the errors as follows*

$$||\widehat{\boldsymbol{\Sigma}} - \boldsymbol{\Sigma}||_F = \sqrt{tr\left(\widehat{\boldsymbol{\Sigma}} - \boldsymbol{\Sigma}\right)'\left(\widehat{\boldsymbol{\Sigma}} - \boldsymbol{\Sigma}\right)}$$

It is known that

$$\lim_{r\to\infty} ||\left(\widehat{\boldsymbol{\Sigma}} - \boldsymbol{\Sigma}\right)^r||_F^{\frac{1}{r}} = \rho\left(\widehat{\boldsymbol{\Sigma}} - \boldsymbol{\Sigma}\right)$$

where $\rho\left(\widehat{\boldsymbol{\Sigma}} - \boldsymbol{\Sigma}\right)$ *is the spectral radius.*

The Frobenius norm of the error of the calculated estimates of $\boldsymbol{\Sigma}$ is given in Table 2.7.

Frobenius norm	Value	Frobenius norm	Value
$\|\|\widehat{\boldsymbol{\mu}}_{MLE} - \boldsymbol{\mu}\|\|_F$	0.5956	$\|\|\widehat{\boldsymbol{\Sigma}}_{MLE} - \boldsymbol{\Sigma}\|\|_F$	0.5984
$\|\|\widehat{\boldsymbol{\mu}}_{SEL,W^{-1}} - \boldsymbol{\mu}\|\|_F$	0.5564	$\|\|\widehat{\boldsymbol{\Sigma}}_{SEL,W^{-1}} - \boldsymbol{\Sigma}\|\|_F$	3.6289
$\|\|\widehat{\boldsymbol{\mu}}_{SEL,W} - \boldsymbol{\mu}\|\|_F$	0.5666	$\|\|\widehat{\boldsymbol{\Sigma}}_{SEL,W} - \boldsymbol{\Sigma}\|\|_F$	0.5362
$\|\|\widehat{\boldsymbol{\mu}}_{MAP,W^{-1}} - \boldsymbol{\mu}\|\|_F$	0.9041	$\|\|\widehat{\boldsymbol{\Sigma}}_{MAP,W^{-1}} - \boldsymbol{\Sigma}\|\|_F$	0.947
$\|\|\widehat{\boldsymbol{\mu}}_{MAP,W} - \boldsymbol{\mu}\|\|_F$	0.8407	$\|\|\widehat{\boldsymbol{\Sigma}}_{MAP,W} - \boldsymbol{\Sigma}\|\|_F$	0.6863

TABLE 2.7: Frobenius norm of the errors of the estimates calculated from the simulated sample

In Figures 2.7 and 2.8, respectively, $|| \left(\widehat{\boldsymbol{\mu}} - \boldsymbol{\mu}\right)^r ||_F^{\frac{1}{r}}$ and $|| \left(\widehat{\boldsymbol{\Sigma}} - \boldsymbol{\Sigma}\right)^r ||_F^{\frac{1}{r}}$ are plotted against r.

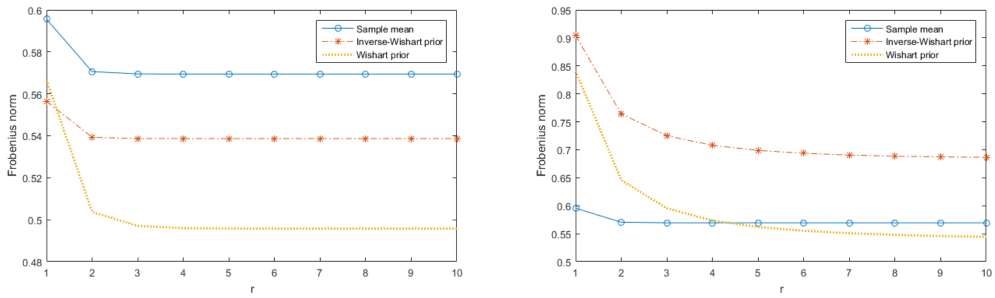


FIGURE 2.7: Frobenius norm for $\widehat{\boldsymbol{\mu}} - \boldsymbol{\mu}$ for the PM (left) and MAP (right) estimators.

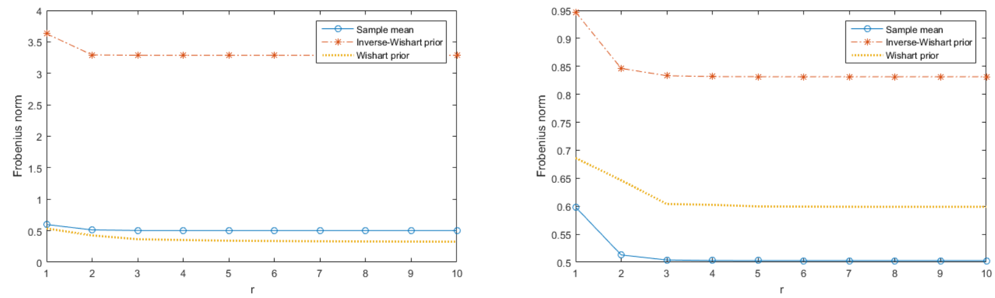


FIGURE 2.8: Frobenius norm for $\widehat{\boldsymbol{\Sigma}} - \boldsymbol{\Sigma}$ for the PM (left) and MAP (right) estimators.

It is clear that the lowest value of the Frobenius norm corresponds to the estimator with the smallest bias. In Figures 2.7 and 2.8 it is evident under the SEL function and the loss function defined by Das and Dey (2010) given in R.61, respectively, that the Bayes estimator from the Wishart prior is superior to the Bayes estimator from the inverse Wishart prior. The estimator under the Wishart prior is competitive when compared to sample (MLE) estimator. Further comparisons of priors are done in Chapter 5.

2.5.2.2.2. Multivariate t model

The sample of size n_1 is simulated from a multivariate t distribution $(p = 1)$ distribution with a dimensionality of m and a zero mean, identity scale matrix and ν_0 degrees of freedom, i.e. $\boldsymbol{X} \sim T_m(\boldsymbol{0}, \boldsymbol{I}, \nu_0)$. The priors are $\boldsymbol{\mu}|\boldsymbol{\Sigma} \sim N_m(\boldsymbol{\theta}, \frac{1}{n_0}\boldsymbol{\Sigma})$ and $\boldsymbol{\Sigma} \sim W_m^{-1}(\boldsymbol{\Phi}, n_{W^{-1}})$ and $\boldsymbol{\Sigma} \sim W_m(\boldsymbol{\Phi}, n_W)$, respectively, with $\boldsymbol{\theta} = 0.5 \times \boldsymbol{1}_m$, $\boldsymbol{\Phi} = 4\boldsymbol{I}_m$, $\boldsymbol{\Phi}_1 = \boldsymbol{\Phi}_2 = \boldsymbol{\Phi}$, $n_1 = 30$, $n_0 = 1$, $m = 3$, $n_{W^{-1}} = m_2 = m_2^* = 7$, $n_W = m_3 = m_3^* = 3$, $\nu_0 = 5$, $w = 0.5$. The initial values are chosen as $\boldsymbol{\mu}_0 = [0.5\ \ 0.5\ \ 0.5]$ and $\boldsymbol{\Sigma}_0 = 1.1\boldsymbol{I}_m$.

Algorithms 3 and 4 (see code in Appendix B.3) are used to simulate posterior samples and convergence of the chains of $\boldsymbol{\mu}$ and $\boldsymbol{\Sigma}$, respectively, are illustrated graphically in Figure 2.9 according to Remark 2.5.2.

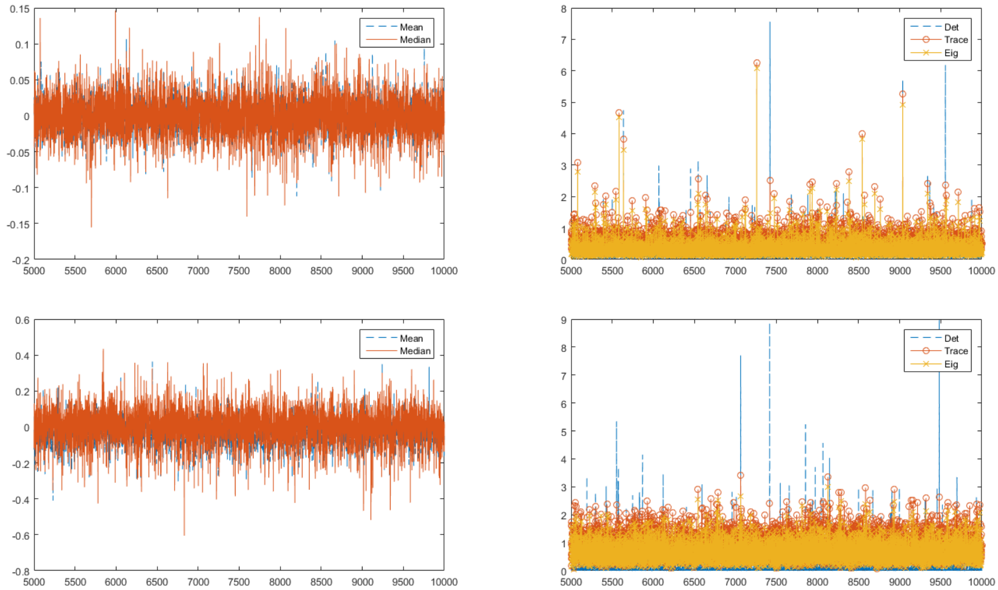


FIGURE 2.9: Convergence measures for the posterior samples of $\boldsymbol{\mu}$ (left) and $\boldsymbol{\Sigma}$ (right) simulated from Algorithms 3 (top) and 4 (bottom)

From Figure 2.9 it is clear that convergence is reached. The Bayes estimates under the SEL (PM estimates) of $\boldsymbol{\mu}$ and $\boldsymbol{\Sigma}$, respectively, for the two priors are:

$$\widehat{\boldsymbol{\mu}}_{SEL,W^{-1}} = \begin{bmatrix} -0.155 & 0.39 & -0.421 \end{bmatrix}, \ \widehat{\boldsymbol{\mu}}_{SEL,W} = \begin{bmatrix} 0.0002 & 0.034 & -0.017 \end{bmatrix}$$

$$\widehat{\boldsymbol{\Sigma}}_{SEL,W^{-1}} = \begin{bmatrix} 1.4208 & 0.0103 & 0.0185 \\ 0.0103 & 1.425 & -0.0036 \\ 0.0185 & -0.0036 & 1.4324 \end{bmatrix}, \ \widehat{\boldsymbol{\Sigma}}_{SEL,W} = \begin{bmatrix} 1.0558 & -0.0002 & -0.0004 \\ -0.0002 & 1.0495 & -0.0022 \\ -0.0004 & -0.0022 & 1.0567 \end{bmatrix}$$

The MAP estimates (see Remark 2.4.2) of $\boldsymbol{\mu}$ and $\boldsymbol{\Sigma}$, respectively, for the two priors are:

$$\widehat{\boldsymbol{\mu}}_{MAP,W^{-1}} = \begin{bmatrix} 0.089 & 0.361 & -0.328 \end{bmatrix}, \ \widehat{\boldsymbol{\mu}}_{MAP,W} = \begin{bmatrix} -0.008 & 0.011 & -0.019 \end{bmatrix}$$

$$\widehat{\boldsymbol{\Sigma}}_{MAP,W^{-1}} = \begin{bmatrix} 0.5877 & 0.5298 & 0.2697 \\ 0.5298 & 1.4511 & -0.486 \\ 0.2697 & -0.486 & 0.7681 \end{bmatrix}, \ \widehat{\boldsymbol{\Sigma}}_{MAP,W} = \begin{bmatrix} 1.3875 & -0.2595 & -0.1943 \\ -0.2595 & 0.8912 & -0.0721 \\ -0.1943 & -0.0721 & 1.1255 \end{bmatrix}$$

The MLE's are given by:

$$\widehat{\boldsymbol{\mu}}_{MLE} = \bar{\boldsymbol{x}} = \begin{bmatrix} 0.0003 & 0.1087 & -0.1516 \end{bmatrix}, \ \widehat{\boldsymbol{\Sigma}}_{MLE} = \begin{bmatrix} 1.6159 & -0.4138 & -0.1025 \\ -0.4138 & 1.6208 & 0.1575 \\ -0.1025 & 0.1575 & 1.2921 \end{bmatrix}$$

In Figures 2.10 and 2.11, respectively, $||\left(\widehat{\boldsymbol{\mu}}-\boldsymbol{\mu}\right)^r||_F^{\frac{1}{r}}$ and $||\left(\widehat{\boldsymbol{\Sigma}}-\boldsymbol{\Sigma}\right)^r||_F^{\frac{1}{r}}$ (see Remark 2.5.4) are plotted against r,

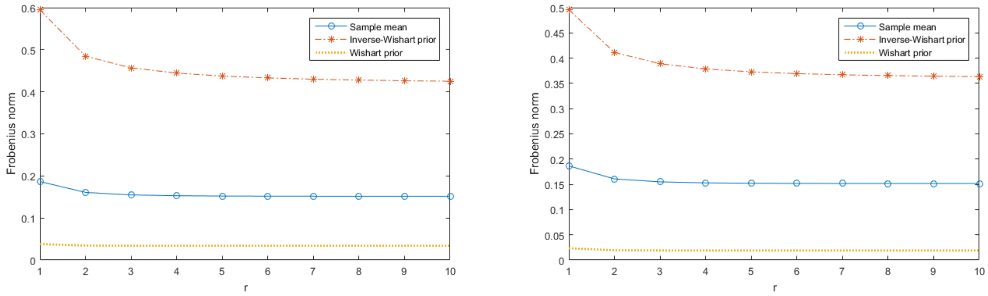


FIGURE 2.10: Frobenius norm for $\widehat{\boldsymbol{\mu}}-\boldsymbol{\mu}$ for the PM (left) and MAP (right) estimators.

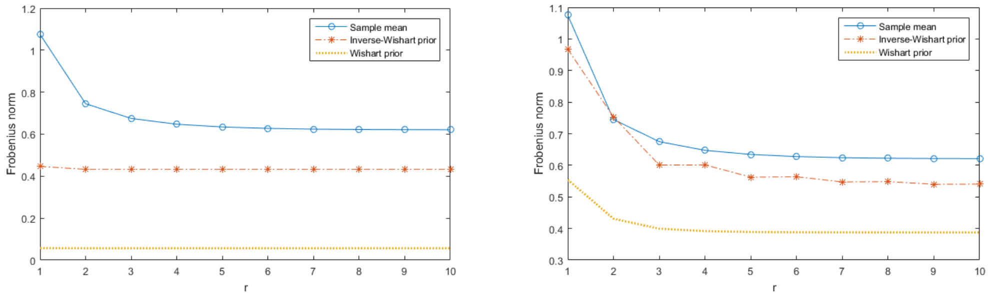


FIGURE 2.11: Frobenius norm for $\widehat{\boldsymbol{\Sigma}}-\boldsymbol{\Sigma}$ for the PM (left) and MAP (right) estimators.

In Figures 2.7 and 2.8 it is evident under the SEL function and the loss function in R.61, respectively, that the Bayes estimator from the Wishart prior is superior to the Bayes estimator from the inverse Wishart prior and also the sample estimator for the underlying multivariate t model.

2.5.2.3. Fisher's Iris dataset

The Iris dataset is a very well-known dataset of dimension four and a sample size of 150. The data set was introduced by Sir Ronald Fisher (1936) as an example of discriminant analysis. The data set consists of 50 samples from each of three species of Iris (Iris setosa, Iris virginica and Iris versicolor) as in Figure 2.12.

FIGURE 2.12: Iris versicolor, virginica and setosa species

Four features were measured from each sample: the length and the width of the sepals and petals, in centimeters. For the purpose of applying the results to this dataset only a subset of the data and the SEL function is used - a subset of the *Iris setosa* subspecies is used, and hence the sample size is $n_1 = 10$. The reason for this is that each of the subspecies' measures have been shown to follow a four dimensional multivariate normal distribution.

Let $\boldsymbol{X}_{4\times 10}$ = sample of *Iris setosa* measurements $\sim N_4(\boldsymbol{\mu}, \boldsymbol{\Sigma})$ with $\boldsymbol{\mu}$ and $\boldsymbol{\Sigma}$ unknown. Assume the prior distributions are $\boldsymbol{\mu}|\boldsymbol{\Sigma} \sim N_4(\boldsymbol{\theta}, \boldsymbol{\Sigma})$ and $\boldsymbol{\Sigma} \sim W_4^{-1}(\boldsymbol{\Phi}, 9)$ and $\boldsymbol{\Sigma} \sim W_4(\boldsymbol{\Phi}, 5)$, respectively, and as illustration $\boldsymbol{\theta} = \overline{\boldsymbol{X}}$ and $\boldsymbol{\Phi} = \frac{1}{10}\sum_{i=1}^{10}(\boldsymbol{X}_i - \overline{\boldsymbol{X}})(\boldsymbol{X}_i - \overline{\boldsymbol{X}})'$.

The posterior distributions are simulated by using Algorithms 1 and 2 (see code in Appendix B.4). The Bayes estimates are calculated as

$$\widehat{\boldsymbol{\mu}}_{B,W^{-1}} = \begin{bmatrix} 4.8950 & 3.2412 & 1.4243 & 0.2329 \end{bmatrix}, \ \widehat{\boldsymbol{\mu}}_{B,W} = \begin{bmatrix} 4.8448 & 3.2401 & 1.4294 & 0.2339 \end{bmatrix}$$

$$\widehat{\boldsymbol{\Sigma}}_{B,W^{-1}} = \begin{bmatrix} 0.0472 & 0.0468 & 0.0023 & -0.0001 \\ 0.0468 & 0.0505 & 0.0015 & -0.0001 \\ 0.0023 & 0.0015 & 0.0006 & 0.0007 \\ -0.0001 & -0.0001 & 0.0007 & 0.0003 \end{bmatrix}$$

$$\widehat{\boldsymbol{\Sigma}}_{B,W} = \begin{bmatrix} 0.1006 & 0.0591 & 0.0290 & 0.0144 \\ 0.0591 & 0.0569 & 0.0188 & 0.0210 \\ 0.0290 & 0.0188 & 0.0174 & 0.0071 \\ 0.0144 & 0.0210 & 0.0071 & 0.0097 \end{bmatrix}$$

The MLE's are given by

$$\widehat{\boldsymbol{\mu}}_{MLE} = \begin{bmatrix} 4.8600 & 3.3100 & 1.4500 & 0.2200 \end{bmatrix}, \ \widehat{\boldsymbol{\Sigma}}_{MLE} = \begin{bmatrix} 0.0849 & 0.0704 & 0.0189 & 0.0087 \\ 0.0704 & 0.0943 & 0.0172 & 0.0164 \\ 0.0189 & 0.0172 & 0.0117 & 0.0044 \\ 0.0087 & 0.0164 & 0.0044 & 0.0062 \end{bmatrix}$$

2.6 Conclusion

Initiated by Bekker and Roux (1995), the question arises whether one should sacrifice complexity of priors for computational efficiency in matrix variate Bayesian analysis. This serves as the motivation for generalizing the use of a complicated prior structure to a broad class of distributions, the matrix variate elliptical model. In this respect, some theoretical and computational achievements are reached.
The main contributions in this chapter are summarized as follows:

Theoretical contributions

- New estimators for the matrix parameters of the matrix variate elliptical model were proposed from a subjective Bayesian viewpoint.
- The normal-inverse Wishart and normal-Wishart priors were considered for the location and scale matrices of the underlying model.

- The Bayes estimators of the parameters, under two loss functions, as well as the joint posterior density functions and marginal posterior density functions, were derived with the matrix variate normal and matrix variate t distribution as particular subfamilies.
- The Bayes estimator under SEL of the location parameter is shown to be a robust estimator in the sense that it is independent of the prior distribution of the scale parameter.

Computational contributions

- Numerical algorithms were developed for the simulation of the posterior samples for the multivariate normal and t distributions, respectively. Some measures of convergence of the algorithms and performance of the estimators were proposed and illustrated.
- For both priors, the posterior distributions for the location and scale matrices of the multivariate normal and t models were simulated, using the four newly developed algorithms. The Bayes estimators under the two loss functions were calculated and compared using the Frobenius norm.
- The estimator derived under the normal-Wishart prior displayed superior performance in the simulation study and this justifies the use of a normal-Wishart prior in the Bayesian analysis of the matrix variate normal and t models.

Chapter 3

Wishart and inverse-Wishart distributions: Generator-types

Motivated from the functional form of the density function of the Wishart distribution, the question arises as to why the functional form is not replaced by a more general functional form that gives the Wishart density function as a special case; hence the hypergeometric Wishart generator distribution is defined and some properties are derived. Special cases of this generator-type distribution, some of which are already available in literature, are mentioned. The significance of this generator distribution is further demonstrated by assuming a special case as a prior in the subjective Bayesian analysis of the matrix variate normal model. From this chapter, Van Niekerk, Bekker, and Arashi (2016) and Bekker, Van Niekerk, and Arashi (2017) resulted. The first paper focuses on the univariate case and the second paper focuses on the matrix variate case.

3.1 Introduction

The Wishart distribution (see Chapter 2, Definition 2.1.1) 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. In this chapter, the hypergeometric Wishart generator distribution is proposed, based on the combination of a generalised matrix variate hypergeometric distribution (Van der Merwe and Roux 1974) and the shape generator $h(.)$. Note this form contains the cases presented by Caro-Lopera, Gonzàlez-Farìas, and Balakrishnan (2014) and Dìaz-Garcìa, Gutièrrez-Jàimez, et al. (2011).

3.2 Construction methodology

In this section the construction methodology of the hypergeometric Wishart generator distribution is given. For $\boldsymbol{X} \in S_m$, Van der Merwe and Roux (1974) defined the hypergeometric function type II distribution with density function

$$f(\boldsymbol{X}) \propto |\boldsymbol{X}|^{\frac{n}{2}-\frac{m+1}{2}} \, {}_2F_1(a_1,a_2;b;\boldsymbol{\Omega X}).$$

where $n \geq m, \boldsymbol{\Omega} \in S_m, ||\boldsymbol{\Omega X}|| < 1$. Roux (1971) defined the generalised hypergeometric function type I distribution for $\boldsymbol{X} \in S_m$ with density function

$$f(\boldsymbol{X}) \propto |\boldsymbol{X}|^{\frac{n}{2}-\frac{m+1}{2}} etr(-\boldsymbol{X}) \, {}_sF_t(\boldsymbol{a};\boldsymbol{b};\boldsymbol{\Omega X}), \tag{3.1}$$

where ${}_sF_t(.;.;.)$ is the hypergeometric function with matrix argument (see R.34),$\boldsymbol{a} = (a_1,\ldots,a_s)$,$\boldsymbol{b} = (b_1,\ldots,b_t), n \geq m, \boldsymbol{\Omega} \in S_m$ and the parameters are resticted to take those values for which the density function is non-negative. For $\boldsymbol{\Omega} = \boldsymbol{0}$, (3.1) reduces to the Wishart density function (Definition 2.1.1). When $s = 0, t = 1$ and $b_1 = \frac{n}{2}$ the noncentral Wishart distribution is obtained with noncentrality parameter $\boldsymbol{\Theta}$ (see Gupta and Nagar (2000, Definition 3.5.1, p.113).

The hypergeometric Wishart generator distribution follows from (3.1) by replacing the $exp(tr(.))$ term with a Borel measurable function $h(.)$ of trace operator as defined as follows.

Definition 3.2.1. *A random matrix $\boldsymbol{X} \in S_m$ is said to have the hypergeometric Wishart generator distribution (HWGD) with parameters $a_1,\ldots,a_s \in \mathbb{C}$, $b_1,\ldots,b_t \in \mathbb{C}$, $(s \leq t)$, $\boldsymbol{\Omega},\boldsymbol{\Sigma} \in S_m$, $n \geq m$ and shape generator $h(\cdot), h(\cdot) \neq 1$, if the density function is given by*

$$f(\boldsymbol{X}) = c_{n,m}^{HWG} |\boldsymbol{\Sigma}|^{-\frac{n}{2}} |\boldsymbol{X}|^{\frac{n}{2}-\frac{m+1}{2}} \, {}_sF_t(\boldsymbol{a};\boldsymbol{b};\boldsymbol{\Omega X}) \, h\left(tr\left(\boldsymbol{\Sigma}^{-1}\boldsymbol{X}\right)\right), \tag{3.2}$$

denoted as $\boldsymbol{X} \sim HWG_m(\boldsymbol{\Sigma},\boldsymbol{\Omega},\boldsymbol{a},\boldsymbol{b},n,h)$, where ${}_sF_t(.;.;.)$ is the hypergeometric function of matrix argument as in R.34. The parameters are restricted to take those values for which the density function is non-negative.

From R.24,

$$\begin{aligned}
\{c_{n,m}^{HWG}\}^{-1} &= |\boldsymbol{\Sigma}|^{-\frac{n}{2}} \int_{S_m} |\boldsymbol{X}|^{\frac{n}{2}-\frac{m+1}{2}} {}_sF_t\left(\boldsymbol{a};\boldsymbol{b};\boldsymbol{\Omega X}\right) h\left(tr\left(\boldsymbol{\Sigma}^{-1}\boldsymbol{X}\right)\right) d\boldsymbol{X} \\
&= |\boldsymbol{\Sigma}|^{-\frac{n}{2}} \sum_{k=0}^{\infty}\sum_{\kappa} \frac{(a_1)_\kappa \ldots (a_s)_\kappa}{(b_1)_\kappa \ldots (b_t)_\kappa} \frac{1}{k!} \\
&\quad \times \int_{S_m} |\boldsymbol{X}|^{\frac{n}{2}-\frac{m+1}{2}} h\left(tr\left(\boldsymbol{\Sigma}^{-1}\boldsymbol{X}\right)\right) C_\kappa(\boldsymbol{\Omega X}) d\boldsymbol{X} \\
&= \Gamma_m\left(\frac{n}{2}\right) \sum_{k=0}^{\infty}\sum_{\kappa} \frac{(a_1)_\kappa \ldots (a_s)_\kappa}{(b_1)_\kappa \ldots (b_t)_\kappa} \frac{\left(\frac{n}{2}\right)_\kappa \gamma_k\left(\frac{n}{2}\right)}{k!\Gamma\left(\frac{nm}{2}+k\right)} C_\kappa(\boldsymbol{\Omega\Sigma}), \qquad (3.3)
\end{aligned}$$

where $C_\kappa(.)$ is the zonal polynomial corresponding to the partition κ as in R.19, $(a_i)_\kappa$ is the generalised Pochammer symbol as in R.10 and $\gamma_k(a) = \int_{\mathbb{R}^+} y^{am+k-1} h(y)dy$. The shape generator $h(.)$ should satisfy the following regularity conditions:

- *$h \neq 1$*
- *$h(.)$ is a Borel measurable function*
- *$h(.)$ admits Taylor's series expansion.*

Definition 3.2.2. *A random matrix $\boldsymbol{X} = \boldsymbol{Y}^{-1} \in S_m$, with $\boldsymbol{Y} \sim HWG_m(\boldsymbol{\Sigma},\boldsymbol{\Omega},\boldsymbol{a},\boldsymbol{b},n,h)$ is said to have the inverse hypergeometric Wishart generator distribution (IHWGD) with parameters $a_1,\ldots,a_s \in \mathbb{C}$, $b_1,\ldots,b_t \in \mathbb{C}$, $(s \leq t)$, $\boldsymbol{\Omega},\boldsymbol{\Sigma} \in S_m$, $n \geq m$ and shape generator $h(\cdot), h(\cdot) \neq 1$, if the density function is given by*

$$f(\boldsymbol{X}) = c_{n,m}^{IHWG} |\boldsymbol{\Sigma}|^{-\frac{n}{2}} |\boldsymbol{X}|^{-\frac{n}{2}-\frac{m+1}{2}} {}_sF_t\left(\boldsymbol{a};\boldsymbol{b};\boldsymbol{\Omega X}^{-1}\right) h\left(tr\left(\boldsymbol{\Sigma}^{-1}\boldsymbol{X}^{-1}\right)\right), \qquad (3.4)$$

denote as $\boldsymbol{X} \sim IHWG_m(\boldsymbol{\Sigma},\boldsymbol{\Omega},\boldsymbol{a},\boldsymbol{b},n,h)$. From R.24 follows that

$$\{c_{n,m}^{IHWG}\}^{-1} = \Gamma_m\left(\frac{n}{2}\right) \sum_{k=0}^{\infty}\sum_{\kappa} \frac{(a_1)_\kappa \ldots (a_s)_\kappa}{(b_1)_\kappa \ldots (b_t)_\kappa} \frac{\left(\frac{n}{2}\right)_\kappa \gamma_k\left(\frac{n}{2}\right)}{k!\Gamma\left(\frac{nm}{2}+k\right)} C_\kappa(\boldsymbol{\Omega\Sigma}),$$

by making the transformation $\boldsymbol{Y}^{-1} = \boldsymbol{X}$ with Jacobian $J(\boldsymbol{Y} \to \boldsymbol{X}) = |\boldsymbol{X}|^{-(m+1)}$ from R.4. The parameters are restricted to take those values for which the density function is non-negative.

3.3 Properties

In this section some properties of the HGWD are derived and illustrated.

Theorem 3.3.1. *For $\boldsymbol{X} \sim HWG_m(\boldsymbol{\Sigma},\boldsymbol{\Omega},\boldsymbol{a},\boldsymbol{b},n,h)$, the r-th moment of $|\boldsymbol{X}|$ is equal to*

$$E\left(|\boldsymbol{X}|^r\right) = \frac{c_{n,m}^{HWG}}{c_{n+2r,m}^{HWG}} |\boldsymbol{\Sigma}|^r,$$

where $c_{n,m}^{HWG}$ and $c_{n+2r,m}^{HWG}$ as in (3.3).

Proof. From (3.2) and (3.3) follows

$$\begin{aligned} E\left(|\boldsymbol{X}|^{r}\right) &= c_{n,m}^{HWG}|\boldsymbol{\Sigma}|^{-\frac{n}{2}}\int_{S_m}|\boldsymbol{X}|^{\frac{n}{2}+r-\frac{m+1}{2}}\,{}_sF_t\left(\boldsymbol{a};\boldsymbol{b};\boldsymbol{\Omega X}\right)h\left(tr\left(\boldsymbol{\Sigma}^{-1}\boldsymbol{X}\right)\right)d\boldsymbol{X} \\ &= c_{n,m}^{HWG}|\boldsymbol{\Sigma}|^{-\frac{n}{2}}(c_{n+2r,m}^{HWG})^{-1}|\boldsymbol{\Sigma}|^{\frac{n}{2}+r} \end{aligned}$$

with $c_{n,m}^{HWG}$ as in (3.3). □

Theorem 3.3.2. *Suppose that $\boldsymbol{X}\sim HWG_m(\boldsymbol{\Sigma},\boldsymbol{\Omega},\boldsymbol{a},\boldsymbol{b},n,h)$. The characteristic function of $\boldsymbol{X}$ is given by*

$$\begin{aligned} \psi_{\boldsymbol{X}}(\boldsymbol{T}) &= c_{n,m}^{HWG}|\boldsymbol{\Sigma}|^{-\frac{n}{2}}\sum_{\phi}\frac{\theta_{\phi}^{\kappa,\rho}(a_1)_{\rho}\ldots(a_s)_{\rho}h^{(k)}(0)}{k!r!(b_1)_{\rho}\ldots(b_t)_{\rho}} \\ &\quad\times\Gamma_m\left(\frac{n}{2},\phi\right)i^{k+r-\frac{mn}{2}}|\boldsymbol{T}|^{-\frac{n}{2}}C_{\phi}^{\kappa,\rho}\left(\boldsymbol{\Sigma}^{-1}\boldsymbol{T},\boldsymbol{\Omega T}\right), \end{aligned}$$

where

$$\sum_{\phi}=\sum_{k=0}^{\infty}\sum_{\kappa}\sum_{r=0}^{\infty}\sum_{\rho}\sum_{\phi\in\kappa,\rho}, \tag{3.5}$$

$c_{n,m}^{HWG}$ as in (3.3), $C_{\phi}^{\kappa,\rho}(.,.)$ the invariant polynomial given in R.25 and $\theta_{\phi}^{\kappa,\rho}=\frac{C_{\phi}^{\kappa,\rho}(\boldsymbol{I}_m,\boldsymbol{I}_m)}{C_{\phi}(\boldsymbol{I}_m)}$.

Proof. From (3.2) follows

$$\begin{aligned} \psi_{\boldsymbol{X}}(\boldsymbol{T}) &= E\left(etr\left(i\boldsymbol{TX}\right)\right) \\ &= c_{n,m}^{HWG}|\boldsymbol{\Sigma}|^{-\frac{n}{2}}\int_{S_m}|\boldsymbol{X}|^{\frac{n}{2}-\frac{m+1}{2}}\,{}_sF_t\left(\boldsymbol{a};\boldsymbol{b};\boldsymbol{\Omega X}\right)h\left(tr\left(\boldsymbol{\Sigma}^{-1}\boldsymbol{X}\right)\right)etr\left(i\boldsymbol{TX}\right)d\boldsymbol{X}. \end{aligned} \tag{3.6}$$

Using the Taylor series expansion (see R.7) of $h(.)$, R.19 and R.34 we have

$$\begin{aligned} &{}_sF_t\left(\boldsymbol{a};\boldsymbol{b};\boldsymbol{\Omega X}\right)h\left(tr\left(\boldsymbol{\Sigma}^{-1}\boldsymbol{X}\right)\right) \\ &= \sum_{k=0}^{\infty}\sum_{\kappa}\sum_{r=0}^{\infty}\sum_{\rho}\frac{(a_1)_{\rho}\ldots(a_s)_{\rho}h^{(k)}(0)}{k!r!(b_1)_{\rho}\ldots(b_t)_{\rho}}C_{\kappa}\left(\boldsymbol{\Sigma}^{-1}\boldsymbol{X}\right)C_{\rho}\left(\boldsymbol{\Omega X}\right). \end{aligned}$$

Hence from (3.6),

$$\begin{aligned} &\int_{S_m}|\boldsymbol{X}|^{\frac{n}{2}-\frac{m+1}{2}}\,{}_sF_t\left(\boldsymbol{a};\boldsymbol{b};\boldsymbol{\Omega X}\right)h\left(tr\left(\boldsymbol{\Sigma}^{-1}\boldsymbol{X}\right)\right)etr\left(i\boldsymbol{TX}\right)d\boldsymbol{X} \\ &= \sum_{k=0}^{\infty}\sum_{\kappa}\sum_{r=0}^{\infty}\sum_{\rho}\frac{(a_1)_{\rho}\ldots(a_s)_{\rho}h^{(k)}(0)}{k!r!(b_1)_{\rho}\ldots(b_t)_{\rho}}\int_{S_m}|\boldsymbol{X}|^{\frac{n}{2}-\frac{m+1}{2}}C_{\kappa}\left(\boldsymbol{\Sigma}^{-1}\boldsymbol{X}\right)C_{\rho}\left(\boldsymbol{\Omega X}\right)etr\left(i\boldsymbol{TX}\right)d\boldsymbol{X} \\ &= \sum_{\phi}\frac{\theta_{\phi}^{\kappa,\rho}(a_1)_{\rho}\ldots(a_s)_{\rho}h^{(k)}(0)}{k!r!(b_1)_{\rho}\ldots(b_t)_{\rho}}\int_{S_m}|\boldsymbol{X}|^{\frac{n}{2}-\frac{m+1}{2}}C_{\phi}^{\kappa,\rho}\left(\boldsymbol{\Sigma}^{-1}\boldsymbol{X},\boldsymbol{\Omega X}\right)etr\left(i\boldsymbol{TX}\right)d\boldsymbol{X} \\ &= \sum_{\phi}\frac{\theta_{\phi}^{\kappa,\rho}(a_1)_{\rho}\ldots(a_s)_{\rho}h^{(k)}(0)}{k!r!(b_1)_{\rho}\ldots(b_t)_{\rho}}\Gamma_m\left(\frac{n}{2},\phi\right)|i\boldsymbol{T}|^{-\frac{n}{2}}C_{\phi}^{\kappa,\rho}\left(i\boldsymbol{\Sigma}^{-1}\boldsymbol{T},i\boldsymbol{\Omega T}\right), \end{aligned}$$

using R.27 and R.28. From R.22, the result follows. □

Theorem 3.3.3. *Let* $\boldsymbol{X} \sim HWG_m(\boldsymbol{\Sigma}, \boldsymbol{\Omega}, \boldsymbol{a}, \boldsymbol{b}, n, h)$*, and* $\boldsymbol{A} \in S_m$*. Then* $\boldsymbol{AXA}' \sim HWG_m\left(\boldsymbol{A\Sigma A}', \boldsymbol{A}'^{-1}\boldsymbol{\Omega A}^{-1}, \boldsymbol{a}, \boldsymbol{b}, n, h\right)$*.*

Proof. Note that the Jacobian of the transformation $\boldsymbol{Y} = \boldsymbol{AXA}'$ is given by $J(\boldsymbol{X} \to \boldsymbol{Y}) = |\boldsymbol{A}|^{-(m+1)}$ from R.3, hence from R.1 the density function of $\boldsymbol{Y} = \boldsymbol{AXA}'$ is given by

$$\begin{aligned}
f(\boldsymbol{Y}) &= f(\boldsymbol{A}^{-1}\boldsymbol{YA}'^{-1})J(\boldsymbol{X} \to \boldsymbol{Y}) \\
&= (c_{n,m}^{HWG})^{-1}|\boldsymbol{\Sigma}|^{-\frac{n}{2}}|\boldsymbol{A}^{-1}\boldsymbol{YA}'^{-1}|^{\frac{n}{2}-\frac{m+1}{2}} \ {}_sF_t\left(\boldsymbol{a};\boldsymbol{b};\boldsymbol{\Omega A}^{-1}\boldsymbol{YA}'^{-1}\right) \\
&\quad \times h\left(tr\left(\boldsymbol{\Sigma}^{-1}\boldsymbol{A}^{-1}\boldsymbol{YA}'^{-1}\right)\right)|\boldsymbol{A}|^{-(m+1)} \\
&= (c_{n,m}^{HWG})^{-1}|\boldsymbol{A\Sigma A}'|^{-\frac{n}{2}}|\boldsymbol{Y}|^{\frac{n}{2}-\frac{m+1}{2}} \ {}_sF_t\left(\boldsymbol{a};\boldsymbol{b};\boldsymbol{\Omega A}^{-1}\boldsymbol{YA}'^{-1}\right) \\
&\quad \times h\left(tr\left(\left(\boldsymbol{A\Sigma A}'\right)^{-1}\boldsymbol{Y}\right)\right),
\end{aligned}$$

where $c_{n,m}^{HWG}$ as in (3.3). □

Remark 3.3.1. *From Theorem 3.3.3, it can be then concluded that if* $\boldsymbol{X} \sim HWG_m(\boldsymbol{\Sigma}, \boldsymbol{\Omega}, \boldsymbol{a}, \boldsymbol{b}, n, h)$ *with* $\boldsymbol{\Sigma} = \boldsymbol{A}^{-1}\boldsymbol{A}'^{-1}$ *and* $\boldsymbol{\Omega} = \boldsymbol{A}'\boldsymbol{A}$*, then* $\boldsymbol{AXA}' \sim HWG_m(\boldsymbol{I}, \boldsymbol{I}, \boldsymbol{a}, \boldsymbol{b}, n, h)$*.*

Theorem 3.3.4. *Let* $\boldsymbol{X} \sim HWG_m(\boldsymbol{\Sigma}, \boldsymbol{\Omega}, \boldsymbol{a}, \boldsymbol{b}, n, h)$*. The joint density function of the eigenvalues* $\boldsymbol{\Lambda} = diag(\lambda_1, ..., \lambda_m)$*,* $\lambda_1 \geq ... \geq \lambda_m > 0$ *of* $\boldsymbol{X}$ *is given by*

$$\begin{aligned}
g'(\boldsymbol{\Lambda}) &= \frac{c_{n,m}^{HWG}\pi^{\frac{m^2}{2}}}{\Gamma_m\left(\frac{m}{2}\right)}\prod_{i<j}^{m}\left(\lambda_i - \lambda_j\right)|\boldsymbol{\Sigma}|^{-\frac{n}{2}}\prod_{i=1}^{m}\lambda_i^{\frac{n}{2}-\frac{m+1}{2}} \\
&\quad \times \sum_{\phi}\theta_{\phi}^{\kappa,\rho}\frac{(a_1)_\rho \ldots (a_s)_\rho h^{(k)}(0)}{k!r!(b_1)_\rho \ldots (b_t)_\rho}\frac{C_{\phi}^{\kappa,\rho}(\boldsymbol{\Sigma}^{-1}, \boldsymbol{\Omega})C_{\phi}(\boldsymbol{\Lambda})}{C_{\phi}(\boldsymbol{I}_m)},
\end{aligned}$$

where $c_{n,m}^{HWG}$ *as in (3.3),* $\sum_{\phi}$ *as in (3.5) and* $\theta_{\phi}^{\kappa,\rho} = \frac{C_{\phi}^{\kappa,\rho}(\boldsymbol{I}_m, \boldsymbol{I}_m)}{C_{\phi}(\boldsymbol{I}_m)}$*.*

Proof. From R.53 the density function of $\boldsymbol{\Lambda}$ is given by

$$g'(\boldsymbol{\Lambda}) = \frac{\pi^{\frac{m^2}{2}}}{\Gamma_m\left(\frac{m}{2}\right)}\prod_{i<j}^{m}\left(\lambda_i - \lambda_j\right)\int_{O_m} f(\boldsymbol{H\Lambda H}')d\boldsymbol{H}. \tag{3.7}$$

Note that from (3.2),

$$\begin{aligned}
\int_{O_m} f(\boldsymbol{H\Lambda H}')d\boldsymbol{H} &= c_{n,m}^{HWG}|\boldsymbol{\Sigma}|^{-\frac{n}{2}}\int_{O_m}|\boldsymbol{H\Lambda H}'|^{\frac{n}{2}-\frac{m+1}{2}}h\left(tr\left(\boldsymbol{\Sigma}^{-1}\boldsymbol{H\Lambda H}'\right)\right) \\
&\quad \times {}_sF_t\left(\boldsymbol{a};\boldsymbol{b};\boldsymbol{\Omega H\Lambda H}'\right)d\boldsymbol{H} \\
&= c_{n,m}^{HWG}|\boldsymbol{\Sigma}|^{-\frac{n}{2}}|\boldsymbol{\Lambda}|^{\frac{n}{2}-\frac{m+1}{2}}\int_{O_m}h\left(tr\left(\boldsymbol{\Sigma}^{-1}\boldsymbol{H\Lambda H}'\right)\right) \\
&\quad \times {}_sF_t\left(\boldsymbol{a};\boldsymbol{b};\boldsymbol{\Omega H\Lambda H}'\right)d\boldsymbol{H}.
\end{aligned} \tag{3.8}$$

Since $h(\cdot)$ admits the Taylor expansion (see R.7) and using R.34, R.26 and R.27,

$$\begin{aligned}
&\int_{O_m} h\left(tr\left(\boldsymbol{\Sigma}^{-1}\boldsymbol{H}\boldsymbol{\Lambda}\boldsymbol{H}'\right)\right)\ {}_sF_t\left(\boldsymbol{a};\boldsymbol{b};\boldsymbol{\Omega}\boldsymbol{H}\boldsymbol{\Lambda}\boldsymbol{H}'\right)d\boldsymbol{H}\\
&= \sum_{k=0}^{\infty}\sum_{\kappa}\sum_{r=0}^{\infty}\sum_{\rho}\frac{(a_1)_\rho\ldots(a_s)_\rho h^{(k)}(0)}{k!r!(b_1)_\rho\ldots(b_t)_\rho}\int_{O_m} C_\kappa\left(\boldsymbol{\Sigma}^{-1}\boldsymbol{H}\boldsymbol{\Lambda}\boldsymbol{H}'\right)C_\rho\left(\boldsymbol{\Omega}\boldsymbol{H}\boldsymbol{\Lambda}\boldsymbol{H}'\right)d\boldsymbol{H}\\
&= \sum_{\phi}\frac{\theta_\phi^{\kappa,\rho}(a_1)_\rho\ldots(a_s)_\rho h^{(k)}(0)}{k!r!(b_1)_\rho\ldots(b_t)_\rho}\int_{O_m} C_\phi^{\kappa,\rho}\left(\boldsymbol{\Sigma}^{-1}\boldsymbol{H}\boldsymbol{\Lambda}\boldsymbol{H}',\boldsymbol{\Omega}\boldsymbol{H}\boldsymbol{\Lambda}\boldsymbol{H}'\right)d\boldsymbol{H}\\
&= \sum_{\phi}\frac{\theta_\phi^{\kappa,\rho}(a_1)_\rho\ldots(a_v)_\rho h^{(k)}(0)}{k!r!(b_1)_\rho\ldots(b_w)_\rho}\frac{C_\phi^{\kappa,\rho}(\boldsymbol{\Sigma}^{-1},\boldsymbol{\Omega})C_\phi(\boldsymbol{\Lambda})}{C_\phi(\boldsymbol{I}_m)}.
\end{aligned}\tag{3.9}$$

From (3.7), (3.8), (3.9) and $|\boldsymbol{\Lambda}|^{\frac{n}{2}-\frac{m+1}{2}} = \prod_{i=1}^{m}\lambda_i^{\frac{n}{2}-\frac{m+1}{2}}$ the theorem is proved. □

Remark 3.3.2. *Note that if $\boldsymbol{\Sigma} = c_1 I_m, \boldsymbol{\Omega} = c_2 I_m$, from (3.7) follows*

$$g'(\boldsymbol{\Lambda}) = \frac{c_{n,m}^{HWG}\pi^{\frac{m^2}{2}}}{\Gamma_m\left(\frac{m}{2}\right)}\prod_{i<j}^{m}(\lambda_i-\lambda_j)\,c_1^{-\frac{nm}{2}}\prod_{i=1}^{m}\lambda_i^{\frac{n}{2}-\frac{m+1}{2}}h\left(c_1^{-1}tr\boldsymbol{\Lambda}\right)\ {}_sF_t\left(\boldsymbol{a};\boldsymbol{b};c_2\boldsymbol{\Lambda}\right). \tag{3.10}$$

Figure 3.1 illustrates the joint density function of the eigenvalues of $\boldsymbol{X}_{m\times m}$ for $m = 2, \boldsymbol{\Sigma} = c_1 I_m, \boldsymbol{\Omega} = c_2 I_m, s = 0, t = 0$, and $h(x) = exp(x)$ for specific values of m and c_1 and different values of n and c_2, while Figure 3.2 illustrates the joint density functions of $\boldsymbol{X}$ for $h(x) = exp(x)$, $h(x) = (1+x)$ and $h(x) = {}_1F_1(1;2;x)$ (see code in Appendix C.1). Note that for $s = 0$ and $t = 0$, (3.10) simplifies to

$$g'(\boldsymbol{\Lambda}) = \frac{c_{n,m}^{HWG}\pi^{\frac{m^2}{2}}}{\Gamma_m\left(\frac{m}{2}\right)}\prod_{i<j}^{m}(\lambda_i-\lambda_j)\,c_1^{-\frac{nm}{2}}\prod_{i=1}^{m}\lambda_i^{\frac{n}{2}-\frac{m+1}{2}}h\left(c_1^{-1}tr\boldsymbol{\Lambda}\right)etr(c_2\boldsymbol{\Lambda}). \tag{3.11}$$

Remark 3.3.3. *Note that if $s = t = 0, h(x) = exp(-\frac{x}{2})$, (3.2) simplifies to the density function of a Wishart distribution with parameters $\left(\frac{c_1}{1-2c_1c_2}\right)\boldsymbol{I}_m$ and n (see Definition 2.1.1), given by*

$$f(\boldsymbol{X}) = \left[2^{\frac{nm}{2}}\Gamma_m\left(\frac{n}{2}\right)\right]^{-1}\left(\frac{c_1}{1-2c_1c_2}\right)^{-\frac{nm}{2}}etr\left(\left(c_2-\frac{1}{2c_1}\right)\boldsymbol{X}\right), \tag{3.12}$$

hence (3.11) simplifies to

$$g'(\boldsymbol{\Lambda}) = \frac{\pi^{\frac{m^2}{2}}(1-2c_1c_2)^{\frac{nm}{2}}}{2^{\frac{nm}{2}}\Gamma_m\left(\frac{m}{2}\right)\Gamma_m\left(\frac{n}{2}\right)}\prod_{i<j}^{m}(\lambda_i-\lambda_j)\,c_1^{-\frac{nm}{2}}\prod_{i=1}^{m}\lambda_i^{\frac{n}{2}-\frac{m+1}{2}}etr\left(\left(c_2-\frac{1}{2c_1}\right)\boldsymbol{\Lambda}\right), \tag{3.13}$$

as given in Remark 2.3.1.

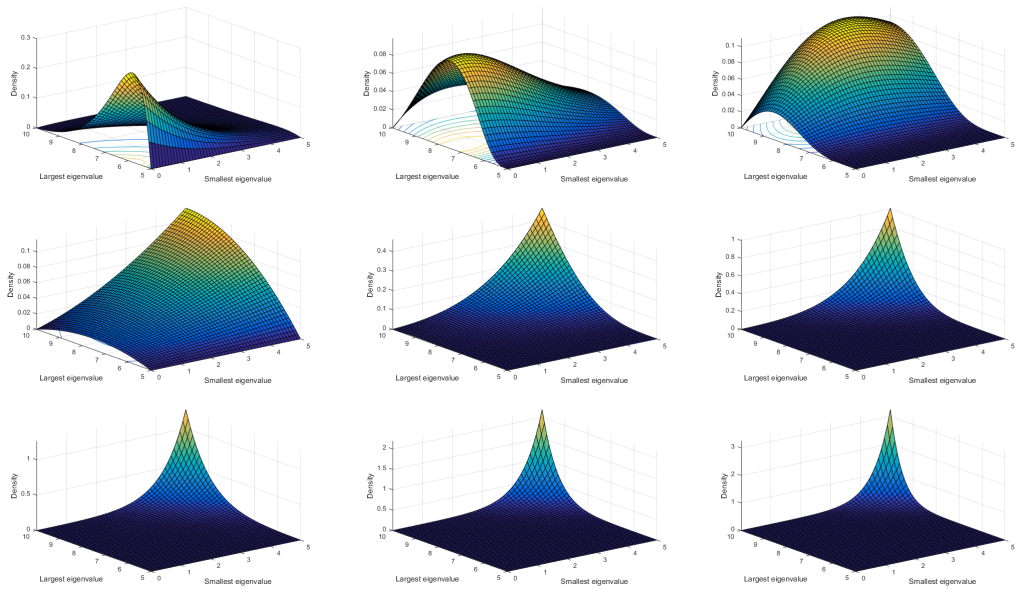


FIGURE 3.1: Joint density function of the largest (x-axis) and smallest eigenvalue (y-axis) for $m = 2$ and $n = 5$ (Left), $n = 10$ (Middle) and $n = 15$ (Right) with $c_1 = 1; c_2 = -2$ (Top), $c_2 = -1$ (Middle) and $c_2 = 0$ (Bottom).

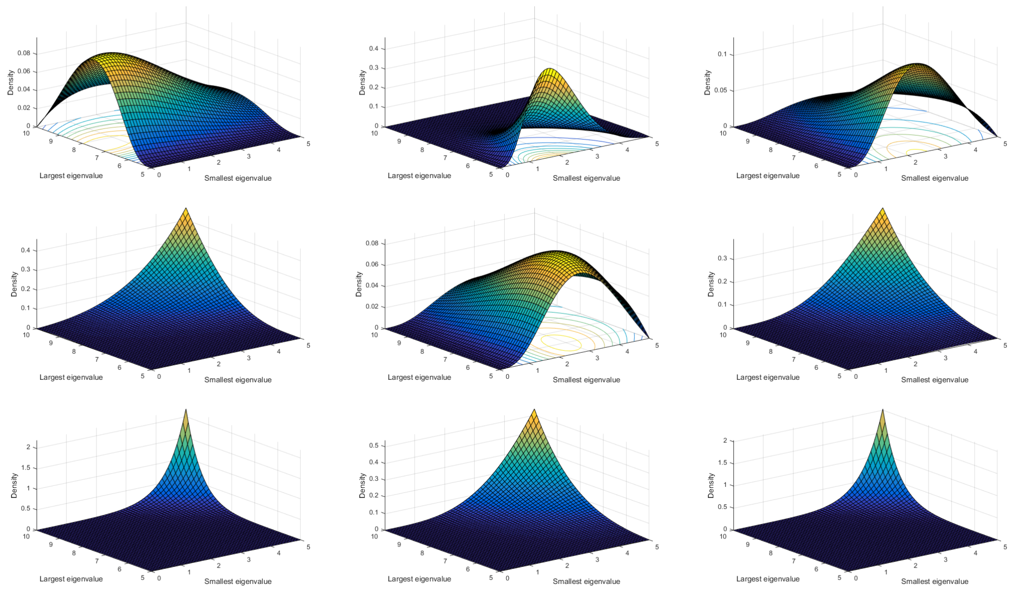


FIGURE 3.2: Joint density function of the largest (x-axis) and smallest eigenvalue (y-axis) for $m = 2$ and $n = 10$ with $h(x) = exp(x)$ (Left), $h(x) = (1+x)$ (Middle), $h(x) = {}_1F_1(1;2;x)$ (Right), $c_1 = 1; c_2 = -2$ (Top), $c_2 = -1$ (Middle) and $c_2 = 0$ (Bottom).

The inherent flexibility of this distribution is clear from Figures 3.1 and 3.2. Positive and negative correlation between the eigenvalues can be obtained by specifying certain $h(.)$ and values for the parameters.

3.4 Some special generator distributions

This section focuses on special cases of (3.2) and link it to existing literature.

Table 3.1 lists some members of the hypergeometric Wishart generator distribution (see (3.2)). Note that this is not an exhaustive list.

3.4.1 Noncentral Wishart generator distribution

As another direct case of Definition 3.2.1, taking $s=0$, $t=1$, $b_1>0$ and $n\geq m$ gives the density function as

$$f(\boldsymbol{X}) = c_{n,m}^{NWG}|\boldsymbol{\Sigma}|^{-\frac{n}{2}}|\boldsymbol{X}|^{\frac{n}{2}-\frac{m+1}{2}}\ {}_0F_1\left(b_1;\boldsymbol{\Omega}\boldsymbol{X}\right)h\left(tr\left(\boldsymbol{\Sigma}^{-1}\boldsymbol{X}\right)\right). \tag{3.14}$$

The normalizing constant is given by

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

The parameters are restricted to take those values for which the density function is non-negative. The density function (3.14) is termed the *non-central Wishart generator distribution* (NWGD) and denoted as $\boldsymbol{X}\sim NWG_m(\boldsymbol{\Sigma},\boldsymbol{\Omega},b_1,n,h)$. This distribution will not be studied in this thesis, but may form part of future research (see Chapter 6).

The density function of $\boldsymbol{Z}=\boldsymbol{Y}'\boldsymbol{Y}$, where $\boldsymbol{Y}$ has a matrix elliptical distribution i.e. $\boldsymbol{Y}\sim E_{p,m}(\boldsymbol{\mu},\boldsymbol{\Sigma},\boldsymbol{I},g'')$, was derived by Teng, Fang, and Deng (1989) (see also Dìaz-Garcìa, Gutièrrez-Jàimez, et al. (2011) and Dìaz-Garcìa and Jàimez (2006)). This is of a similar functional form as in Definition 3.2.1 if we take $s=0$, $t=1$, $b_1=\frac{n}{2}$, $\boldsymbol{\Omega}=\boldsymbol{\Sigma}^{-\frac{1}{2}}\boldsymbol{\Upsilon}\boldsymbol{\Sigma}^{-\frac{1}{2}}$ and $h(X)=g^{(2)}\left(X+tr\left(\boldsymbol{\Upsilon}\right)\right)$.

3.4.2 Wishart generator distribution

Note that for $s=0, t=0, n\geq m$ and $\boldsymbol{\Omega}=\boldsymbol{0}$ in Definition 3.2.1, the density function simplifies to

$$f(\boldsymbol{X}) = c_{n,m}^{WG}|\boldsymbol{\Sigma}|^{-\frac{n}{2}}|\boldsymbol{X}|^{\frac{n}{2}-\frac{m+1}{2}}h\left(tr\left(\boldsymbol{\Sigma}^{-1}\boldsymbol{X}\right)\right), \tag{3.15}$$

with

$$(c_{n,m}^{WG})^{-1} = \frac{\Gamma_m(\frac{n}{2})\gamma_0(\frac{n}{2})}{\Gamma(\frac{nm}{2})}, \quad \gamma_0\left(\frac{n}{2}\right) = \int_{\mathbb{R}^+} y^{\frac{nm}{2}-1}h(y)dy, \tag{3.16}$$

from R.24, provided that the above integral exists. The parameters are restricted to take those values for which the density function is non-negative. The density function in (3.15) is termed the *Wishart generator distribution* (WGD) and denoted as $\boldsymbol{X}\sim WG_m(\boldsymbol{\Sigma},n,h)$. The name of the distribution is due to the fact that for $h(x)=exp(-\frac{x}{2})$ in (3.15), yields the Wishart distribution (see Definition 2.1.1) with $c_{n,m}^{WG}=\frac{\Gamma(\frac{nm}{2})}{\Gamma_m(\frac{n}{2})\gamma_0(\frac{n}{2})}$ and $\gamma_0\left(\frac{n}{2}\right)=2^{\frac{nm}{2}}\Gamma\left(\frac{nm}{2}\right)$. This specific form is also called the Wishart elliptical distribution by Caro-Lopera, Gonzàlez-Farìas, and Balakrishnan (2014).

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

TABLE 3.1: Some special cases of the hypergeometric Wishart generator distribution (3.2)

3.4.3 Hypergeometric Wishart distribution

The hypergeometric Wishart distribution is considered in Section 3.5 as a prior. From (3.2) with $s=0, t=0, n \geq m, \mathbf{\Omega} = -\mathbf{\Sigma}^{-1}$ and $h(x) = {}_vF_w(\boldsymbol{a}^*;\boldsymbol{b}^*;cx), v \leq w, \boldsymbol{a}^* = \{a_1^*,...,a_v^*\}, \boldsymbol{b}^* = \{b_1^*,...,b_w^*\}$, the hypergeometric Wishart density function is given by

$$\begin{aligned} f(\boldsymbol{X}) &= \frac{1}{\Gamma_m\left(\frac{n}{2}\right) {}_{v+1}F_w\left(\frac{nm}{2},\boldsymbol{a}^*;\boldsymbol{b}^*;c\right)} |\mathbf{\Sigma}|^{-\frac{n}{2}} |\boldsymbol{X}|^{\frac{n}{2}-\frac{m+1}{2}} \\ &\quad \times {}_vF_w\{\boldsymbol{a}^*;\boldsymbol{b}^*;c\, tr(\mathbf{\Sigma}^{-1}\boldsymbol{X})\} etr(-\mathbf{\Sigma}^{-1}\boldsymbol{X}) \end{aligned} \quad (3.17)$$

as in Table 3.1, denoted as $\boldsymbol{X} \sim HW_m(\mathbf{\Sigma},\boldsymbol{a}^*,\boldsymbol{b}^*,c,n)$. The parameters are restricted to take those values for which the density function is non-negative.

Remark 3.4.1. *Suppose $\boldsymbol{X} \sim HW_m(\mathbf{\Sigma},\boldsymbol{a}^*,\boldsymbol{b}^*,c,n)$, then the r^{th} moment of $|\boldsymbol{X}|$ follows from (3.17) as*

$$E\left(|\boldsymbol{X}|^r\right) = \frac{\Gamma_m\left(\frac{n+2r}{2}\right) {}_{v+1}F_w\left(\frac{(n+2r)m}{2},\boldsymbol{a}^*;\boldsymbol{b}^*;c\right)}{\Gamma_m\left(\frac{n}{2}\right) {}_{v+1}F_w\left(\frac{nm}{2},\boldsymbol{a}^*;\boldsymbol{b}^*;c\right)} |\mathbf{\Sigma}|^r. \quad (3.18)$$

Remark 3.4.2. *Note that for $v=w=0$ the density function in (3.17) simplifies to*

$$\begin{aligned} &\frac{1}{\Gamma_m\left(\frac{n}{2}\right) {}_1F_0\left(\frac{nm}{2};c\right)} |\mathbf{\Sigma}|^{-\frac{n}{2}} |\boldsymbol{X}|^{\frac{n}{2}-\frac{m+1}{2}} {}_0F_0\left(c tr\left(\mathbf{\Sigma}^{-1}\boldsymbol{X}\right)\right) etr(-\mathbf{\Sigma}^{-1}\boldsymbol{X}) \\ &= \left(2^{\frac{nm}{2}}\Gamma_m\left(\frac{n}{2}\right)\right)^{-1} [2(1-c)]^{\frac{nm}{2}} |\mathbf{\Sigma}|^{-\frac{n}{2}} |\boldsymbol{X}|^{\frac{n}{2}-\frac{m+1}{2}} etr\left(-\frac{1}{2}2[1-c]\mathbf{\Sigma}^{-1}\boldsymbol{X}\right), \end{aligned}$$

for $|c|<1$ from R.35 and R.36. This can be identified as the Wishart density function (see Definition 2.1.1) with parameters $\frac{1}{2(1-c)}\mathbf{\Sigma}$ and n.

Figure 3.3 displays the density functions of the hypergeometric Wishart distribution (left) with $v=w=1, a_1=1, b_1=2, m=2, n=5$ and the Wishart distribution (right) with $m=2, n=5$ for different c (see code in Appendix C.2) as presented in Remark 3.4.2 for

$$\boldsymbol{X} = \begin{bmatrix} x_1 & 0 \\ 0 & x_2 \end{bmatrix}$$

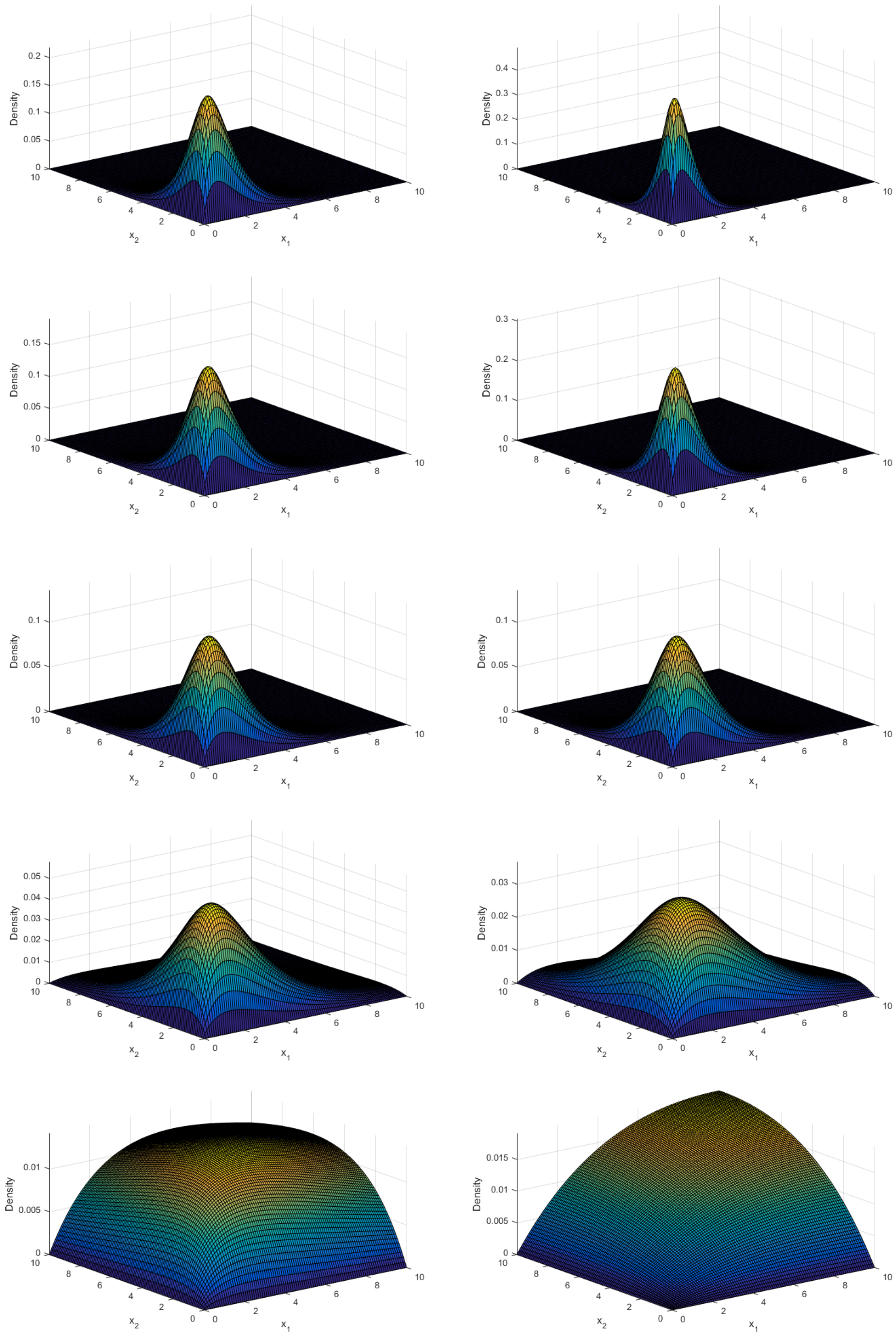


FIGURE 3.3: Density functions of the hypergeometric Wishart (left) and Wishart (right) distributions for $c = -0.9, -0.5, 0, 0.5, 0.9$ (Top to bottom)

It is apparent from Figure 3.3 that this hypergeometric Wishart distribution has heavier tails than the Wishart distribution for $c < 0$ and lighter tails for $c \geq 0$.

The hypergeometric gamma distribution is the univariate counterpart of (3.17), as given in the following definition.

Definition 3.4.1. *The hypergeometric gamma distribution (HGD) with parameters $\boldsymbol{a}^*, \boldsymbol{b}^*, c, \varphi$ and n has density function*

$$\begin{aligned} f(X) &= \frac{\varphi^{-\frac{n}{2}}}{\Gamma(\frac{n}{2})\ {}_{v+1}F_w\left(\frac{n}{2}, \boldsymbol{a}^*; \boldsymbol{b}^*; c\right)} \\ &\quad \times X^{\frac{n}{2}-1} exp\left(-\varphi^{-1}X\right)\ {}_vF_w\left(\boldsymbol{a}^*; \boldsymbol{b}^*; c\varphi^{-1}X\right), \end{aligned} \tag{3.19}$$

and is denoted by $X \sim HG(\boldsymbol{a}^*, \boldsymbol{b}^*, c, \varphi, n)$ for $X > 0, \varphi > 0$. The values of the parameters are chosen such that the density function is non-negative with ${}_vF_w(.)$ defined in R.29 and $v \leq w$. The r^{th} moment of X is given by

$$E[X^r] = \frac{\varphi^r \Gamma(\frac{n}{2}+r)\ {}_{v+1}F_w\left(\frac{n}{2}+r, \boldsymbol{a}^*; \boldsymbol{b}^*; c\right)}{\Gamma(\frac{n}{2})\ {}_{v+1}F_w\left(\frac{n}{2}, \boldsymbol{a}^*; \boldsymbol{b}^*; c\right)}, \tag{3.20}$$

Remark 3.4.3. *For $v = w = 0$ and $|c| < 1$ the density function (3.19) reduces to*

$$\begin{aligned} f(X) &= \frac{\varphi^{-\frac{n}{2}} X^{\frac{n}{2}-1}}{\Gamma(\frac{n}{2})\ (1-c)^{-\frac{n}{2}}} exp\left(-(1-c)\varphi^{-1}X\right) \\ &= \frac{\left(\frac{(1-c)}{\varphi}\right)^{\frac{n}{2}}}{\Gamma(\frac{n}{2})} X^{\frac{n}{2}-1} exp\left(-\frac{(1-c)}{\varphi}X\right), \end{aligned}$$

where $X > 0, \varphi > 0$, which is the density function of a gamma random variable with parameters $\frac{n}{2}$ and $\frac{(1-c)}{\varphi}$ (see R.41), denoted by $G(\frac{n}{2}, \frac{(1-c)}{\varphi})$, since ${}_1F_0\left(\frac{n}{2}; c\right) = (1-c)^{-\frac{n}{2}}$ from R.31 and ${}_0F_0\left(c\varphi^{-1}\sigma^2\right) = exp\left(c\varphi^{-1}\sigma^2\right)$ from R.30.

Remark 3.4.4. *Definition 3.4.1 can also be rewritten as an infinite mixture of gamma density functions by using the series expansion of ${}_vF_w(.)$ (see R.29) as follows*

$$\begin{aligned} f(X) &= \frac{\varphi^{-\frac{n}{2}}}{\Gamma(\frac{n}{2})\ {}_{v+1}F_w\left(\frac{n}{2}, \boldsymbol{a}^*; \boldsymbol{b}^*; c\right)} \\ &\quad \times \sum_{k=0}^{\infty} \frac{(a_1^*)_k \cdots (a_v^*)_k}{k!\,(b_1^*)_k \cdots (b_w^*)_k} c^k \varphi^{-(\frac{n}{2}+k)} X^{\frac{n}{2}+k-1} exp\left(-\varphi^{-1}X\right). \end{aligned}$$

3.5 Bayesian applications

In this section a special case of the HWGD, the *hypergeometric Wishart* distribution (3.17) and the *hypergeometric gamma* distribution (3.19), is applied in the Bayesian analysis of the matrix variate and univariate normal models respectively as subjective priors, as outlayed in Figure 3.4. The hypergeometric gamma distribution exists in literature, but has never been applied as a prior.

In Section 2.4 the underlying model was assumed to be elliptical instead of normal. In this section the normal model is assumed. In the matrix variate case, i.e. $\boldsymbol{X} \sim N_{m\times p}(\boldsymbol{\mu}, \boldsymbol{\Sigma} \otimes \boldsymbol{\Omega})$, the hypergeometric Wishart prior (3.17) is assumed for $\boldsymbol{\Sigma}$. In the univariate case, i.e. $X \sim N(\mu, \sigma^2)$, the hypergeometric gamma (3.19) prior is assumed for σ^2.

Note that the SEL function (see R.60) is used throughout this chapter.

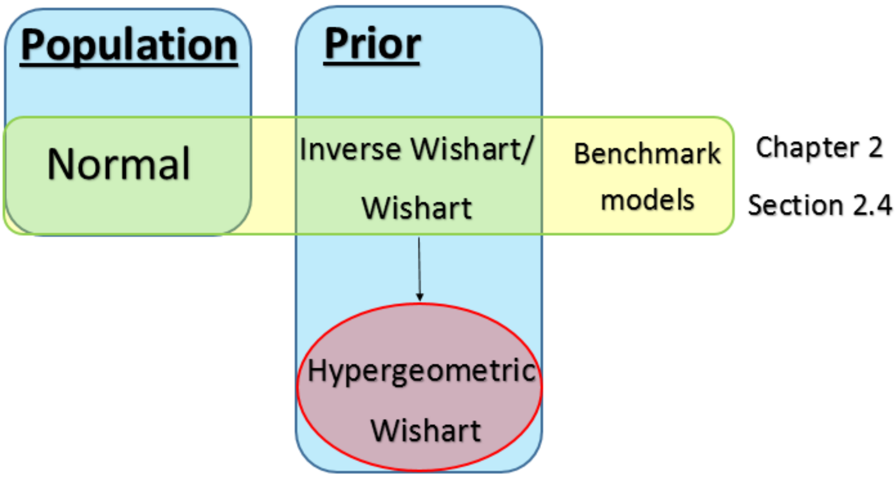


FIGURE 3.4: Outlay of Section 3.5

3.5.1 Univariate Bayesian analysis

Consider a sample of n_1 observations, $\boldsymbol{X} = \{X_1, \ldots, X_{n_1}\}$, from the normal model $N(\mu, \sigma^2)$, (see R.39) where both parameters are unknown, with the following likelihood function

$$\begin{aligned} L(\mu, \sigma^2 | \boldsymbol{X}) &= \left(2\pi\sigma^2\right)^{-\frac{n_1}{2}} exp\left(-\frac{1}{2\sigma^2}\sum_{i=1}^{n_1}(X_i - \mu)^2\right) \\ &\propto \left(\sigma^2\right)^{-\frac{n_1}{2}} exp\left(-\frac{1}{2\frac{\sigma^2}{n_1}}\left(\left(\mu - \overline{X}\right)^2 + \frac{1}{n_1}\sum_{i=1}^{n_1} X_i^2 - \overline{X}^2\right)\right). \end{aligned} \quad (3.21)$$

Further assume an improper objective prior (see R.55) for $\mu, \pi(\mu) = 1$, and independently the hypergeometric gamma distribution (see (3.19)) as the subjective prior for σ^2 such that the joint prior density function is given by

$$\begin{aligned}\pi\left(\mu,\sigma^2\right) &= \frac{\varphi^{-\frac{n}{2}}\left(\sigma^2\right)^{\frac{n}{2}-1}}{\Gamma(\frac{n}{2})\ {}_{v+1}F_w\left(\frac{n}{2},\boldsymbol{a}^*;\boldsymbol{b}^*;c\right)} exp\left(-\varphi^{-1}\sigma^2\right)\\ &\quad \times {}_vF_w\left(\boldsymbol{a}^*;\boldsymbol{b}^*;c\varphi^{-1}\sigma^2\right),\end{aligned} \tag{3.22}$$

with $\pi\left(\mu,\sigma^2\right)=\pi\left(\mu\right)\pi\left(\sigma^2\right)$.

Then, the joint posterior density function is obtained from (3.21) and (3.22) as follows

$$\begin{aligned}q\left(\mu,\sigma^2|\boldsymbol{X}\right) &\propto \left(\sigma^2\right)^{\frac{n}{2}-\frac{n_1}{2}-1}\ {}_vF_w\left(\boldsymbol{a}^*;\boldsymbol{b}^*;c\varphi^{-1}\sigma^2\right)\\ &\quad \times exp\left(-\varphi^{-1}\sigma^2\right)exp\left(-\frac{1}{2\frac{\sigma^2}{n_1}}\left(\left(\mu-\overline{X}\right)^2+\frac{1}{n_1}\sum_{i=1}^{n_1}X_i^2-\overline{X}^2\right)\right).\end{aligned} \tag{3.23}$$

Remark 3.5.1. *From* (3.23) *it follows that*

$$q\left(\mu|\sigma^2,\boldsymbol{X}\right)\propto exp\left(-\frac{1}{2\frac{\sigma^2}{n_1}}\left(\mu-\overline{X}\right)^2\right),$$

hence $\mu|\sigma^2,\boldsymbol{X}\sim N(\overline{X},\frac{\sigma^2}{n_1})$, *and*

$$q\left(\sigma^2|\mu,\boldsymbol{X}\right)\propto\left(\sigma^2\right)^{\frac{n}{2}-\frac{n_1}{2}-1}\ {}_vF_w\left(\boldsymbol{a}^*;\boldsymbol{b}^*;c\varphi^{-1}\sigma^2\right)exp\left(-\varphi^{-1}\sigma^2\right)exp\left(-\frac{1}{2\sigma^2}\sum_{i=1}^{n_1}\left(X_i-\mu\right)^2\right). \tag{3.24}$$

To obtain the marginal posterior density functions, the mathematical solutions of the complicated integrals lead to unattractive expressions for the posterior density functions. In the derivation of the marginal posterior density function of μ, the following intergral is obtained,

$$\begin{aligned}&\int_0^\infty\left(\sigma^2\right)^{\frac{n}{2}-\frac{m}{2}-1}\ {}_vF_w\left(\boldsymbol{a}^*;\boldsymbol{b}^*;c\varphi^{-1}\sigma^2\right)\\ &\times exp\left(-\varphi^{-1}\sigma^2\right)exp\left(-\frac{1}{2\sigma^2}\sum_{i=1}^{m}\left(x_i-\mu\right)^2\right)d\sigma^2\end{aligned}$$

which can be solved mathematically by using the series expansion of ${}_vF_w(.)$ (see R.29) and the definition of the Bessel function of the third kind (see R.14), as

$$\sum_{k=0}^{\infty}K_{\frac{n}{2}-\frac{m}{2}+k}\left(2\sqrt{\sum_{i=1}^{m}\left(x_i-\mu\right)^2\left(1-c\right)\varphi^{-1}}\right).$$

This leads to a computationally challenging posterior density function which is not implementable and Gibbs sampling will have to be used to approximate this posterior density function.
However, if the integrand is viewed as the product of a gamma density function and some function of σ^2, the integral can be expressed as an expected value. A novel approach is proposed to

obtain the analytical expressions in situations where the usual mathematical approach leads to computationally complicated forms in the following remark.

Remark 3.5.2. *In general, let $h(\boldsymbol{\theta}|\boldsymbol{X})$ be the likelihood function of $\boldsymbol{\theta} = (\theta_1, \theta_2), \theta_i \in \Omega_i, i = 1,2$ for given $\boldsymbol{x}$. The prior density assumed is $\pi(\boldsymbol{\theta})$. Consider the marginal posterior density function of θ_i:*

$$\begin{aligned} q(\theta_i, \boldsymbol{X}) &= \int_{\Omega_j} h(\boldsymbol{\theta}|\boldsymbol{X})\pi(\boldsymbol{\theta})d\theta_j \\ &= \int_{\Omega*} g(\theta_i, \phi|\boldsymbol{X}) f(\phi) d\phi \\ &= E_\phi \left(g(\theta_i, \phi|\boldsymbol{X})\right) \end{aligned}$$

with ϕ a function of θ_j, $\phi \in \Omega*$, $g(.)$ a function of θ_i, ϕ and $\boldsymbol{X}$, and $f(\phi)$ a known density function from which random variates can be simulated. This result can then be computationally calculated using the central limit theorem as

$$E_\phi \left(\widehat{g(\theta_i, \phi}|\boldsymbol{X})\right) = \frac{1}{n_1} \sum_{k=1}^{n_1} g(\theta_i, \phi_k|\boldsymbol{x})$$

where $\phi_k, k = 1, \ldots, n_1$ is a random variate from the distribution with density function $f(\phi)$.

Theorem 3.5.1. *The marginal posterior density function of μ for the normal model with prior (3.22) is*

$$\begin{aligned} q(\mu|\boldsymbol{X}) = &\frac{\Gamma\left(\frac{n}{2}\right)}{\left(2\pi n_1^{-1}\varphi\right)^{\frac{1}{2}} \Gamma\left(\frac{n}{2}+\frac{1}{2}\right)} \\ &\times \frac{E_{\sigma_1^2}\left(\left(\sigma_1^2\right)^{-\frac{n_1}{2}} exp\left(-\frac{1}{2\sigma_1^2}\sum_{i=1}^{n_1}\left(X_i-\mu\right)^2\right) {}_vF_w\left(\boldsymbol{a}^*;\boldsymbol{b}^*;c\varphi^{-1}\sigma_1^2\right)\right)}{E_{\sigma_2^2}\left(\left(\sigma_2^2\right)^{-\frac{n_1}{2}} exp\left(-\frac{1}{2\sigma_2^2}\left(\sum_{i=1}^{n_1}X_i^2 - n_1\overline{X}^2\right)\right) {}_vF_w\left(\boldsymbol{a}^*;\boldsymbol{b}^*;c\varphi^{-1}\sigma_2^2\right)\right)}, \end{aligned} \tag{3.25}$$

where $\sigma_1^2 \sim G\left(\frac{n}{2}, \varphi^{-1}\right)$ and $\sigma_2^2 \sim G(\frac{n}{2}+\frac{1}{2}, \varphi^{-1})$, provided $\varphi > 0$ and $E_{\sigma_i^2}(\cdot)$ denotes the expected value with respect to the distribution of $\sigma_i^2, i = 1,2$.

Proof. From (3.23), Remark 3.5.2 and R.41, the marginal posterior density function of μ is proportional to

$$\begin{aligned} &\int_0^\infty \left(\sigma^2\right)^{\frac{n}{2}-\frac{n_1}{2}-1} {}_vF_w\left(\boldsymbol{a}^*;\boldsymbol{b}^*;c\varphi^{-1}\sigma^2\right) \\ &\times exp\left(-\varphi^{-1}\sigma^2\right) exp\left(-\frac{1}{2\sigma^2}\sum_{i=1}^{n_1}\left(X_i-\mu\right)^2\right) d\sigma^2 \\ \propto\ & E_{\sigma_1^2}\left(\left(\sigma_1^2\right)^{-\frac{n_1}{2}} exp\left(-\frac{1}{2\sigma_1^2}\sum_{i=1}^{n_1}\left(X_i-\mu\right)^2\right) {}_vF_w\left(\boldsymbol{a}^*;\boldsymbol{b}^*;c\varphi^{-1}\sigma_1^2\right)\right), \end{aligned}$$

where $\sigma_1^2 \sim G\left(\frac{n}{2}, \varphi^{-1}\right)$. Hence

$$q(\mu|\boldsymbol{X}) = c_\mu E_{\sigma_1^2}\left((\sigma_1^2)^{-\frac{n_1}{2}} exp\left(-\frac{1}{2\sigma_1^2}\sum_{i=1}^{n_1}(X_i-\mu)^2\right) {}_vF_w\left(\boldsymbol{a}^*;\boldsymbol{b}^*;c\varphi^{-1}\sigma_1^2\right)\right). \tag{3.26}$$

Expressing the expected value in integral form using R.41, the following is obtained,

$$\begin{aligned}
c_\mu^{-1} &= \int_{-\infty}^{\infty} E_{\sigma_1^2}\left((\sigma_1^2)^{-\frac{n_1}{2}} exp\left(-\frac{1}{2\sigma_1^2}\sum_{i=1}^{n_1}(X_i-\mu)^2\right) {}_vF_w\left(\boldsymbol{a}^*;\boldsymbol{b}^*;c\varphi^{-1}\sigma_1^2\right)\right) d\mu \\
&= \frac{\varphi^{-\frac{n}{2}}}{\Gamma\left(\frac{n}{2}\right)}\int_0^\infty (\sigma^2)^{\frac{n}{2}-\frac{n_1}{2}-1} {}_vF_w\left(\boldsymbol{a}^*;\boldsymbol{b}^*;c\varphi^{-1}\sigma^2\right) \\
&\quad \times exp\left(-\varphi^{-1}\sigma^2\right)\int_{-\infty}^{\infty} exp\left(-\frac{1}{2\frac{\sigma^2}{n_1}}\left((\mu-\overline{X})^2+\frac{1}{n_1}\sum_{i=1}^{n_1}X_i^2-\overline{X}^2\right)\right) d\mu d\sigma^2 \\
&= \int_0^\infty \frac{\varphi^{-\frac{n}{2}}}{\Gamma\left(\frac{n}{2}\right)}(\sigma^2)^{\frac{n}{2}-\frac{n_1}{2}-1} {}_vF_w\left(\boldsymbol{a}^*;\boldsymbol{b}^*;c\varphi^{-1}\sigma^2\right) \\
&\quad \times exp\left(-\varphi^{-1}\sigma^2\right) exp\left(-\frac{1}{2\sigma^2}\left(\sum_{i=1}^{n_1}X_i^2-n_1\overline{X}^2\right)\right)\int_{-\infty}^{\infty} exp\left(-\frac{1}{2\frac{\sigma^2}{n_1}}(\mu-\overline{X})^2\right) d\mu d\sigma^2 \\
&= \int_0^\infty \left(2\pi\frac{\sigma^2}{n_1}\right)^{\frac{1}{2}} \frac{\varphi^{-\frac{n}{2}}}{\Gamma\left(\frac{n}{2}\right)}(\sigma^2)^{\frac{n}{2}-\frac{n_1}{2}-1} {}_vF_w\left(\boldsymbol{a}^*;\boldsymbol{b}^*;c\varphi^{-1}\sigma^2\right) \\
&\quad \times exp\left(-\varphi^{-1}\sigma^2\right) exp\left(-\frac{1}{2\sigma^2}\left(\sum_{i=1}^{n_1}X_i^2-n_1\overline{X}^2\right)\right) d\sigma^2 \\
&= \left(2\pi n_1^{-1}\varphi\right)^{\frac{1}{2}}\frac{\Gamma\left(\frac{n}{2}+\frac{1}{2}\right)}{\Gamma\left(\frac{n}{2}\right)}\int_0^\infty \frac{\varphi^{-\frac{n}{2}-\frac{1}{2}}}{\Gamma\left(\frac{n}{2}+\frac{1}{2}\right)} {}_vF_w\left(\boldsymbol{a}^*;\boldsymbol{b}^*;c\varphi^{-1}\sigma^2\right) \\
&\quad \times (\sigma^2)^{\frac{n}{2}-\frac{n_1}{2}-\frac{1}{2}} exp\left(-\varphi^{-1}\sigma^2\right) exp\left(-\frac{1}{2\sigma^2}\left(\sum_{i=1}^{n_1}X_i^2-n_1\overline{X}^2\right)\right) d\sigma^2 \\
&= \left(2\pi n_1^{-1}\varphi\right)^{\frac{1}{2}}\frac{\Gamma\left(\frac{n}{2}+\frac{1}{2}\right)}{\Gamma\left(\frac{n}{2}\right)} \\
&\quad \times E_{\sigma_2^2}\left((\sigma_2^2)^{-\frac{n_1}{2}} {}_vF_w\left(\boldsymbol{a}^*;\boldsymbol{b}^*;c\varphi^{-1}\sigma_2^2\right) exp\left(-\frac{1}{2\sigma_2^2}\left(\sum_{i=1}^{n_1}X_i^2-n_1\overline{X}^2\right)\right)\right),
\end{aligned} \tag{3.27}$$

using Remark 3.5.2 and R.41 with $\sigma_2^2 \sim G(\frac{n}{2}+\frac{1}{2}, \varphi^{-1})$. Hence (3.25) follows from (3.26) and (3.27). □

Theorem 3.5.2. *The marginal posterior density function of σ^2 for the normal model with prior (3.22) is*

$$\begin{aligned} & q\left(\sigma^2|\boldsymbol{X}\right) \\ = & \left(E_{\sigma_2^2}\left(\left(\sigma_2^2\right)^{-\frac{n_1}{2}}\, {}_vF_w\left(\boldsymbol{a}^*;\boldsymbol{b}^*;c\varphi^{-1}\sigma_2^2\right) exp\left(-\frac{1}{2\sigma_2^2}\left(\sum_{i=1}^{n_1}X_i^2-n_1\overline{X}^2\right)\right)\right)\right)^{-1} \\ & \times\frac{1}{\Gamma\left(\frac{n}{2}+\frac{1}{2}\right)\varphi^{\frac{n}{2}+\frac{1}{2}}}\left(\sigma^2\right)^{\frac{n}{2}+\frac{1}{2}-1}\, {}_vF_w\left(\boldsymbol{a}^*;\boldsymbol{b}^*;c\varphi^{-1}\sigma^2\right) \\ & \times exp\left(-\varphi^{-1}\sigma^2\right) exp\left(-\frac{1}{2\sigma^2}\left(\sum_{i=1}^{n_1}X_i^2-n_1\overline{X}^2\right)\right), \end{aligned} \tag{3.28}$$

with $\sigma_2^2 \sim G(\frac{n}{2}+\frac{1}{2},\varphi^{-1})$ *provided* $\varphi>0$.

Proof. From (3.23) and R.39 the marginal posterior density function of σ^2 is given by

$$\begin{aligned} & q\left(\sigma^2|\boldsymbol{X}\right) \\ \propto & \left(\sigma^2\right)^{\frac{n}{2}-\frac{n_1}{2}-1}\, {}_vF_w\left(\boldsymbol{a}^*;\boldsymbol{b}^*;c\varphi^{-1}\sigma^2\right) \\ & \times exp\left(-\varphi^{-1}\sigma^2\right) exp\left(-\frac{1}{2\sigma^2}\left(\sum_{i=1}^{n_1}X_i^2-n_1\overline{X}^2\right)\right)\int_{-\infty}^{\infty} exp\left(-\frac{1}{2\frac{\sigma^2}{n_1}}\left(\mu-\overline{X}\right)^2\right)d\mu \\ \propto & \left(\sigma^2\right)^{\frac{n}{2}-\frac{n_1}{2}+\frac{1}{2}-1}\, {}_vF_w\left(\boldsymbol{a}^*;\boldsymbol{b}^*;c\varphi^{-1}\sigma^2\right) exp\left(-\varphi^{-1}\sigma^2\right) exp\left(-\frac{1}{2\sigma^2}\left(\sum_{i=1}^{n_1}X_i^2-n_1\overline{X}^2\right)\right), \end{aligned} \tag{3.29}$$

with the integrating constant

$$\begin{aligned} c_{\sigma^2}^{-1} = & \int_0^{\infty}\left(\sigma^2\right)^{\frac{n}{2}-\frac{n_1}{2}+\frac{1}{2}-1}\, {}_vF_w\left(\boldsymbol{a}^*;\boldsymbol{b}^*;c\varphi^{-1}\sigma^2\right) \\ & \times exp\left(-\varphi^{-1}\sigma^2\right) exp\left(-\frac{1}{2\sigma^2}\left(\sum_{i=1}^{n_1}X_i^2-n_1\overline{X}^2\right)\right)d\sigma^2 \\ = & \Gamma\left(\frac{n}{2}+\frac{1}{2}\right)\varphi^{\frac{n}{2}+\frac{1}{2}} \\ & \times E_{\sigma_2^2}\left(\left(\sigma_2^2\right)^{-\frac{n_1}{2}}\, {}_vF_w\left(\boldsymbol{a}^*;\boldsymbol{b}^*;c\varphi^{-1}\sigma_2^2\right) exp\left(-\frac{1}{2\sigma_2^2}\left(\sum_{i=1}^{n_1}X_i^2-n_1\overline{X}^2\right)\right)\right), \end{aligned} \tag{3.30}$$

from Remark 3.5.2 and R.41, with $\sigma_2^2 \sim G(\frac{n}{2}+\frac{1}{2},\varphi^{-1})$. Therefore (3.28) follows from (3.29) and (3.30). □

The following theorems provide the Bayes estimators of μ and σ^2.

Theorem 3.5.3. *The Bayes estimator of μ for the normal model with prior (3.22) is*

$$\widehat{\mu}_{B,HG} = E[\mu|\boldsymbol{X}] = \overline{X}. \tag{3.31}$$

Proof. Consider that the expected value of $\mu - \overline{X}$ is given by

$$\begin{aligned}
& E[\mu - \overline{X}|\boldsymbol{X}] \\
= & \int_{-\infty}^{\infty} (\mu - \overline{X})\, q(\mu|\boldsymbol{X})\, d\mu \\
= & \frac{\varphi^{-\frac{1}{2}} \left(2\pi n_1^{-1}\right)^{-\frac{1}{2}} \Gamma\left(\frac{n}{2}\right) \left(\Gamma\left(\frac{n}{2}+\frac{1}{2}\right)\right)^{-1}}{E_{\sigma_2^2}\left((\sigma_2^2)^{-\frac{n_1}{2}} exp\left(-\frac{1}{2\sigma_2^2}\left(\sum_{i=1}^{n_1} X_i^2 - n_1\overline{X}^2\right)\right)\ {}_vF_w\left(\boldsymbol{a}^*;\boldsymbol{b}^*;c\varphi^{-1}\sigma_2^2\right)\right)} \\
& \times \int_{-\infty}^{\infty} (\mu - \overline{X})\, E_{\sigma_1^2}\left((\sigma_1^2)^{-\frac{n_1}{2}} exp\left(-\frac{1}{2\sigma_1^2}\sum_{i=1}^{n_1}(X_i-\mu)^2\right)\ {}_vF_w\left(\boldsymbol{a}^*;\boldsymbol{b}^*;c\varphi^{-1}\sigma_1^2\right)\right) d\mu,
\end{aligned}$$

from (3.25). It is quite clear that the integrand is an odd function and therefore

$$\int_{-\infty}^{\infty} (\mu - \overline{X})\, E_{\sigma_1^2}\left(exp\left((\sigma_1^2)^{-\frac{n_1}{2}} - \frac{1}{2\sigma_1^2}\sum_{i=1}^{n_1}(X_i-\mu)^2\right)\ {}_vF_w\left(\boldsymbol{a}^*;\boldsymbol{b}^*;c\varphi^{-1}\sigma_1^2\right)\right) d\mu = 0.$$

It can then be concluded that $E[\mu - \overline{X}|\boldsymbol{X}] = 0$ and hence

$$E[\mu|\boldsymbol{X}] = \widehat{\mu}_{B,HG} = \overline{X}.$$

□

Theorem 3.5.4. *The Bayes estimator of σ^2 for the normal model with prior (3.22) is given by*

$$\widehat{\sigma}^2_{B,HG} = \frac{\left(\frac{n}{2}+\frac{1}{2}\right)\varphi E_{\sigma_3^2}\left((\sigma_3^2)^{-\frac{n_1}{2}}\ {}_vF_w\left(\boldsymbol{a}^*;\boldsymbol{b}^*;c\varphi^{-1}\sigma_3^2\right) exp\left(-\frac{1}{2\sigma_3^2}\left(\sum_{i=1}^{n_1} X_i^2 - n_1\overline{X}^2\right)\right)\right)}{E_{\sigma_2^2}\left((\sigma_2^2)^{-\frac{n_1}{2}}\ {}_vF_w\left(\boldsymbol{a}^*;\boldsymbol{b}^*;c\varphi^{-1}\sigma_2^2\right) exp\left(-\frac{1}{2\sigma_2^2}\left(\sum_{i=1}^{n_1} X_i^2 - n_1\overline{X}^2\right)\right)\right)}, \tag{3.32}$$

with $\sigma_2^2 \sim G(\frac{n}{2}+\frac{1}{2},\varphi^{-1})$ and $\sigma_3^2 \sim G(\frac{n}{2}+\frac{3}{2},\varphi^{-1})$.

Proof. From (3.28), Remark 3.5.2 and R.41, follows that

$$\begin{aligned}
&\widehat{\sigma}^2_{B,HG} = E[\sigma^2|\boldsymbol{X}] \\
= &\left(E_{\sigma_2^2}\left((\sigma_2^2)^{-\frac{n_1}{2}}\ {}_vF_w\left(\boldsymbol{a}^*;\boldsymbol{b}^*;c\varphi^{-1}\sigma_2^2\right) exp\left(-\frac{1}{2\sigma_2^2}\left(\sum_{i=1}^{n_1}X_i^2-n_1\overline{X}^2\right)\right)\right)\right)^{-1} \\
&\times\int_0^\infty \frac{1}{\Gamma\left(\frac{n}{2}+\frac{1}{2}\right)\varphi^{\frac{n}{2}+\frac{1}{2}}}\left(\sigma^2\right)^{\frac{n}{2}+\frac{3}{2}-1}\ {}_vF_w\left(\boldsymbol{a}^*;\boldsymbol{b}^*;c\varphi^{-1}\sigma^2\right) \\
&\times exp\left(-\varphi^{-1}\sigma^2\right) exp\left(-\frac{1}{2\sigma^2}\left(\sum_{i=1}^{n_1}X_i^2-n_1\overline{X}^2\right)\right) d\sigma^2 \\
= &\frac{\left(\frac{n}{2}+\frac{1}{2}\right)\ \varphi E_{\sigma_3^2}\left((\sigma_3^2)^{-\frac{n_1}{2}}\ {}_vF_w\left(\boldsymbol{a}^*;\boldsymbol{b}^*;c\varphi^{-1}\sigma_3^2\right) exp\left(-\frac{1}{2\sigma_3^2}\left(\sum_{i=1}^{n_1}X_i^2-n_1\overline{X}^2\right)\right)\right)}{E_{\sigma_2^2}\left((\sigma_2^2)^{-\frac{n_1}{2}}\ {}_vF_w\left(\boldsymbol{a}^*;\boldsymbol{b}^*;c\varphi^{-1}\sigma_2^2\right) exp\left(-\frac{1}{2\sigma_2^2}\left(\sum_{i=1}^{n_1}X_i^2-n_1\overline{X}^2\right)\right)\right)}.
\end{aligned}$$

where $\sigma_2^2 \sim G(\frac{n}{2}+\frac{1}{2},\varphi^{-1})$ and $\sigma_3^2 \sim G(\frac{n}{2}+\frac{3}{2},\varphi^{-1})$. □

3.5.1.1. Special case

Referring to Remark 3.4.3 the posterior distributions of μ and σ^2 when assuming the gamma prior can be obtained from (3.25) and (3.28).

Theorem 3.5.5. *The marginal posterior density functions of μ and σ^2 of the normal model with the gamma prior is*

$$q\left(\mu|\boldsymbol{X}\right) = \frac{(1-c)^{\frac{1}{4}}2^{\frac{1}{4}}\left(\sum_{i=1}^{n_1}\left(X_i-\mu\right)^2\right)^{\frac{n}{4}-\frac{n_1}{4}} K_{\frac{n}{2}-\frac{n_1}{2}}\left(2\sqrt{\sum_{i=1}^{n_1}\left(X_i-\mu\right)^2(1-c)\varphi^{-1}}\right)}{\left(2\pi n_1^{-1}\right)^{\frac{1}{2}}\left(\sum_{i=1}^{n_1}X_i^2-n_1\overline{X}^2\right)^{\frac{n}{4}-\frac{n_1}{4}+\frac{1}{4}} K_{\frac{n}{2}-\frac{n_1}{2}+\frac{1}{2}}\left(2\sqrt{\left(\sum_{i=1}^{n_1}X_i^2-n_1\overline{X}^2\right)(1-c)\varphi^{-1}}\right)},$$

and

$$\begin{aligned}
q\left(\sigma^2|\boldsymbol{X}\right) \quad = \quad &\frac{\left(-\frac{\sum_{i=1}^{n_1}X_i^2-n_1\overline{X}^2}{2c\varphi^{-1}}\right)^{-\frac{n}{4}+\frac{n_1}{4}-\frac{1}{4}}}{2K_{\frac{n}{4}-\frac{m}{4}+\frac{1}{4}}\left(2\sqrt{-\left(\sum_{i=1}^{n_1}X_i^2-n_1\overline{X}^2\right)c\varphi^{-1}}\right)}\left(\sigma^2\right)^{\frac{n}{2}-\frac{n_1}{2}+\frac{1}{2}-1} \\
&\times exp\left(-(1-c)\varphi^{-1}\sigma^2\right) exp\left(-\frac{1}{2\sigma^2}\left(\sum_{i=1}^{n_1}X_i^2-n_1\overline{X}^2\right)\right),
\end{aligned}$$

where $K_v(.)$ is the Bessel function of the third kind (see R.14).

Proof. Since $\sigma_1^2 \sim Gamma\left(\frac{n}{2}, \varphi^{-1}\right)$ and $\sigma_2^2 \sim Gamma(\frac{n}{2}+\frac{1}{2}, \varphi^{-1})$, the posterior density function of μ can be obtained from (3.25) as

$$
\begin{aligned}
& q(\mu|\boldsymbol{X}) \\
= & \frac{\Gamma\left(\frac{n}{2}\right) E_{\sigma_1^2}\left((\sigma_1^2)^{-\frac{n_1}{2}} exp\left(-\frac{1}{2\sigma_1^2}\sum_{i=1}^{n_1}(X_i-\mu)^2\right)\, exp\left(c\varphi^{-1}\sigma_1^2\right)\right)}{\left(2\pi n_1^{-1}\varphi\right)^{\frac{1}{2}}\Gamma\left(\frac{n}{2}+\frac{1}{2}\right) E_{\sigma_2^2}\left((\sigma_2^2)^{-\frac{n_1}{2}} exp\left(-\frac{1}{2\sigma_2^2}\left(\sum_{i=1}^{n_1}X_i^2-n_1\overline{X}^2\right)\right)\, exp\left(c\varphi^{-1}\sigma_2^2\right)\right)} \\
= & \frac{\Gamma\left(\frac{n}{2}\right)\frac{1}{\Gamma\left(\frac{n}{2}\right)\varphi^{\frac{n}{2}}}2\left(\frac{\sum_{i=1}^{n_1}(X_i-\mu)^2}{2(1-c)\varphi^{-1}}\right)^{\frac{n}{4}-\frac{n_1}{4}}}{\left(2\pi n_1^{-1}\varphi\right)^{\frac{1}{2}}\Gamma\left(\frac{n}{2}+\frac{1}{2}\right)\frac{1}{\Gamma\left(\frac{n}{2}+\frac{1}{2}\right)\varphi^{\frac{n}{2}+\frac{1}{2}}}2\left(\frac{\sum_{i=1}^{n_1}X_i^2-n_1\overline{X}^2}{2(1-c)\varphi^{-1}}\right)^{\frac{n}{4}-\frac{n_1}{4}+\frac{1}{4}}} \\
& \times\frac{K_{\frac{n}{2}-\frac{n_1}{2}}\left(2\sqrt{\sum_{i=1}^{n_1}(X_i-\mu)^2(1-c)\varphi^{-1}}\right)}{K_{\frac{n}{2}-\frac{n_1}{2}+\frac{1}{2}}\left(2\sqrt{\left(\sum_{i=1}^{n_1}X_i^2-n_1\overline{X}^2\right)(1-c)\varphi^{-1}}\right)} \\
= & \frac{(1-c)^{\frac{1}{4}}2^{\frac{1}{4}}\left(\sum_{i=1}^{n_1}(X_i-\mu)^2\right)^{\frac{n}{4}-\frac{n_1}{4}}K_{\frac{n}{2}-\frac{n_1}{2}}\left(2\sqrt{\sum_{i=1}^{n_1}(X_i-\mu)^2(1-c)\varphi^{-1}}\right)}{\left(2\pi n_1^{-1}\right)^{\frac{1}{2}}\left(\sum_{i=1}^{n_1}X_i^2-n_1\overline{X}^2\right)^{\frac{n}{4}-\frac{n_1}{4}+\frac{1}{4}}K_{\frac{n}{2}-\frac{n_1}{2}+\frac{1}{2}}\left(2\sqrt{\left(\sum_{i=1}^{n_1}X_i^2-n_1\overline{X}^2\right)(1-c)\varphi^{-1}}\right)},
\end{aligned}
$$

since from R.14,

$$
\begin{aligned}
& E_{\sigma_1^2}\left((\sigma_1^2)^{-\frac{n_1}{2}} exp\left(-\frac{1}{2\sigma_1^2}\sum_{i=1}^{n_1}(X_i-\mu)^2\right)\, exp\left(c\varphi^{-1}\sigma_1^2\right)\right) \\
= & \int_0^\infty \frac{1}{\Gamma\left(\frac{n}{2}\right)\varphi^{\frac{n}{2}}}(\sigma_1^2)^{\frac{n}{2}-\frac{n_1}{2}} exp\left(-\frac{1}{2\sigma_1^2}\sum_{i=1}^{n_1}(X_i-\mu)^2\right)\, exp\left(-(1-c)\varphi^{-1}\sigma_1^2\right)d\sigma_1^2 \\
= & \frac{1}{\Gamma\left(\frac{n}{2}\right)\varphi^{\frac{n}{2}}}2\left(\frac{\sum_{i=1}^{n_1}(X_i-\mu)^2}{2(1-c)\varphi^{-1}}\right)^{\frac{n}{4}-\frac{n_1}{4}}K_{\frac{n}{2}-\frac{n_1}{2}}\left(2\sqrt{\sum_{i=1}^{n_1}(X_i-\mu)^2(1-c)\varphi^{-1}}\right).
\end{aligned}
$$

Similarly it follows that

$$
\begin{aligned}
& E_{\sigma_2^2}\left((\sigma_2^2)^{-\frac{n_1}{2}} exp\left(-\frac{1}{2\sigma_2^2}\left(\sum_{i=1}^{n_1}X_i^2-n_1\overline{X}^2\right)\right)\, exp\left(c\varphi^{-1}\sigma_2^2\right)\right) \\
= & \frac{1}{\Gamma\left(\frac{n}{2}+\frac{1}{2}\right)\varphi^{\frac{n}{2}+\frac{1}{2}}}2\left(\frac{\sum_{i=1}^{n_1}X_i^2-n_1\overline{X}^2}{2(1-c)\varphi^{-1}}\right)^{\frac{n}{4}-\frac{n_1}{4}+\frac{1}{4}} \\
\times & K_{\frac{n}{2}-\frac{n_1}{2}+\frac{1}{2}}\left(2\sqrt{\left(\sum_{i=1}^{n_1}X_i^2-n_1\overline{X}^2\right)(1-c)\varphi^{-1}}\right)
\end{aligned}
$$

From (3.28) follows that

$$
\begin{aligned}
& q\left(\sigma^2|\boldsymbol{X}\right) \\
= & \left(E_{\sigma_2^2}\left((\sigma_2^2)^{-\frac{n_1}{2}} exp\left(c\varphi^{-1}\sigma_2^2\right) exp\left(-\frac{1}{2\sigma_2^2}\left(\sum_{i=1}^{n_1} X_i^2 - n_1\overline{X}^2\right)\right)\right)\right)^{-1} \\
& \times\frac{1}{\Gamma\left(\frac{n}{2}+\frac{1}{2}\right)\varphi^{\frac{n}{2}+\frac{1}{2}}}\left(\sigma^2\right)^{\frac{n}{2}-\frac{n_1}{2}+\frac{1}{2}-1} exp\left(c\varphi^{-1}\sigma^2\right) \\
& \times exp\left(-\varphi^{-1}\sigma^2\right) exp\left(-\frac{1}{2\sigma^2}\left(\sum_{i=1}^{n_1} X_i^2 - n_1\overline{X}^2\right)\right) \\
= & \frac{\left(-\frac{\sum_{i=1}^{n_1} X_i^2 - n_1\overline{X}^2}{2c\varphi^{-1}}\right)^{-\frac{n}{4}+\frac{n_1}{4}-\frac{1}{4}}}{2K_{\frac{n}{4}-\frac{m}{4}+\frac{1}{4}}\left(2\sqrt{-\left(\sum_{i=1}^{n_1} X_i^2 - n_1\overline{X}^2\right)c\varphi^{-1}}\right)}\left(\sigma^2\right)^{\frac{n}{2}-\frac{n_1}{2}+\frac{1}{2}-1} \\
& \times exp\left(-(1-c)\varphi^{-1}\sigma^2\right) exp\left(-\frac{1}{2\sigma^2}\left(\sum_{i=1}^{n_1} X_i^2 - n_1\overline{X}^2\right)\right),
\end{aligned}
$$

since

$$
\begin{aligned}
& E_{\sigma_2^2}\left((\sigma_2^2)^{-\frac{n_1}{2}} exp\left(-\frac{1}{2\sigma_2^2}\left(\sum_{i=1}^{n_1} X_i^2 - n_1\overline{X}^2\right)\right) exp\left(c\varphi^{-1}\sigma_2^2\right)\right) \\
= & \frac{1}{\Gamma\left(\frac{n}{2}+\frac{1}{2}\right)\varphi^{\frac{n}{2}+\frac{1}{2}}} 2\left(-\frac{\sum_{i=1}^{n_1} X_i^2 - n_1\overline{X}^2}{2c\varphi^{-1}}\right)^{\frac{n}{4}-\frac{n_1}{4}+\frac{1}{4}} \\
& \times K_{\frac{n}{4}-\frac{n_1}{4}+\frac{1}{4}}\left(2\sqrt{-\left(\sum_{i=1}^{n_1} X_i^2 - n_1\overline{X}^2\right)c\varphi^{-1}}\right)
\end{aligned}
$$

from R.14. □

Theorem 3.5.6. *The Bayes estimators of* μ *and* σ^2 *for the normal model with the gamma prior follows are*

$$\widehat{\mu}_{B,G} = \overline{X},$$

and

$$\widehat{\sigma}^2_{B,G} = \left(-\frac{\sum_{i=1}^{n_1} X_i^2 - n_1\overline{X}^2}{2c\varphi^{-3}}\right)^{\frac{1}{2}} \frac{K_{\frac{n}{4}-\frac{n_1}{4}+\frac{3}{4}}\left(2\sqrt{-\left(\sum_{i=1}^{n_1} X_i^2 - n_1\overline{X}^2\right)c\varphi^{-1}}\right)}{K_{\frac{n}{4}-\frac{n_1}{4}+\frac{1}{4}}\left(2\sqrt{-\left(\sum_{i=1}^{n_1} X_i^2 - n_1\overline{X}^2\right)c\varphi^{-1}}\right)}.$$

Proof. From (3.31) and (3.32) the Bayes estimators are given by

$$\widehat{\mu}_{B,G} = \overline{X}$$

and

$$\begin{aligned}
\widehat{\sigma}^2_{B,G} &= \frac{\left(\frac{n}{2}+\frac{1}{2}\right)\varphi E_{\sigma_3^2}\left(\left(\sigma_3^2\right)^{-\frac{n_1}{2}} exp\left(c\varphi^{-1}\sigma_3^2\right) exp\left(-\frac{1}{2\sigma_3^2}\left(\sum_{i=1}^{n_1} X_i^2 - n_1\overline{X}^2\right)\right)\right)}{E_{\sigma_2^2}\left(\left(\sigma_2^2\right)^{-\frac{n_1}{2}} exp\left(c\varphi^{-1}\sigma_2^2\right) exp\left(-\frac{1}{2\sigma_2^2}\left(\sum_{i=1}^{n_1} X_i^2 - n_1\overline{X}^2\right)\right)\right)} \\
&= \left(-\frac{\sum_{i=1}^{n_1} X_i^2 - n_1\overline{X}^2}{2c\varphi^{-3}}\right)^{\frac{1}{2}} \frac{K_{\frac{n}{4}-\frac{n_1}{4}+\frac{3}{4}}\left(2\sqrt{-\left(\sum_{i=1}^{n_1} X_i^2 - n_1\overline{X}^2\right)c\varphi^{-1}}\right)}{K_{\frac{n}{4}-\frac{n_1}{4}+\frac{1}{4}}\left(2\sqrt{-\left(\sum_{i=1}^{n_1} X_i^2 - n_1\overline{X}^2\right)c\varphi^{-1}}\right)}
\end{aligned}$$

since

$$\begin{aligned}
& E_{\sigma_2^2}\left(\left(\sigma_2^2\right)^{-\frac{n_1}{2}} exp\left(-\frac{1}{2\sigma_2^2}\left(\sum_{i=1}^{n_1} X_i^2 - n_1\overline{X}^2\right)\right) exp\left(c\varphi^{-1}\sigma_2^2\right)\right) \\
&= \frac{1}{\Gamma\left(\frac{n}{2}+\frac{1}{2}\right)\varphi^{\frac{n}{2}+\frac{1}{2}}} 2\left(-\frac{\sum_{i=1}^{n_1} X_i^2 - n_1\overline{X}^2}{2c\varphi^{-1}}\right)^{\frac{n}{4}-\frac{n_1}{4}+\frac{1}{4}} \\
&\quad \times K_{\frac{n}{4}-\frac{n_1}{4}+\frac{1}{4}}\left(2\sqrt{-\left(\sum_{i=1}^{n_1} X_i^2 - n_1\overline{X}^2\right)c\varphi^{-1}}\right)
\end{aligned}$$

and

$$\begin{aligned}
& E_{\sigma_3^2}\left(\left(\sigma_3^2\right)^{-\frac{n_1}{2}} exp\left(c\varphi^{-1}\sigma_3^2\right) exp\left(-\frac{1}{2\sigma_3^2}\left(\sum_{i=1}^{n_1} X_i^2 - m\overline{X}^2\right)\right)\right) \\
&= \frac{1}{\Gamma\left(\frac{n}{2}+\frac{3}{2}\right)\varphi^{\frac{n}{2}+\frac{3}{2}}} 2\left(-\frac{\sum_{i=1}^{n_1} X_i^2 - n_1\overline{X}^2}{2c\varphi^{-1}}\right)^{\frac{n}{4}-\frac{n_1}{4}+\frac{3}{4}} \\
&\quad \times K_{\frac{n}{4}-\frac{n_1}{4}+\frac{3}{4}}\left(2\sqrt{-\left(\sum_{i=1}^{n_1} X_i^2 - n_1\overline{X}^2\right)c\varphi^{-1}}\right)
\end{aligned}$$

from R.14. □

Remark 3.5.3. *The results in Theorems 3.5.5 and 3.5.6 were derived by Bhattacharya and Saxena (1989) and Van Niekerk (2012).*

3.5.2 Multivariate Bayesian analysis

Consider the matrix variate normal model, $\boldsymbol{X} \sim N_{m,p}(\boldsymbol{\mu}, \boldsymbol{\Sigma} \otimes \boldsymbol{\Omega})$ with density function R.49. Based on a random sample of size n_1 from this distribution, $\boldsymbol{X} = \{\boldsymbol{X}_1, \ldots, \boldsymbol{X}_{n_1}\}$, the likelihood function is obtained from R.49 as follows

$$\begin{aligned}
& L(\boldsymbol{\mu}, \boldsymbol{\Sigma}|\boldsymbol{X}) \\
= & \prod_{i=1}^{n_1} (2\pi)^{-\frac{mp}{2}} |\boldsymbol{\Sigma}|^{-\frac{p}{2}} |\boldsymbol{\Omega}|^{-\frac{m}{2}} etr\left(-\frac{1}{2}(\boldsymbol{X}_i - \boldsymbol{\mu})' \boldsymbol{\Sigma}^{-1} (\boldsymbol{X}_i - \boldsymbol{\mu}) \boldsymbol{\Omega}^{-1}\right) \\
\propto & |\boldsymbol{\Sigma}|^{-\frac{n_1 p}{2}} etr\left(-\frac{1}{2}\sum_{i=1}^{n_1} (\boldsymbol{X}_i - \boldsymbol{\mu})' \boldsymbol{\Sigma}^{-1} (\boldsymbol{X}_i - \boldsymbol{\mu}) \boldsymbol{\Omega}^{-1}\right) \\
= & |\boldsymbol{\Sigma}|^{-\frac{n_1 p}{2}} etr\left(-\frac{1}{2}\boldsymbol{\Sigma}^{-1}\left(\boldsymbol{V} + n_1(\overline{\boldsymbol{X}} - \boldsymbol{\mu})\boldsymbol{\Omega}^{-1}(\overline{\boldsymbol{X}} - \boldsymbol{\mu})'\right)\right), \qquad (3.33)
\end{aligned}$$

with $\boldsymbol{V}$ as defined in (2.7). Now consider an improper objective prior (see R.55) for $\boldsymbol{\mu}$, $\pi(\boldsymbol{\mu}) = 1$, and independently the hypergeometric Wishart distribution (3.17) with parameters $\boldsymbol{\Theta}, \boldsymbol{a}^*, \boldsymbol{b}^*, c$ and n as the prior for $\boldsymbol{\Sigma}$ such that the joint prior density function is

$$\begin{aligned}
\pi(\boldsymbol{\mu}, \boldsymbol{\Sigma}) = & \frac{|\boldsymbol{\Theta}|^{-\frac{n}{2}} |\boldsymbol{\Sigma}|^{\frac{n}{2} - \frac{m+1}{2}}}{\Gamma_m\left(\frac{n}{2}\right) {}_{v+1}F_w\left(\frac{nm}{2}, \boldsymbol{a}^*, \boldsymbol{b}^*; c\right)} etr\left(-\boldsymbol{\Theta}^{-1}\boldsymbol{\Sigma}\right) \\
& \times {}_vF_w\left(\boldsymbol{a}^*, \boldsymbol{b}^*; ctr\left(\boldsymbol{\Theta}^{-1}\boldsymbol{\Sigma}\right)\right). \qquad (3.34)
\end{aligned}$$

Remark 3.5.4. *For $v = w = 0$ and $c = \frac{1}{2}$ in (3.34) the prior density function of $\boldsymbol{\Sigma}$ simplifies as*

$$\pi(\boldsymbol{\Sigma}) = \frac{|\boldsymbol{\Theta}|^{-\frac{n}{2}} |\boldsymbol{\Sigma}|^{\frac{n}{2} - \frac{m+1}{2}}}{\Gamma_m\left(\frac{n}{2}\right) {}_1F_0\left(\frac{nm}{2}; \frac{1}{2}\right)} {}_0F_0\left(\frac{1}{2}tr\left(\boldsymbol{\Theta}^{-1}\boldsymbol{\Sigma}\right)\right) etr\left(-\boldsymbol{\Theta}^{-1}\boldsymbol{\Sigma}\right).$$

Note that ${}_1F_0\left(\frac{nm}{2}; \frac{1}{2}\right) = 2^{\frac{nm}{2}}$ from R.31 and ${}_0F_0\left(\frac{1}{2}tr\left(\boldsymbol{\Theta}^{-1}\boldsymbol{\Sigma}\right)\right) = etr\left(\frac{1}{2}\boldsymbol{\Theta}^{-1}\boldsymbol{\Sigma}\right)$ from R.35, hence (3.34) simplifies to

$$\pi(\boldsymbol{\mu}, \boldsymbol{\Sigma}) = \frac{1}{2^{\frac{nm}{2}} \Gamma_m\left(\frac{n}{2}\right)} |\boldsymbol{\Theta}|^{-\frac{n}{2}} |\boldsymbol{\Sigma}|^{\frac{n}{2} - \frac{m+1}{2}} etr\left(-\frac{1}{2}\boldsymbol{\Theta}^{-1}\boldsymbol{\Sigma}\right),$$

which is the Wishart density function with parameters $\boldsymbol{\Theta}$ and n (Definition 2.1.1). This prior was considered in Section 2.4 for the elliptical model.

From (3.33) and (3.34) the joint posterior density function is

$$\begin{aligned}
q(\boldsymbol{\mu}, \boldsymbol{\Sigma}|\boldsymbol{X}) \propto & |\boldsymbol{\Sigma}|^{-\frac{n_1 p}{2}} etr\left(-\frac{1}{2}\boldsymbol{\Sigma}^{-1}\left(\boldsymbol{V} + n_1(\overline{\boldsymbol{X}} - \boldsymbol{\mu})\boldsymbol{\Omega}^{-1}(\overline{\boldsymbol{X}} - \boldsymbol{\mu})'\right)\right) \\
& \times \frac{|\boldsymbol{\Theta}|^{-\frac{n}{2}} |\boldsymbol{\Sigma}|^{\frac{n}{2} - \frac{m+1}{2}}}{\Gamma_m\left(\frac{n}{2}\right) {}_{v+1}F_w\left(\frac{nm}{2}, \boldsymbol{a}^*, \boldsymbol{b}^*; c\right)} {}_vF_w\left(\boldsymbol{a}^*, \boldsymbol{b}^*; ctr\left(\boldsymbol{\Theta}^{-1}\boldsymbol{\Sigma}\right)\right) etr\left(-\boldsymbol{\Theta}^{-1}\boldsymbol{\Sigma}\right) \\
\propto & |\boldsymbol{\Sigma}|^{\frac{n - n_1 p}{2} - \frac{m+1}{2}} etr\left(-\frac{1}{2}\boldsymbol{\Sigma}^{-1}\left(\boldsymbol{V} + n_1(\overline{\boldsymbol{X}} - \boldsymbol{\mu})\boldsymbol{\Omega}^{-1}(\overline{\boldsymbol{X}} - \boldsymbol{\mu})'\right)\right) \\
& \times {}_vF_w\left(\boldsymbol{a}^*, \boldsymbol{b}^*; ctr\left(\boldsymbol{\Theta}^{-1}\boldsymbol{\Sigma}\right)\right) etr\left(-\boldsymbol{\Theta}^{-1}\boldsymbol{\Sigma}\right), \qquad (3.35)
\end{aligned}$$

with $\boldsymbol{V}$ as defined in (2.7). Note that from (3.35) follows that

$$q(\boldsymbol{\mu}|\boldsymbol{\Sigma},\boldsymbol{X}) \propto etr\left(-\frac{1}{2}\boldsymbol{\Sigma}^{-1}\left(n_1(\overline{\boldsymbol{X}}-\boldsymbol{\mu})\boldsymbol{\Omega}^{-1}(\overline{\boldsymbol{X}}-\boldsymbol{\mu})'\right)\right),$$

hence $\boldsymbol{\mu}|\boldsymbol{\Sigma},\boldsymbol{X} \sim N(\overline{\boldsymbol{X}},\boldsymbol{\Omega}\otimes\frac{1}{n_1}\boldsymbol{\Sigma})$, and

$$\begin{aligned} q(\boldsymbol{\Sigma}|\boldsymbol{\mu},\boldsymbol{X}) \propto\ & |\boldsymbol{\Sigma}|^{\frac{n-n_1p}{2}-\frac{m+1}{2}}\ {}_vF_w\left(\boldsymbol{a}^*,\boldsymbol{b}^*;ctr\left(\boldsymbol{\Theta}^{-1}\boldsymbol{\Sigma}\right)\right) \\ & \times etr\left(-\frac{1}{2}\boldsymbol{\Sigma}^{-1}\left(\boldsymbol{V}+n_1(\overline{\boldsymbol{X}}-\boldsymbol{\mu})\boldsymbol{\Omega}^{-1}(\overline{\boldsymbol{X}}-\boldsymbol{\mu})'\right)-\boldsymbol{\Theta}^{-1}\boldsymbol{\Sigma}\right). \end{aligned} \tag{3.36}$$

Theorem 3.5.7. *The marginal posterior density function of $\boldsymbol{\mu}$ for the normal model with prior (3.34) is*

$$\begin{aligned} & q(\boldsymbol{\mu}|\boldsymbol{X}) \\ = & \frac{\Gamma_m(\frac{n}{2})2^{\frac{nm}{2}}|\boldsymbol{\Omega}|^{-\frac{m}{2}}|\boldsymbol{\Theta}|^{-\frac{n+p}{2}}}{\left(\frac{2\pi}{n_1}\right)^{\frac{mp}{2}}\Gamma_m(\frac{n+p}{2})E_{\boldsymbol{\Sigma}_2}\left(|\boldsymbol{\Sigma}_2|^{-\frac{n_1p}{2}}\ {}_vF_w\left(\boldsymbol{a}^*,\boldsymbol{b}^*;ctr\left(\boldsymbol{\Theta}^{-1}\boldsymbol{\Sigma}_2\right)\right)etr\left(-\frac{1}{2}\boldsymbol{\Sigma}_2^{-1}\boldsymbol{V}\right)\right)} \\ & \times E_{\boldsymbol{\Sigma}_1}\left(|\boldsymbol{\Sigma}_1|^{-\frac{n_1p}{2}}etr\left(-\frac{1}{2}\boldsymbol{\Sigma}_1^{1}\left(\boldsymbol{V}+n_1(\overline{\boldsymbol{X}}-\boldsymbol{\mu})\boldsymbol{\Omega}^{-1}(\overline{\boldsymbol{X}}-\boldsymbol{\mu})'\right)\right)\ {}_vF_w\left(\boldsymbol{a}^*,\boldsymbol{b}^*;ctr\left(\boldsymbol{\Theta}^{-1}\boldsymbol{\Sigma}_1\right)\right)\right), \end{aligned} \tag{3.37}$$

where $\boldsymbol{\Sigma}_1 \sim W_m\left(\frac{1}{2}\boldsymbol{\Theta},n\right), \boldsymbol{\Sigma}_2 \sim W_m\left(\frac{1}{2}\boldsymbol{\Theta},n+p\right)$ and $\boldsymbol{V}$ as defined in (2.7).

Proof. From (3.35) the marginal posterior density function of $\boldsymbol{\mu}$ is

$$\begin{aligned} & q(\boldsymbol{\mu}|\boldsymbol{X}) \\ \propto & \int_{S_m}|\boldsymbol{\Sigma}|^{\frac{n-n_1p}{2}-\frac{m+1}{2}}etr\left(-\frac{1}{2}\boldsymbol{\Sigma}^{-1}\left(\boldsymbol{V}+n_1(\overline{\boldsymbol{X}}-\boldsymbol{\mu})\boldsymbol{\Omega}^{-1}(\overline{\boldsymbol{X}}-\boldsymbol{\mu})'\right)\right) \\ & \times {}_vF_w\left(\boldsymbol{a}^*,\boldsymbol{b}^*;ctr\left(\boldsymbol{\Theta}^{-1}\boldsymbol{\Sigma}\right)\right)etr\left(-\boldsymbol{\Theta}^{-1}\boldsymbol{\Sigma}\right)d\boldsymbol{\Sigma} \\ \propto & \int_{S_m}\frac{1}{\Gamma_m(\frac{n}{2})2^{\frac{nm}{2}}}|\boldsymbol{\Sigma}|^{\frac{n-n_1p}{2}-\frac{m+1}{2}}\left|\frac{1}{2}\boldsymbol{\Theta}\right|^{-\frac{n}{2}}etr\left(-\boldsymbol{\Theta}^{-1}\boldsymbol{\Sigma}\right) \\ & \times etr\left(-\frac{1}{2}\boldsymbol{\Sigma}^{-1}\left(\boldsymbol{V}+n_1(\overline{\boldsymbol{X}}-\boldsymbol{\mu})\boldsymbol{\Omega}^{-1}(\overline{\boldsymbol{X}}-\boldsymbol{\mu})'\right)\right)\ {}_vF_w\left(\boldsymbol{a}^*,\boldsymbol{b}^*;ctr\left(\boldsymbol{\Theta}^{-1}\boldsymbol{\Sigma}\right)\right)d\boldsymbol{\Sigma} \\ = & E_{\boldsymbol{\Sigma}_1}\left(|\boldsymbol{\Sigma}_1|^{-\frac{n_1p}{2}}etr\left(-\frac{1}{2}\boldsymbol{\Sigma}_1^{-1}\left(\boldsymbol{V}+n_1(\overline{\boldsymbol{X}}-\boldsymbol{\mu})\boldsymbol{\Omega}^{-1}(\overline{\boldsymbol{X}}-\boldsymbol{\mu})'\right)\right)\ {}_vF_w\left(\boldsymbol{a}^*,\boldsymbol{b}^*;ctr\left(\boldsymbol{\Theta}^{-1}\boldsymbol{\Sigma}_1\right)\right)\right), \end{aligned}$$

from Remark 3.5.2 and Definition 2.1.1, where $\boldsymbol{\Sigma}_1 \sim W_m\left(\frac{1}{2}\boldsymbol{\Theta},n\right)$. Hence

$$\begin{aligned} & q(\boldsymbol{\mu}|\boldsymbol{X}) \\ = & c_{\boldsymbol{\mu}}E_{\boldsymbol{\Sigma}_1}\left(|\boldsymbol{\Sigma}_1|^{-\frac{n_1p}{2}}etr\left(-\frac{1}{2}\boldsymbol{\Sigma}_1^{-1}\left(\boldsymbol{V}+n_1(\overline{\boldsymbol{X}}-\boldsymbol{\mu})\boldsymbol{\Omega}^{-1}(\overline{\boldsymbol{X}}-\boldsymbol{\mu})'\right)\right)\ {}_vF_w\left(\boldsymbol{a}^*,\boldsymbol{b}^*;ctr\left(\boldsymbol{\Theta}^{-1}\boldsymbol{\Sigma}_1\right)\right)\right), \end{aligned} \tag{3.38}$$

where

$$
\begin{aligned}
c_{\boldsymbol{\mu}}^{-1} &= \int_{\mathbb{R}^{m\times p}} E_{\boldsymbol{\Sigma}_1}(|\boldsymbol{\Sigma}_1|^{-\frac{n_1 p}{2}} etr\left(-\frac{1}{2}\boldsymbol{\Sigma}_1^{-1}\left(\boldsymbol{V}+n_1(\overline{\boldsymbol{X}}-\boldsymbol{\mu})\boldsymbol{\Omega}^{-1}(\overline{\boldsymbol{X}}-\boldsymbol{\mu})'\right)\right) \\
&\quad \times {}_vF_w\left(\boldsymbol{a}^*,\boldsymbol{b}^*;ctr\left(\boldsymbol{\Theta}^{-1}\boldsymbol{\Sigma}_1\right)\right))d\boldsymbol{\mu} \\
&= \int_{\mathbb{R}^{m\times p}}\int_{S_m} \frac{1}{\Gamma_m(\frac{n}{2})2^{\frac{nm}{2}}}|\boldsymbol{\Sigma}|^{\frac{n-n_1p}{2}-\frac{m+1}{2}}|\frac{1}{2}\boldsymbol{\Theta}|^{-\frac{n}{2}} etr\left(-\boldsymbol{\Theta}^{-1}\boldsymbol{\Sigma}\right) \\
&\quad \times etr\left(-\frac{1}{2}\boldsymbol{\Sigma}^{-1}\left(\boldsymbol{V}+n_1(\overline{\boldsymbol{X}}-\boldsymbol{\mu})\boldsymbol{\Omega}^{-1}(\overline{\boldsymbol{X}}-\boldsymbol{\mu})'\right)\right)\ {}_vF_w\left(\boldsymbol{a}^*,\boldsymbol{b}^*;ctr\left(\boldsymbol{\Theta}^{-1}\boldsymbol{\Sigma}\right)\right)d\boldsymbol{\Sigma}d\boldsymbol{\mu} \\
&= \int_{S_m} \frac{1}{\Gamma_m(\frac{n}{2})2^{\frac{nm}{2}}}|\boldsymbol{\Sigma}|^{\frac{n-n_1p}{2}-\frac{m+1}{2}}|\frac{1}{2}\boldsymbol{\Theta}|^{-\frac{n}{2}} etr\left(-\boldsymbol{\Theta}^{-1}\boldsymbol{\Sigma}\right) etr\left(-\frac{1}{2}\boldsymbol{\Sigma}^{-1}\boldsymbol{V}\right) \\
&\quad \times\ {}_vF_w\left(\boldsymbol{a}^*,\boldsymbol{b}^*;ctr\left(\boldsymbol{\Theta}^{-1}\boldsymbol{\Sigma}\right)\right)\int_{\mathbb{R}^{m\times p}} etr\left(-\frac{1}{2}\left(\frac{1}{n_1}\boldsymbol{\Sigma}\right)^{-1}(\overline{\boldsymbol{X}}-\boldsymbol{\mu})\boldsymbol{\Omega}^{-1}(\overline{\boldsymbol{X}}-\boldsymbol{\mu})'\right)d\boldsymbol{\mu}d\boldsymbol{\Sigma} \\
&= \int_{S_m} \frac{\left(\frac{2\pi}{n_1}\right)^{\frac{mp}{2}}|\boldsymbol{\Omega}|^{\frac{m}{2}}}{\Gamma_m(\frac{n}{2})2^{\frac{nm}{2}}}|\boldsymbol{\Sigma}|^{\frac{n-n_1p+p}{2}-\frac{m+1}{2}}|\frac{1}{2}\boldsymbol{\Theta}|^{-\frac{n}{2}} etr\left(-\boldsymbol{\Theta}^{-1}\boldsymbol{\Sigma}\right) etr\left(-\frac{1}{2}\boldsymbol{\Sigma}^{-1}\boldsymbol{V}\right) \\
&\quad \times {}_vF_w\left(\boldsymbol{a}^*,\boldsymbol{b}^*;ctr\left(\boldsymbol{\Theta}^{-1}\boldsymbol{\Sigma}\right)\right)d\boldsymbol{\Sigma} \\
&= \frac{\left(\frac{2\pi}{n_1}\right)^{\frac{mp}{2}}|\boldsymbol{\Omega}|^{\frac{m}{2}}\Gamma_m(\frac{n+p}{2})2^{\frac{(n+p)m}{2}}|\frac{1}{2}\boldsymbol{\Theta}|^{\frac{n+p}{2}}}{\Gamma_m(\frac{n}{2})2^{\frac{nm}{2}}}\int_{S_m}\frac{1}{\Gamma_m(\frac{n+p}{2})2^{\frac{(n+p)m}{2}}}|\frac{1}{2}\boldsymbol{\Theta}|^{-\frac{n+p}{2}} \\
&\quad \times|\boldsymbol{\Sigma}|^{\frac{n-n_1p+p}{2}-\frac{m+1}{2}} etr\left(-\boldsymbol{\Theta}^{-1}\boldsymbol{\Sigma}\right) etr\left(-\frac{1}{2}\boldsymbol{\Sigma}^{-1}\boldsymbol{V}\right)\ {}_vF_w\left(\boldsymbol{a}^*,\boldsymbol{b}^*;ctr\left(\boldsymbol{\Theta}^{-1}\boldsymbol{\Sigma}\right)\right)d\boldsymbol{\Sigma} \\
&= \frac{\left(\frac{2\pi}{n_1}\right)^{\frac{mp}{2}}|\boldsymbol{\Omega}|^{\frac{m}{2}}\Gamma_m(\frac{n+p}{2})|\boldsymbol{\Theta}|^{\frac{n+p}{2}}}{\Gamma_m(\frac{n}{2})2^{\frac{nm}{2}}} \\
&\quad \times E_{\boldsymbol{\Sigma}_2}\left(|\boldsymbol{\Sigma}_2|^{-\frac{n_1p}{2}}\ {}_vF_w\left(\boldsymbol{a}^*,\boldsymbol{b}^*;ctr\left(\boldsymbol{\Theta}^{-1}\boldsymbol{\Sigma}_2\right)\right)etr\left(-\frac{1}{2}\boldsymbol{\Sigma}_2^{-1}\boldsymbol{V}\right)\right),
\end{aligned}
\tag{3.39}
$$

from R.49, Remark 3.5.2 and Definition 2.1.1, where $\boldsymbol{\Sigma}_2 \sim W_m\left(\frac{1}{2}\boldsymbol{\Theta}, n+p\right)$. Hence (3.37) follows from (3.38) and (3.39). □

Theorem 3.5.8. *The marginal posterior density function of* $\boldsymbol{\Sigma}$ *for the normal model with prior* (3.34) *is*

$$
\begin{aligned}
& q\left(\boldsymbol{\Sigma}|\boldsymbol{X}\right) \\
&= \frac{|\boldsymbol{\Theta}|^{-\frac{n+p}{2}}|\boldsymbol{\Sigma}|^{\frac{n-n_1p+p}{2}-\frac{m+1}{2}}\ {}_vF_w\left(\boldsymbol{a}^*,\boldsymbol{b}^*;ctr\left(\boldsymbol{\Theta}^{-1}\boldsymbol{\Sigma}\right)\right)etr\left(-\frac{1}{2}\boldsymbol{\Sigma}^{-1}\boldsymbol{V}\right)etr\left(-\boldsymbol{\Theta}^{-1}\boldsymbol{\Sigma}\right)}{\Gamma_m(\frac{n+p}{2})E_{\boldsymbol{\Sigma}_2}\left(|\boldsymbol{\Sigma}_2|^{-\frac{n_1p}{2}}\ {}_vF_w(\boldsymbol{a}^*,\boldsymbol{b}^*;ctr(\boldsymbol{\Theta}^{-1}\boldsymbol{\Sigma}_2))\,etr\left(-\frac{1}{2}\boldsymbol{\Sigma}_2^{-1}\boldsymbol{V}\right)\right)},
\end{aligned}
\tag{3.40}
$$

where $\boldsymbol{\Sigma}_2 \sim W_m\left(\frac{1}{2}\boldsymbol{\Theta}, n+p\right)$ *and* $\boldsymbol{V}$ *as defined in* (2.7).

Proof. From (3.35) and the R.49 the marginal posterior density of $\mathbf{\Sigma}$ is given by

$$\begin{aligned} q(\mathbf{\Sigma}|\mathbf{X}) &\propto |\mathbf{\Sigma}|^{\frac{n-n_1p}{2}-\frac{m+1}{2}} {}_vF_w\left(\mathbf{a}^*,\mathbf{b}^*;ctr\left(\mathbf{\Theta}^{-1}\mathbf{\Sigma}\right)\right) \\ &\quad \times etr\left(-\mathbf{\Theta}^{-1}\mathbf{\Sigma}\right) etr\left(-\frac{1}{2}\mathbf{\Sigma}^{-1}\mathbf{V}\right) \\ &\quad \times \int_{\mathbb{R}^{m\times p}} etr\left(-\frac{1}{2}\left(\frac{1}{n_1}\mathbf{\Sigma}\right)^{-1}(\overline{\mathbf{X}}-\boldsymbol{\mu})\mathbf{\Omega}^{-1}(\overline{\mathbf{X}}-\boldsymbol{\mu})'\right) d\boldsymbol{\mu} \\ &\propto |\mathbf{\Sigma}|^{\frac{n-n_1p+p}{2}-\frac{m+1}{2}} {}_vF_w\left(\mathbf{a}^*,\mathbf{b}^*;ctr\left(\mathbf{\Theta}^{-1}\mathbf{\Sigma}\right)\right) etr\left(-\frac{1}{2}\mathbf{\Sigma}^{-1}\mathbf{V}\right) etr\left(-\mathbf{\Theta}^{-1}\mathbf{\Sigma}\right). \end{aligned} \tag{3.41}$$

Hence

$$q(\mathbf{\Sigma}|\mathbf{X}) = c_{\mathbf{\Sigma}}^{-1}|\mathbf{\Sigma}|^{\frac{n-n_1p+p}{2}-\frac{m+1}{2}} {}_vF_w\left(\mathbf{a}^*,\mathbf{b}^*;ctr\left(\mathbf{\Theta}^{-1}\mathbf{\Sigma}\right)\right) etr\left(-\frac{1}{2}\mathbf{\Sigma}^{-1}\mathbf{V}\right) etr\left(-\mathbf{\Theta}^{-1}\mathbf{\Sigma}\right),$$

with

$$\begin{aligned} & c_{\mathbf{\Sigma}}^{-1} \\ &= \int_{S_m} |\mathbf{\Sigma}|^{\frac{n-n_1p+p}{2}-\frac{m+1}{2}} {}_vF_w\left(\mathbf{a}^*,\mathbf{b}^*;ctr\left(\mathbf{\Theta}^{-1}\mathbf{\Sigma}\right)\right) etr\left(-\frac{1}{2}\mathbf{\Sigma}^{-1}\mathbf{V}\right) etr\left(-\mathbf{\Theta}^{-1}\mathbf{\Sigma}\right) d\mathbf{\Sigma} \\ &= \Gamma_m\left(\frac{n+p}{2}\right)2^{\frac{(n+p)m}{2}}\left|\frac{1}{2}\mathbf{\Theta}\right|^{\frac{n+p}{2}} E_{\mathbf{\Sigma}_2}\left(|\mathbf{\Sigma}_2|^{-\frac{n_1p}{2}} {}_vF_w\left(\mathbf{a}^*,\mathbf{b}^*;ctr\left(\mathbf{\Theta}^{-1}\mathbf{\Sigma}_2\right)\right) etr\left(-\frac{1}{2}\mathbf{\Sigma}_2^{-1}\mathbf{V}\right)\right) \\ &= \Gamma_m\left(\frac{n+p}{2}\right)|\mathbf{\Theta}|^{\frac{n+p}{2}} E_{\mathbf{\Sigma}_2}\left(|\mathbf{\Sigma}_2|^{-\frac{n_1p}{2}} {}_vF_w\left(\mathbf{a}^*,\mathbf{b}^*;ctr\left(\mathbf{\Theta}^{-1}\mathbf{\Sigma}_2\right)\right) etr\left(-\frac{1}{2}\mathbf{\Sigma}_2^{-1}\mathbf{V}\right)\right) \end{aligned} \tag{3.42}$$

from Remark 3.5.2 and Definition 2.1.1, where $\mathbf{\Sigma}_2 \sim W_m\left(\frac{1}{2}\mathbf{\Theta}, n+p\right)$. Therefore, (3.40) follows from (3.41) and (3.42). □

Theorem 3.5.9. *The Bayes estimator of $\boldsymbol{\mu}$ for the normal model with prior (3.34) is*

$$\widehat{\boldsymbol{\mu}}_{B,HW} = \overline{\mathbf{X}}. \tag{3.43}$$

Proof. The Bayes estimator of $\boldsymbol{\mu}$ is

$$\widehat{\boldsymbol{\mu}}_{B,HW} = E[\boldsymbol{\mu}|\mathbf{X}]$$

The expected value of $\boldsymbol{\mu} - \boldsymbol{X}$ is from (3.37) as follows

$$\begin{aligned}
&E[\boldsymbol{\mu} - \overline{\boldsymbol{X}}|\boldsymbol{X}] \\
= &\int_{\mathbb{R}^{m\times p}} (\boldsymbol{\mu} - \boldsymbol{X})\, q(\boldsymbol{\mu}|\boldsymbol{X})\, d\boldsymbol{\mu} \\
= &\frac{\Gamma_m(\frac{n}{2}) 2^{\frac{nm}{2}} |\boldsymbol{\Omega}|^{-\frac{m}{2}} |\boldsymbol{\Theta}|^{-\frac{n+p}{2}}}{\left(\frac{2\pi}{n_1}\right)^{\frac{mp}{2}} \Gamma_m(\frac{n+p}{2}) E_{\boldsymbol{\Sigma}_2}\left(|\boldsymbol{\Sigma}_2|^{-\frac{n_1 p}{2}}\, {}_vF_w\left(\boldsymbol{a}^*, \boldsymbol{b}^*; ctr\left(\boldsymbol{\Theta}^{-1}\boldsymbol{\Sigma}_2\right)\right) etr\left(-\frac{1}{2}\boldsymbol{\Sigma}_2^{-1}\boldsymbol{V}\right)\right)} \\
&\times \int_{\mathbb{R}^{m\times p}} (\boldsymbol{\mu} - \overline{\boldsymbol{X}}) E_{\boldsymbol{\Sigma}_1}(|\boldsymbol{\Sigma}_1|^{-\frac{n_1 p}{2}} etr\left(-\frac{1}{2}\boldsymbol{\Sigma}_1^{-1}\left(\boldsymbol{V} + n_1(\overline{\boldsymbol{X}} - \boldsymbol{\mu})\boldsymbol{\Omega}^{-1}(\overline{\boldsymbol{X}} - \boldsymbol{\mu})'\right)\right) \\
&\times {}_vF_w\left(\boldsymbol{a}^*, \boldsymbol{b}^*; ctr\left(\boldsymbol{\Theta}^{-1}\boldsymbol{\Sigma}_1\right)\right)) d\boldsymbol{\mu},
\end{aligned}$$

with $\boldsymbol{\Sigma}_1 \sim W_m\left(\frac{1}{2}\boldsymbol{\Theta}, n\right)$ and $\boldsymbol{\Sigma}_2 \sim W_m\left(\frac{1}{2}\boldsymbol{\Theta}, n+p\right)$. It is quite clear that the integrand is an odd function and therefore

$$\begin{aligned}
0 \;=\; &\int_{\mathbb{R}^{m\times p}} (\boldsymbol{\mu} - \overline{\boldsymbol{X}}) E_{\boldsymbol{\Sigma}_1}(|\boldsymbol{\Sigma}_1|^{-\frac{n_1 p}{2}} etr\left(-\frac{1}{2}\boldsymbol{\Sigma}_1^{-1}\left(\boldsymbol{V} + n_1(\overline{\boldsymbol{X}} - \boldsymbol{\mu})\boldsymbol{\Omega}^{-1}(\overline{\boldsymbol{X}} - \boldsymbol{\mu})'\right)\right) \\
&\times {}_vF_w\left(\boldsymbol{a}^*, \boldsymbol{b}^*; ctr\left(\boldsymbol{\Theta}^{-1}\boldsymbol{\Sigma}_1\right)\right)) d\boldsymbol{\mu}.
\end{aligned}$$

Therefore $E[\boldsymbol{\mu} - \overline{\boldsymbol{X}}|\boldsymbol{X}] = 0$ and hence $E[\boldsymbol{\mu}|\boldsymbol{X}] = \widehat{\boldsymbol{\mu}}_{B,HW} = \overline{\boldsymbol{X}}$. □

Theorem 3.5.10. *The Bayes estimator of $|\boldsymbol{\Sigma}|^r$ for the normal model with prior (3.34) is*

$$\widehat{|\boldsymbol{\Sigma}|^r}_{B,HW} = \frac{\Gamma_m(\frac{n+p+2r}{2}) E_{\boldsymbol{\Sigma}_3}\left(|\boldsymbol{\Sigma}_3|^{-\frac{n_1 p}{2}}\, {}_vF_w\left(\boldsymbol{a}^*, \boldsymbol{b}^*; ctr\left(\boldsymbol{\Theta}^{-1}\boldsymbol{\Sigma}_3\right)\right) etr\left(-\frac{1}{2}\boldsymbol{\Sigma}_3^{-1}\boldsymbol{V}\right)\right) |\boldsymbol{\Theta}|^r}{\Gamma_m(\frac{n+p}{2}) E_{\boldsymbol{\Sigma}_2}\left(|\boldsymbol{\Sigma}_2|^{-\frac{n_1 p}{2}}\, {}_vF_w\left(\boldsymbol{a}^*, \boldsymbol{b}^*; ctr\left(\boldsymbol{\Theta}^{-1}\boldsymbol{\Sigma}_2\right)\right) etr\left(-\frac{1}{2}\boldsymbol{\Sigma}_2^{-1}\boldsymbol{V}\right)\right)} \tag{3.44}$$

where $\boldsymbol{\Sigma}_2 \sim W_m(\frac{1}{2}\boldsymbol{\Theta}, n+p)$, $\boldsymbol{\Sigma}_3 \sim W_m(\frac{1}{2}\boldsymbol{\Theta}, n+p+2r)$ and $\boldsymbol{V}$ as defined in (2.7).

Proof. From (3.40), using Remark 3.5.2 and Definition 2.1.1, the Bayes estimator of $|\boldsymbol{\Sigma}|^r$ is

$$\begin{aligned}
E[|\boldsymbol{\Sigma}|^r|\boldsymbol{X}] \;=\; &\frac{|\boldsymbol{\Theta}|^{-\frac{n+p}{2}}}{\Gamma_m(\frac{n+p}{2}) E_{\boldsymbol{\Sigma}_2}\left(|\boldsymbol{\Sigma}_2|^{-\frac{n_1 p}{2}}\, {}_vF_w\left(\boldsymbol{a}^*, \boldsymbol{b}^*; ctr\left(\boldsymbol{\Theta}^{-1}\boldsymbol{\Sigma}_2\right)\right) etr\left(-\frac{1}{2}\boldsymbol{\Sigma}_2^{-1}\boldsymbol{V}\right)\right)} \\
&\times \int_{S_m} |\boldsymbol{\Sigma}|^{\frac{n-n_1p+p+2r}{2} - \frac{m+1}{2}}\, {}_vF_w\left(\boldsymbol{a}^*, \boldsymbol{b}^*; ctr\left(\boldsymbol{\Theta}^{-1}\boldsymbol{\Sigma}\right)\right) \\
&\times etr\left(-\frac{1}{2}\boldsymbol{\Sigma}^{-1}\boldsymbol{V}\right) etr\left(-\boldsymbol{\Theta}^{-1}\boldsymbol{\Sigma}\right) d\boldsymbol{\Sigma} \\
= \;&\frac{\Gamma_m(\frac{n+p+2r}{2}) E_{\boldsymbol{\Sigma}_3}\left(|\boldsymbol{\Sigma}_3|^{-\frac{n_1 p}{2}}\, {}_vF_w\left(\boldsymbol{a}^*, \boldsymbol{b}^*; ctr\left(\boldsymbol{\Theta}^{-1}\boldsymbol{\Sigma}_3\right)\right) etr\left(-\frac{1}{2}\boldsymbol{\Sigma}_3^{-1}\boldsymbol{V}\right)\right) |\boldsymbol{\Theta}|^r}{\Gamma_m(\frac{n+p}{2}) E_{\boldsymbol{\Sigma}_2}\left(|\boldsymbol{\Sigma}_2|^{-\frac{n_1 p}{2}}\, {}_vF_w\left(\boldsymbol{a}^*, \boldsymbol{b}^*; ctr\left(\boldsymbol{\Theta}^{-1}\boldsymbol{\Sigma}_2\right)\right) etr\left(-\frac{1}{2}\boldsymbol{\Sigma}_2^{-1}\boldsymbol{V}\right)\right)},
\end{aligned}$$

where $\boldsymbol{\Sigma}_2 \sim W_m(\frac{1}{2}\boldsymbol{\Theta}, n+p)$ and $\boldsymbol{\Sigma}_3 \sim W_m(\frac{1}{2}\boldsymbol{\Theta}, n+p+2r)$. □

3.6 Numerical study

The newly developed results will be applied in a simulation and real context, and compared with the sample estimates. The priors under consideration are the hypergeometric gamma prior (see (3.19) with $v = w = 1$) and the hypergeometric Wishart prior (see (3.17) with $v = w = 1$) as special cases of the HWGD (see Definition 3.2.1). These priors are studied for the first time in literature.

To obtain the computational results, the algorithms developed in Section 2.5 can be adjusted using the results derived in Section 3.5.

3.6.1 Univariate case

In this section the hypergeometric gamma distribution (see (3.19) with $v = w = 1$) is applied as a prior for the variance of the univariate normal model.

3.6.1.1. Algorithm

In this section, the analytical posterior density functions (3.25) and (3.28), as well as the Bayes estimators (3.31) and (3.32) for the hypergeometric gamma prior (3.19) with $v = w = 1$ are evaluated, using Gibbs sampling.

For this purpose, a Gibbs sampling algorithm with a Metropolis-Hastings algorithm is developed, to simulate samples from (3.37) and (3.40) as follows:

Algorithm 5

1. Initialize μ_0 and σ_0^2
2. Repeat the following steps for $t = 1, ..., 100\,000$ times:
   (a) Generate $\mu_t \sim N(\overline{X}, \frac{1}{n_1}\sigma_{t-1}^2)$
   (b) Calculate $A_t = \sum_{i=1}^{n_1}(X_i - \mu_t)^2$
   (c) Metropolis-Hastings algorithm:
      i. Generate the random numbers $\sigma^{2*} \sim G(\frac{n}{2}, \varphi^{-1})$
      ii. If $min\left(\frac{f^*[\sigma^{2*}|\mu_t]}{f^*[\sigma_{t-1}^2|\mu_t]}, 1\right) > u$ where u is a random uniform$(0,1)$ variate, then $\sigma_t^2 = \sigma^{2*}$ else $\sigma_t^2 = \sigma_{t-1}^2$, with $f^*[\sigma^2|\mu_t] \propto (\sigma^2)^{\frac{n}{2}-\frac{n_1}{2}-1} {}_vF_w\left(a_1^*; b_1^*; c\varphi^{-1}\sigma^2\right) exp\left(-\varphi^{-1}\sigma^2\right) exp\left(-\frac{1}{2\sigma^2}A_t\right)$ from (3.24).
3. Burn-in the first couple of observations, i.e. the posterior observations are $\mu_{1000}, ..., \mu_{100\,000}$ and $\sigma_{1000}, ..., \sigma_{100\,000}$.

A sample of 1000 posterior values are simulated with a burn-in of 10%. The empirical histograms are obtained from these samples and can be observed in Figure 3.5. The density functions (3.25) and (3.28), are calculated over a grid of μ and σ^2 values and displayed in Figure 3.5 (see code in Appendix C.3).

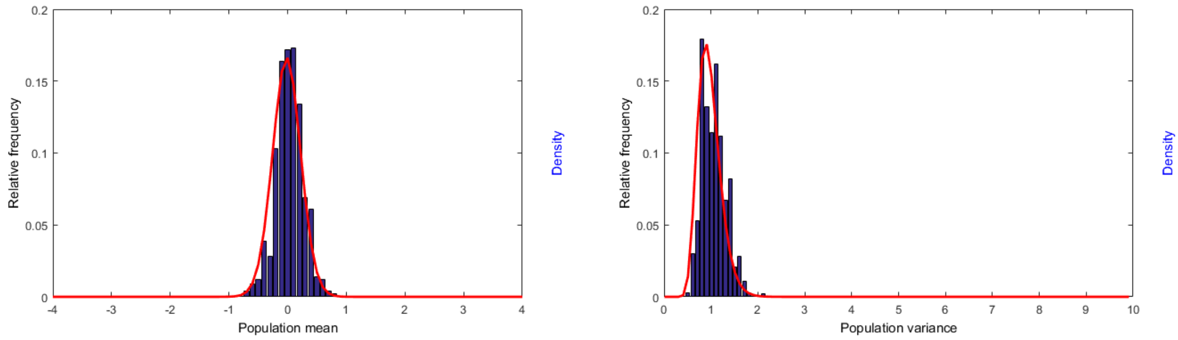


FIGURE 3.5: Analytical posterior density functions (3.25) (Left) and (3.28) (Right) and histograms based on Gibbs sampling for μ (Left) and σ^2 (Right)

Remark 3.6.1. *The computational time to construct the histograms is 812.56 seconds, whereas the curves using the proposed analytical expressions is 68.52 seconds.*

It is clear that the analytical posterior density functions (3.25) and (3.28) are supported by results obtained from the Gibbs sampling method. The advantage of the proposed methodology in Remark 3.5.2, is that no specialized software or advanced programming abilities are needed. This approach provides an alternative to approximations such as Gibbs sampling (see R.62) and INLA (see Rue, Martino, and Chopin (2009)), since the implementation is computationally efficient, especially for complicated prior structures such as the case under consideration.

3.6.1.2. The choice of hyperparameter values

The choice of the prior parameters is an important aspect of any subjective Bayesian analysis. An intuitive methodology is proposed to choose the parameter values of this specific prior structure, similar to Duran and Booker (1988). Assuming the hypergeometric gamma distribution ((3.19) with $v = w = 1$) as the prior for σ^2, gives the prior density function

$$\pi\left(\sigma^2\right) = \frac{\varphi^{-\frac{n}{2}}\left(\sigma^2\right)^{\frac{n}{2}-1}}{\Gamma(\frac{n}{2})\, {}_2F_1\left(\frac{n}{2}, a_1^*; b_1^*; c\right)} exp\left(-\varphi^{-1}\sigma^2\right)\, {}_1F_1\left(a_1^*; b_1^*; c\varphi^{-1}\sigma^2\right), \tag{3.45}$$

where ${}_1F_1(.)$ is the confluent hypergeometric function (see R.32) and ${}_2F_1(.)$ is the Gauss hypergeometric function with $|c| < 1$ (see R.33), with first moment

$$E[\sigma^2] = \frac{\varphi\Gamma(\frac{n}{2}+1)\, {}_2F_1\left(\frac{n}{2}+1, a_1^*; b_1^*; c\right)}{\Gamma(\frac{n}{2})\, {}_2F_1\left(\frac{n}{2}, a_1^*; b_1^*; c\right)},$$

from (3.20). Consider the the following prior belief

$$H_0 : \sigma^2 = \sigma_0^2.$$

Setting the expected value of σ^2 equal to σ_0^2 yields

$$\begin{aligned}\sigma_0^2 &= E[\sigma^2] \\ &= \frac{\varphi\Gamma(\frac{n}{2}+1)\ {}_2F_1\left(\frac{n}{2}+1, a_1^*; b_1^*; c\right)}{\Gamma(\frac{n}{2})\ {}_2F_1\left(\frac{n}{2}, a_1^*; b_1^*; c\right)} \\ &= \frac{(\frac{n}{2})\varphi\ {}_2F_1\left(\frac{n}{2}+1, a_1^*; b_1^*; c\right)}{{}_2F_1\left(\frac{n}{2}, a_1^*; b_1^*; c\right)}. \end{aligned} \tag{3.46}$$

The ratio $\frac{{}_2F_1\left(\frac{n}{2}+1,a_1^*;b_1^*;c\right)}{{}_2F_1\left(\frac{n}{2},a_1^*;b_1^*;c\right)}$ is a slow varying function as can be seen in Figure 3.6 (see code in Appendix C.4).

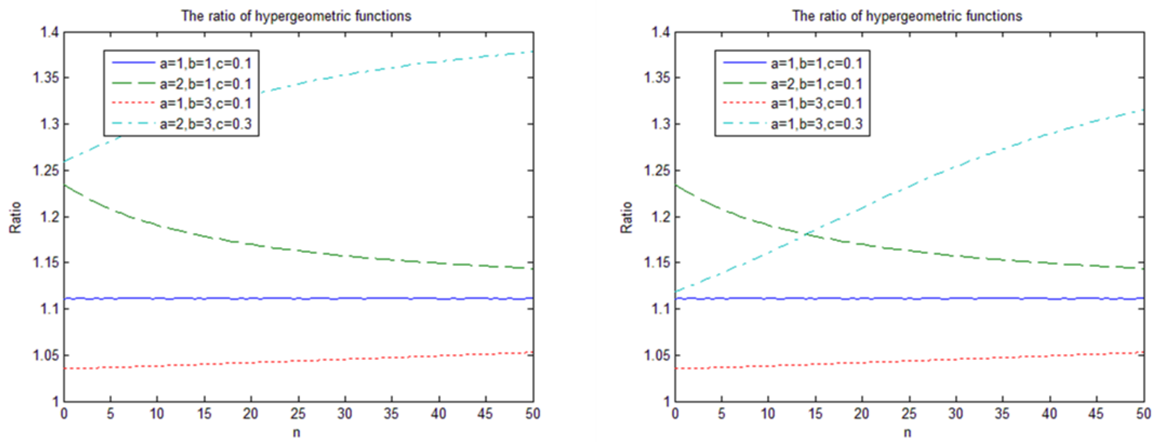


FIGURE 3.6: $\frac{{}_2F_1\left(\frac{n}{2}+1,a_1^*;b_1^*;c\right)}{{}_2F_1\left(\frac{n}{2},a_1^*;b_1^*;c\right)}$ for various parameter values

Hence

$$\sigma_0^2 \approx (\frac{n}{2})\varphi. \tag{3.47}$$

This enables the user to choose φ in an informed manner, based on the choice of n.

3.6.1.3. Simulation study

In this section, a normal sample of size n_1 is simulated with known mean and variance. The newly proposed Bayes estimators ((3.31) and (3.32) with $v = 1$ and $w = 1$) for μ and σ^2, respectively, are calculated as well as the coverage probability.

For the purpose of this study, a normal sample of size 18 is simulated with parameters $\mu = 0$ and $\sigma^2 = 1$. The hyperparameters are chosen according to the above methodology as $n = 19, \varphi = 0.105, a_1^* = 1, b_1^* = 2$ and $c = 0.02$ such that $E(\sigma^2) = 0.9$. The choice of $\sigma_0^2 = 0.9$ not being analytically $\sigma^2 = 1$, illustrates that this prior performs well even under an incorrect prior belief. The sample estimates which are the MLE's are given by

$$\widehat{\mu} = \overline{X} = -0.0166, \qquad \widehat{\sigma^2} = s^2 = 0.8397.$$

The Bayes estimates calculated using the sample, from (3.31) and (3.32), respectively, are

$$\widehat{\mu}_{B,HG} = \overline{X} = -0.0166, \qquad \widehat{\sigma^2}_{B,HG} = 0.9701.$$

It is evident from the comparison of the estimates that the Bayes estimate for σ^2 is closer to the target parameter value than the MLE. Figure 3.7 displays the analytical posterior density functions as in (3.25) and (3.28) with $v = 1$ and $w = 1$ (see code in Appendix C.5).

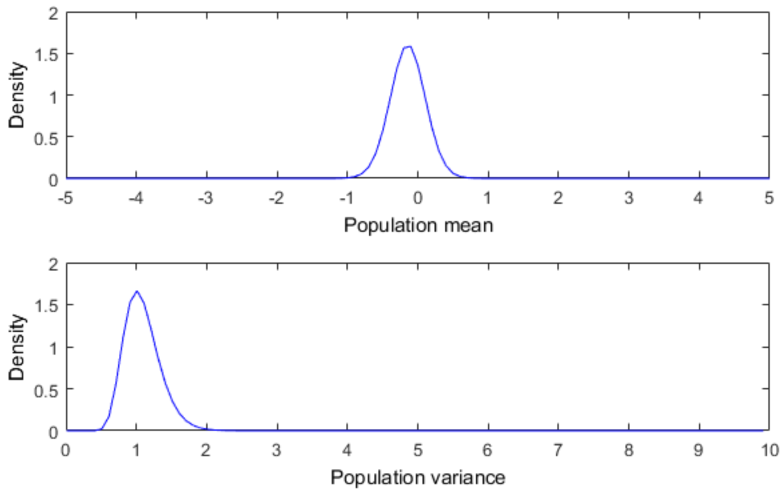


FIGURE 3.7: Analytical posterior density functions of μ (3.25) and σ^2 (3.28)

To calculate the coverage probability, 1000 samples were simulated from this underlying model and in each case the 95% credible interval was calculated so that the coverage probability can be calculated.

The coverage probability of this prior is 97% with a median credible interval width of 0.8.

3.6.1.4. Forestry dataset

A forester wishes to estimate the volume of merchantable timber in a population of trees, based on a sample (Grafen and Hails 2002, p.28).

FIGURE 3.8: A forest of trees (left) and the circumference measurement (right) needed for the calculation of the merchantable timber

To this purpose he selected 31 trees and measured their volume. Since this is a small sample, most probably not random due to location constraints, we propose a Bayesian approach with a hypergeometric gamma prior for the estimation of the variance and an improper objective prior for the mean. The data is found to be non-normal and a log-transformation corrected this. Let X be the volume of the tree and $Y = log(X)$ be the log-transformed variable, then $Y \sim N(\mu, \sigma^2)$.

Prior (3.22) with $v = w = 1$ is assumed, and the prior information is specified as $E[\sigma^2] = 0.3$ based on expert information. The hyperparameters are chosen as $\varphi = 0.01875, c = 0.01, n = 32, a_1^* = 1, b_1^* = 1$, according to (3.47).

The resulting sample estimate of σ^2 is 0.277 and the Bayes estimate is 0.2933 for the hypergeometric gamma prior with credible interval $(0.21; 0.385)$. The likelihood, prior and posterior density functions of σ^2 for this dataset are displayed in Figure 3.9 (see code in Appendix C.6). Since the posterior is more concentrated than the prior, this prior is suitable for the dataset (see Lesaffre and Lawson (2012)).

A comparative study using different priors for this dataset, is done in Section 5.2.2.

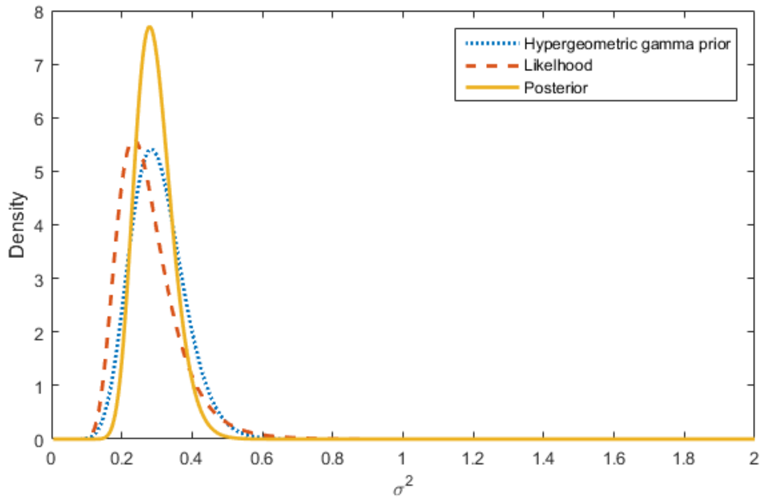


FIGURE 3.9: Likelihood, prior and analytical posterior density functions of σ^2

3.6.2 Multivariate case

In this section, the results of the preceding section will be applied to a simulated and real dataset to illustrate the novelty of the hypergeometric Wishart distribution (3.17) with $v = w = 1$ as a prior distribution for the matrix variate normal model.

3.6.2.1. Algorithm

The algorithms proposed in Section 2.5 are adjusted to develop a Gibbs sampling algorithm for the simulation of posterior samples of $\boldsymbol{\mu}$ and $\boldsymbol{\Sigma}$ under the prior $\boldsymbol{\Sigma} \sim HW_m(\boldsymbol{\Phi}, a_1^*, b_1^*, c, n)$ (see (3.17)). The algorithm is as follows:

Algorithm 6

1. Initialize $\boldsymbol{\mu}_0$ and $\boldsymbol{\Sigma}_0$.
2. Repeat the following steps for $t = 1, ..., 100\,000$ times:
   (a) Generate $\boldsymbol{\mu}_t \sim N_{m,p}(\overline{\boldsymbol{X}}, \frac{1}{n_1}\boldsymbol{\Sigma}_{t-1} \otimes \boldsymbol{\Omega})$.
   (b) Calculate $\boldsymbol{A}_t = \boldsymbol{V} + n_1(\overline{\boldsymbol{X}} - \boldsymbol{\mu}_t)(\overline{\boldsymbol{X}} - \boldsymbol{\mu}_t)'$.
   (c) Metropolis-Hastings algorithm:
      i. Generate the random matrices $\boldsymbol{\Sigma}_1^* \sim W_m(\boldsymbol{\Theta}_1, m_4)$ and $\boldsymbol{\Sigma}_2^* \sim W_m^{-1}(\boldsymbol{\Theta}_2, m_4^*)$ such that $E[\boldsymbol{\Sigma}_1^*] = cE[\boldsymbol{\Sigma}_2^*]$.
      ii. Calculate $\boldsymbol{\Sigma}^* = w\boldsymbol{\Sigma}_1^* + (1-w)\boldsymbol{\Sigma}_2^*$ for some $0 < w < 1$.
      iii. If $min\left(\frac{f^*[\boldsymbol{\Sigma}^*|\boldsymbol{\mu}_t]}{f^*[\boldsymbol{\Sigma}_{t-1}|\boldsymbol{\mu}_t]}, 1\right) > u$ where u is a random uniform$(0,1)$ variate, then $\boldsymbol{\Sigma}_t = \boldsymbol{\Sigma}^*$ else $\boldsymbol{\Sigma}_t = \boldsymbol{\Sigma}_{t-1}$, with
      $f^*[\boldsymbol{\Sigma}|\boldsymbol{\mu}_t] \propto |\boldsymbol{\Sigma}|^{-0.5(n-n_1p-m-1)}\ {}_vF_w\left(\boldsymbol{a}^*; \boldsymbol{b}^*; ctr\left(\boldsymbol{\Theta}^{-1}\boldsymbol{\Sigma}\right)\right) etr\left(-\frac{1}{2}\boldsymbol{\Sigma}\boldsymbol{\Theta}^{-1}\right) etr\left(-\frac{1}{2}\boldsymbol{\Sigma}^{-1}\boldsymbol{A}_t\right)$
      from (3.36).
3. Burn-in the first couple of observations, i.e. the posterior observations are $\boldsymbol{\mu}_{1000}, ..., \boldsymbol{\mu}_{100\,000}$ and $\boldsymbol{\Sigma}_{1000}, ..., \boldsymbol{\Sigma}_{100\,000}$.

3.5.2.2. Simulation study

A sample of size n_1 is simulated from a multivariate normal distribution ($p = 1$) (see R.45) with a zero mean vector and identity covariance matrix, i.e. $\boldsymbol{X} \sim N_m(\boldsymbol{0}, \boldsymbol{I}_m)$.

The prior distribution is assumed as $\boldsymbol{\Sigma} \sim HW_m(\boldsymbol{\Phi}, a_1, b_1, c, n)$ (see (3.17)), with $\boldsymbol{\Phi} = 4\boldsymbol{I}_m, m = 3, a_1^* = 1, b_1^* = 2,\ c = 1, n_1 = 20, n = m_4 = m_4^* = 3$. The initial values are chosen as $\boldsymbol{\mu}_0 = [0.5\ \ 0.5\ \ 0.5]$ and $\boldsymbol{\Sigma}_0 = 1.1\boldsymbol{I}_m$.

A posterior sample of size 10 000 with a burn-in of 10%, is simulated using Algorithm 6 and the convergence of the chain is illustrated in Figure 3.10.

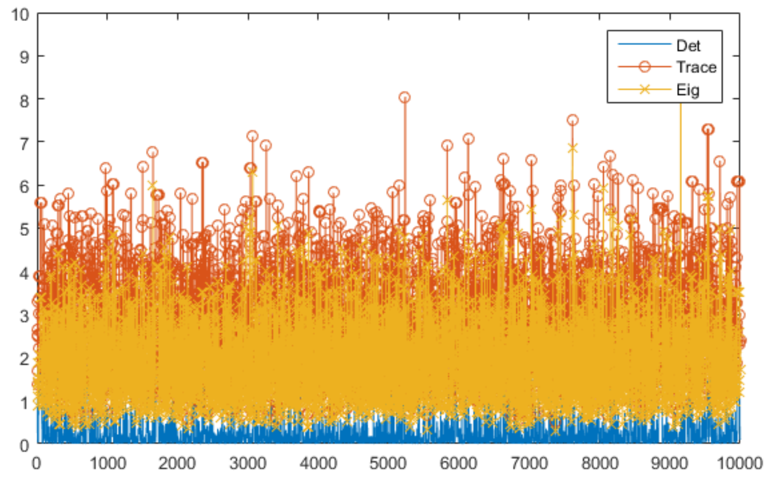


FIGURE 3.10: Convergence measures of the posterior sample simulated using Algorithm 6

The Bayes estimate calculated for $\mathbf{\Sigma}$ under this prior using Algorithm 6 (see code in Appendix C.7) as well as the MLE are

$$\widehat{\mathbf{\Sigma}}_{MLE} = \begin{bmatrix} 3.089 & 1.852 & -0.669 \\ 1.852 & 4.158 & 2.612 \\ -0.669 & 2.612 & 7.191 \end{bmatrix}, \widehat{\mathbf{\Sigma}}_{B,HW} = \begin{bmatrix} 0.9517 & 0.1143 & -0.0178 \\ 0.1143 & 0.9268 & 0.0775 \\ -0.0178 & 0.0775 & 1.054 \end{bmatrix}$$

The above estimates are obtained for one posterior sample. The Frobenius norm of the errors (see Remark 2.5.4) are calculated for each estimate and given in Table 3.2. The hypergeometric Wishart prior estimate results in the smallest Frobenius norm of the error and is hence preferred to the MLE.

Frobenius norm	Value
$\|\|\widehat{\mathbf{\Sigma}}_{MLE} - \mathbf{\Sigma}\|\|_F$	0.5241
$\|\|\widehat{\mathbf{\Sigma}}_{HW} - \mathbf{\Sigma}\|\|_F$	0.1151

TABLE 3.2: Frobenius norm of the estimates calculated from the simulated sample

3.6.2.3. 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, which is a very time-consuming task. Other measurements, which can be obtained easily, can give an indication of the number of rings.

FIGURE 3.11: An abalone with meat (left) and without meat (right) where the rings are exposed

The abalone dataset is from Lichman (2013) and contains 8 variables and has a sample size of 4177. The six continuous variables are used: 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 (Royston 1983). Hence, $\boldsymbol{X}_{1\times 6} \sim N(\boldsymbol{\mu}_{1\times 6}, \boldsymbol{\Sigma}_{6\times 6})$. It is essential to estimate $\boldsymbol{\Sigma}$ accurately to use this information to estimate the age of the abalone.

Assume the prior (3.34), $v = w = 1$, with $\boldsymbol{\Phi} = 0.5\boldsymbol{S}_1, m = 6, a_1^* = 1, b_1^* = 2, c = 1, n = m_4 = m_4^* = 6$ where $\boldsymbol{S}_1$ is the sample covariance matrix of the available data excluding the 20 observations in the chosen subsample.

The Bayes estimate calculated for $\boldsymbol{\Sigma}$ under this prior using Algorithm 6 (see code in Appendix C.8), as well as the MLE are

$$\widehat{\boldsymbol{\Sigma}}_{MLE} = \begin{bmatrix} 0.014 & 0.012 & 0.004 & 0.054 & 0.015 & 0.012 \\ 0.012 & 0.01 & 0.003 & 0.045 & 0.013 & 0.01 \\ 0.004 & 0.003 & 0.002 & 0.017 & 0.005 & 0.004 \\ 0.054 & 0.045 & 0.017 & 0.24 & 0.065 & 0.052 \\ 0.015 & 0.013 & 0.005 & 0.065 & 0.019 & 0.014 \\ 0.012 & 0.01 & 0.004 & 0.052 & 0.014 & 0.012 \end{bmatrix}$$

$$\widehat{\boldsymbol{\Sigma}}_{B,HW} = \begin{bmatrix} 0.011 & 0.009 & 0.003 & 0.045 & 0.012 & 0.01 \\ 0.009 & 0.008 & 0.003 & 0.038 & 0.01 & 0.008 \\ 0.003 & 0.003 & 0.001 & 0.013 & 0.004 & 0.003 \\ 0.045 & 0.038 & 0.013 & 0.209 & 0.055 & 0.045 \\ 0.012 & 0.01 & 0.004 & 0.055 & 0.015 & 0.012 \\ 0.01 & 0.008 & 0.003 & 0.045 & 0.012 & 0.011 \end{bmatrix}$$

The convergence of Algorithm 6 is clear from Figure 3.12.

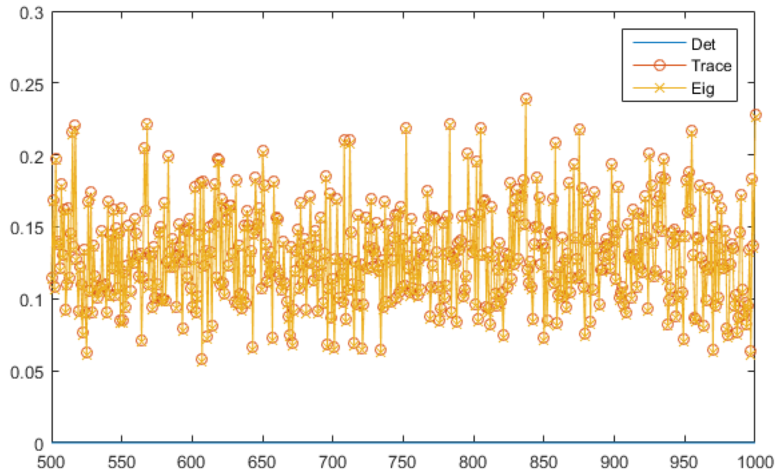


FIGURE 3.12: Convergence measures of the posterior sample (updates 500-1000) simulated using Algorithm 6

A comparative study using different priors for this dataset, is done in Section 5.3.2.

3.7 Conclusion

Since the Wishart distribution plays such an important role in multivariate analysis, a more general framework is developed, which contains the Wishart distribution, for wider applicability as illustrated in Sections 3.5 and 3.6. This leads to the definition of the hypergeometric Wishart generator distribution. The main contributions in this chapter are summarized as follows:

Theoretical contributions

- The hypergeometric Wishart generator distribution (HWGD) was defined as a generator-type distribution which gives rise to amongst others, the hypergeometric Wishart and Wishart distributions.
- Various properties of the HWGD were derived and graphically illustrated.
- Univariate and multivariate Bayesian analysis of the normal model were done with a special case of the HWGD as the prior, which has not been considered before, using an innovative approach circumventing the use of sampling algorithms like Gibbs sampling.

Computational contributions

- New algorithms were developed for the application of a special case of the HGWD as the prior for the univariate and matrix variate normal models.

Chapter 4

Weighted-type Wishart distributions

In this chapter, the weighted-type Wishart distributions are proposed. The methodology involves coupling the density function of the Wishart distribution with a Borel measurable function as a weight. Three different weight functions are proposed by considering trace and determinant operators on matrices. The characteristics of the proposed weighted-type Wishart distributions are derived and used to illustrate the enrichment of this approach. The significance of these new results is demonstrated by applying a special case of the weighted-type I Wishart distribution as a prior in the Bayesian analysis of the univariate and multivariate normal models. Algorithms are developed for the computational use of this prior. Arashi, Bekker, and Van Niekerk (2017) resulted from this chapter.

4.1 Introduction

To generalize the Wishart distribution (see Chapter 2) different from what exists in literature, a weighted-type Wishart distribution, making use of the mathematical mechanism frequently used in proposing weighted-type distributions, is developed in this chapter. The focus is on three general types of weighted-type Wishart distributions, originating from three different types of weight functions. Some properties of each type are derived and the weighted-type I Wishart distribution is applied as a prior for the covariance matrix of the matrix variate normal model.

In Section 4.3, the weighted-type Wishart distribution that originates from a weight function of the trace operator and some of its important properties, are proposed and illustrated. The enrichment of this approach is illustrated by the graphical display of the joint density function of the eigenvalues for certain cases. In Sections 4.4 and 4.5, the weighted-type Wishart distributions emanating from weight functions of determinant, and a combination of trace and determinant forms, are proposed with their properties. The significance of these new models are demonstrated in Section 4.6 by assuming special cases as priors for the underlying univariate and multivariate normal model.

4.2 Construction methodology

In this section the construction methodology of the weighted-type Wishart distribution is considered. The building block of the construction is an extension of the mathematical formulation of univariate (see Rao (1965), Nanda and Jain (1999)) and multivariate (see Jain and Nanda (1995), Navarro, Ruiz, and Aguila (2006)) weighted-type distributions to matrix variate weighted-type distributions.

Let $f(x;\sigma^2)$ be the main density function which is imposed by a scalar weight function $h(x;\psi)$ (not necessarily positive), then the weighted-type distribution proposed by Rao (1965) has the form

$$g(x;\boldsymbol{\theta}) = \frac{h(x;\phi)f(x;\sigma^2)}{E[h(X;\phi)]}, \quad \boldsymbol{\theta} = (\sigma^2,\phi), \tag{4.1}$$

where the expectation $E[.]$ is taken over the same probability measure as $f(.)$. The parameter ϕ can be seen as an *enriching parameter*.

In this chapter, this concept is extended to the matrix variate case by using a matrix variate density function for $f(.)$ in (4.1). Further, the parameter space can be multi-dimensional although the weight function $h(.)$ should remain of a scalar form. Using (4.1) as departure, matrix variate weighted-type distributions are defined, from where new matrix variate distributions originate.

Consider a random matrix variate $\boldsymbol{X} \in S_m$ having a density function $f(.;\boldsymbol{\Psi})$ with parameter $\boldsymbol{\Psi}$. Matrix variate distributions, with density functions $g(.;\boldsymbol{\Theta})$[1], are constructed where $\boldsymbol{\Theta} = (\boldsymbol{\Psi},\boldsymbol{\Phi})$, and $\boldsymbol{\Phi} \in S_m$ is an enrichment parameter, by utilizing one of the following mechanisms:

1. (Loading with a weight of trace form)
$$g(\boldsymbol{X};\boldsymbol{\Theta}) = C_1 h_1(tr[\boldsymbol{X}\boldsymbol{\Phi}])f(\boldsymbol{X};\boldsymbol{\Psi}), \quad \boldsymbol{\Theta} = (\boldsymbol{\Psi},\boldsymbol{\Phi}). \tag{4.2}$$

2. (Loading with a weight of determinant form)
$$g(\boldsymbol{X};\boldsymbol{\Theta}) = C_2 h_2(|\boldsymbol{X}\boldsymbol{\Phi}|)f(\boldsymbol{X};\boldsymbol{\Psi}), \quad \boldsymbol{\Theta} = (\boldsymbol{\Psi},\boldsymbol{\Phi}). \tag{4.3}$$

3. (Loading with a mixture of weights of trace and determinant forms)
$$g(\boldsymbol{X};\boldsymbol{\Theta}) = C_3 h_1(tr[\boldsymbol{X}\boldsymbol{\Phi}_1])h_2(|\boldsymbol{X}\boldsymbol{\Phi}_2|)f(\boldsymbol{X};\boldsymbol{\Psi}), \quad \boldsymbol{\Theta} = (\boldsymbol{\Psi},\boldsymbol{\Phi}_1,\boldsymbol{\Phi}_2), \tag{4.4}$$

where $h_i(.), i = 1,2$ is a Borel measurable function (weight function) which admits Taylor's series expansion, C_j is a normalizing constant and $f(.)$ can be referred to as the main density function.

In this chapter, $f(.;\boldsymbol{\Psi})$ in (4.2)-(4.4) is considered to be the density function of the Wishart distribution with parameters n and $\boldsymbol{\Sigma}$ given in Definition 2.1.1 and a weight function as given by (4.2)-(4.4) is incorporated. Other density functions can be considered for $f(.;\boldsymbol{\Psi})$ in future research (see Chapter 6).

4.3 Weighted-type I Wishart Distribution

In this section the weighted-type I Wishart distribution is defined according to (4.2) and specific properties of this distribution are derived.

Definition 4.3.1. *The random matrix $\boldsymbol{X} \in S_m$ is said to have a weighted-type I Wishart distribution (W1WD) with parameters $\boldsymbol{\Psi}$, $\boldsymbol{\Phi} \in S_m$ and the weight function $h_1(.)$, if it has the following density function*

$$\begin{aligned} g(\boldsymbol{X};\boldsymbol{\Theta}) &= \frac{h_1(tr[\boldsymbol{X}\boldsymbol{\Phi}])f(\boldsymbol{X};\boldsymbol{\Psi})}{E\left[h_1(tr[\boldsymbol{X}\boldsymbol{\Phi}])\right]} \\ &= c^I_{n,m}(\boldsymbol{\Theta})|\boldsymbol{\Sigma}|^{-\frac{n}{2}}|\boldsymbol{X}|^{\frac{n}{2}-\frac{m+1}{2}} etr\left(-\frac{1}{2}\boldsymbol{\Sigma}^{-1}\boldsymbol{X}\right) h_1(tr[\boldsymbol{X}\boldsymbol{\Phi}]), \quad \boldsymbol{\Theta} = (\boldsymbol{\Psi},\boldsymbol{\Phi}), \end{aligned} \tag{4.5}$$

with

$$\{c^I_{n,m}(\boldsymbol{\Theta})\}^{-1} = \int_{S_m} |\boldsymbol{\Sigma}|^{-\frac{n}{2}}|\boldsymbol{X}|^{\frac{n}{2}-\frac{m+1}{2}} etr\left(-\frac{1}{2}\boldsymbol{\Sigma}^{-1}\boldsymbol{X}\right) h_1(tr[\boldsymbol{X}\boldsymbol{\Phi}])d\boldsymbol{X} \tag{4.6}$$

$$= 2^{\frac{nm}{2}}\Gamma_m\left(\frac{n}{2}\right)\sum_{k=0}^{\infty}\frac{2^k h_1^{(k)}(0)}{k!}\sum_{\kappa}\left(\frac{n}{2}\right)_{\kappa} C_{\kappa}(\boldsymbol{\Phi}\boldsymbol{\Sigma}), \tag{4.7}$$

[1] In this chapter, $g(.;\boldsymbol{\Theta})$ indicates the weighted-type density function with a set of parameters $\boldsymbol{\Theta}$ consisting of the parameters from the main density function as well as the weight function

denoted as $\boldsymbol{X} \sim W_m^I(\boldsymbol{\Sigma},\boldsymbol{\Phi},n)$*, where* $f(\boldsymbol{X};\boldsymbol{\Psi})$ *is the density function of the Wishart distribution,* $(W_m(\boldsymbol{\Sigma},n))$ *(see Definition 2.1.1) i.e.* $\boldsymbol{\Psi}=(\boldsymbol{\Sigma},n)$*,* $n \geq m$*,* $\boldsymbol{\Sigma} \in S_m$ *and* $h_1(.)$ *is a Borel measurable function that admits Taylor's series expansion (see R.7). The parameters are restricted to take those values for which the density function is non-negative.*

Note that using Taylor's series expansion for $h_1(.)$ in (4.5) it follows that

$$\begin{aligned} h_1(tr[\boldsymbol{X\Phi}]) &= \sum_{k=0}^{\infty} \frac{h_1^{(k)}(0)}{k!} tr(\boldsymbol{X\Phi})^k \\ &= \sum_{k=0}^{\infty} \sum_{\kappa} \frac{h_1^{(k)}(0)}{k!} C_\kappa(\boldsymbol{X\Phi}), \end{aligned} \tag{4.8}$$

from R.7 and R.19. Using R.22, R.23 and (4.8) follows that

$$\begin{aligned} E\left[h_1(tr[\boldsymbol{X\Phi}])\right] &= \int_{S_m} h_1(tr[\boldsymbol{X\Phi}]) f(\boldsymbol{X};\boldsymbol{\Sigma}) d\boldsymbol{X} \\ &= \sum_{k=0}^{\infty} \frac{h_1^{(k)}(0)}{k!} \int_{S_m} C_\kappa(\boldsymbol{X\Phi}) f(\boldsymbol{X};\boldsymbol{\Sigma}) d\boldsymbol{X} \\ &= \frac{|\boldsymbol{\Sigma}|^{-\frac{n}{2}}}{2^{\frac{nm}{2}} \Gamma_m\left(\frac{n}{2}\right)} \sum_{k=0}^{\infty} \frac{h_1^{(k)}(0)}{k!} \sum_{\kappa} \int_{S_m} |\boldsymbol{X}|^{\frac{n}{2}-\frac{m+1}{2}} etr\left(-\frac{1}{2}\boldsymbol{\Sigma}^{-1}\boldsymbol{X}\right) C_\kappa(\boldsymbol{X\Phi}) d\boldsymbol{X} \\ &= \frac{|\boldsymbol{\Sigma}|^{-\frac{n}{2}}}{2^{\frac{nm}{2}} \Gamma_m\left(\frac{n}{2}\right)} \sum_{k=0}^{\infty} \frac{2^{\frac{nm}{2}+k} \Gamma_m\left(\frac{n}{2}\right) |\boldsymbol{\Sigma}|^{\frac{n}{2}} h_1^{(k)}(0)}{k!} \sum_{\kappa} \left(\frac{n}{2}\right)_\kappa C_\kappa(\boldsymbol{\Sigma\Phi}) \\ &= \sum_{k=0}^{\infty} \frac{2^k h_1^{(k)}(0)}{k!} \sum_{\kappa} \left(\frac{n}{2}\right)_\kappa C_\kappa(\boldsymbol{\Sigma\Phi}), \end{aligned}$$

with $C_\kappa(.)$ the zonal polynomial corresponding to κ (see R.19).

Remark 4.3.1. *Here some thoughts related to Definition 4.3.1 and (4.5) are considered.*

(1) *For* $h_1(tr[\boldsymbol{X\Phi}]) = etr(\boldsymbol{X\Phi})$ *in (4.5) we obtain an enriched Wishart distribution with scale matrix* $\boldsymbol{\Sigma}^{-1}+\boldsymbol{\Phi}$*.*

(2) *For* $h_1(x\phi) = exp(\phi x)$ *and* $m=1$ *in (4.5) the density function simplifies to*

$$g(x;\theta) = c_n^I(\theta)(\sigma^2)^{-\frac{n}{2}} x^{\frac{n}{2}-1} exp\left(-\left(\frac{1}{2\sigma^2}-\phi\right)x\right) \quad \theta=(\sigma^2,\phi), \tag{4.9}$$

with $c_n^I(\theta) = \frac{\left(\frac{1}{2\sigma^2}-\phi\right)^{\frac{n}{2}}}{\Gamma\left(\frac{n}{2}\right)}$ *which is the density function of a gamma random variable (see R.41) with parameters* $\frac{n}{2}$ *and* $\frac{1}{2\sigma^2}-\phi$*, written as* $X \sim G(\frac{n}{2}, \frac{1}{2\sigma^2}-\phi)$*.*

(3) *For* $h_1(x) = x$ *and* $m=1$ *in (4.5) the density function simplifies to*

$$\begin{aligned} g(x;\theta) &= c_n^I(\theta)(\sigma^2)^{-\frac{n}{2}} x^{\frac{n}{2}-1} exp\left(-\frac{1}{2\sigma^2}x\right) x \\ &= c_n^I(\theta)(\sigma^2)^{-\frac{n}{2}} x^{\frac{n}{2}} exp\left(-\frac{1}{2\sigma^2}x\right) \quad \theta=(\sigma^2,\phi), \end{aligned} \tag{4.10}$$

with $c_n^I(\theta) = \frac{\left(\frac{1}{2\sigma^2}-\phi\right)^{\frac{n}{2}+1}}{\Gamma\left(\frac{n}{2}+1\right)}$, *hence* $X \sim G(\frac{n}{2}+1, \frac{1}{2\sigma^2}-\phi)$. *This is also called the length-biased or size-biased gamma distribution (Patil and Ord 1976).*

The following example illustrates the practical origin of the length-biased gamma distribution.

Example 1. *Consider an exhibition hosted in a venue which can accommodate many people. The organizers of the exhibition are investigating the overall time spent by attendees at the exhibition for those who remain until the end.*

Every minute one hundred attendees arrive at the exhibition and stay for a random time period which follows a $G(\alpha,\beta)$. *The exhibition lasts for a total of* 200 *minutes. At the end the exhibition the organizers will host a competition for those who are still at the exhibition. The total time spent at the exhibition of all the attendees still present after* 200 *minutes will not be distributed according to a gamma distribution anymore, but will be weighted by their time spent at the exhibition. This then leads to the length-biased or weighted gamma distribution* (4.10).

Theoretically, the time spent by the attendees at the exhibition for those who are still at the exhibition after 200 *minutes, will follow a length-biased gamma distribution which is equivalent to a* $G(\alpha+1,\beta)$ *distribution.*

This is illustrated in Figure 4.1 (see code in Appendix D.1); the length-biased gamma distribution fits the data visually better than the original gamma distribution. The maximum likelihood estimates of the parameters are $\widehat{\alpha} = 11.09$ *and* $\widehat{\beta} = 9.95$ *(Funke 2014).*

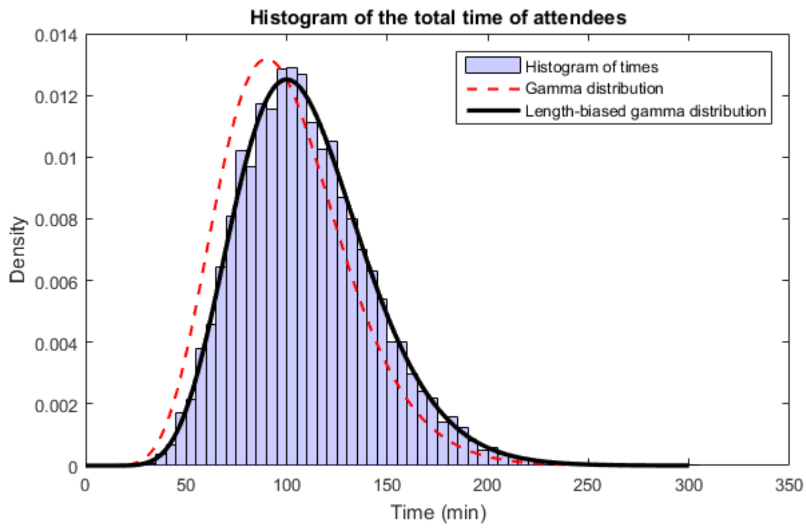


FIGURE 4.1: Histogram of time spent at the exhibition of those attendees still at the exhibition after 200 minutes with the gamma and length-biased gamma densities.

(4) *For* $h_1(tr[\boldsymbol{X\Phi}]) = (1+tr(\boldsymbol{X\Phi}))$ *in* (4.5) *the density function simplifies to*

$$g(\boldsymbol{X};\boldsymbol{\Theta}) = c^I_{n,m}(\boldsymbol{\Theta})|\boldsymbol{\Sigma}|^{-\frac{n}{2}}|\boldsymbol{X}|^{\frac{n}{2}-\frac{m+1}{2}} etr\left(-\frac{1}{2}\boldsymbol{\Sigma}^{-1}\boldsymbol{X}\right)(1+tr(\boldsymbol{X\Phi})), \quad \boldsymbol{\Theta}=(\boldsymbol{\Sigma},\boldsymbol{\Phi}) \quad (4.11)$$

with $c^I_{n,m}(\boldsymbol{\Theta})$ *as in* (4.7), *defined as the Kummer Wishart distribution and denoted as* $\boldsymbol{X} \sim KW_m(\boldsymbol{\Sigma},\boldsymbol{\Phi},n)$.

(5) *For* $h_1(x\phi) = (1+x\phi)^{\omega}$, *where* ω *is a known fixed constant, and* $m=1$ *in (4.5) the density function simplifies to*

$$g(x;\theta) = c_n(\theta)(\sigma^2)^{-\frac{n}{2}}x^{\frac{n}{2}-1}exp\left(-\frac{1}{2\sigma^2}x\right)(1+\phi x)^{\omega}, \quad \theta=(\sigma^2,\phi) \quad (4.12)$$

with $c_n(\theta) = 2^{\frac{n}{2}}\Gamma\left(\frac{n}{2}\right)\sum_{k=0}^{\infty}\frac{(2\phi^2\sigma^2)^k\omega!}{(\omega-k)!k!}\sum_{\kappa}\left(\frac{n}{2}\right)_{\kappa}$. *If* $\phi=1$ *then this is also known as the Kummer gamma or generalized gamma distribution, written as* $KG(\alpha=\frac{n}{2},\beta=\frac{1}{2\sigma^2},\omega)$, *by expanding the term* $(1+x)^{\omega}$ *(see Pauw, Bekker, and Roux (2010)).*

(6) *Various functional forms of* $h_1(.)$ *will be employed and the joint density of their eigenvalues graphically illustrated to show the flexibility achieved by this construction, see Tables 4.1-4.3.*

4.3.1 Properties

In this section some statistical properties of the W1WD (see Definition 4.3.1) are derived.

Theorem 4.3.1. *Let* $\boldsymbol{X} \sim W^I_m(n,\boldsymbol{\Sigma},\boldsymbol{\Phi})$, *then the* r^{th} *moment of* $|\boldsymbol{X}|$ *is given by*

$$E\left(|\boldsymbol{X}|^r\right) = \frac{c^I_{n,m}(\boldsymbol{\Theta})}{c^I_{2(\frac{n}{2}+r),m}(\boldsymbol{\Theta})}|\boldsymbol{\Sigma}|^r,$$

where $c^I_{n,m}(\boldsymbol{\Theta})$ *and* $c^I_{2(\frac{n}{2}+r),m}(\boldsymbol{\Theta})$ *are as in* (4.7).

Proof. Using (4.5) and (4.6), it follows that

$$\begin{aligned} E\left(|\boldsymbol{X}|^r\right) &= c^I_{n,m}(\boldsymbol{\Theta})|\boldsymbol{\Sigma}|^{-\frac{n}{2}}\int_{S_m}|\boldsymbol{X}|^{r+\frac{n}{2}-\frac{m+1}{2}}etr\left(-\frac{1}{2}\boldsymbol{\Sigma}^{-1}\boldsymbol{X}\right)h_1(tr[\boldsymbol{X\Phi}])d\boldsymbol{X} \\ &= c^I_{n,m}(\boldsymbol{\Theta})|\boldsymbol{\Sigma}|^{-\frac{n}{2}}\left(c^I_{2(\frac{n}{2}+r),m}(\boldsymbol{\Theta})|\boldsymbol{\Sigma}|^{-\frac{n}{2}-r}\right)^{-1} \end{aligned}$$

which gives the result. □

Theorem 4.3.2. *Let* $\boldsymbol{X} \sim W^I_m(\boldsymbol{\Sigma},\boldsymbol{\Phi},n)$, *then the chracteristic function of* $\boldsymbol{X}$ *is given by*

$$\psi_{\boldsymbol{X}}(\boldsymbol{T}) = c^I_{n,m}(\boldsymbol{\Theta})d_{n,m}|\boldsymbol{I}_m - 2i\boldsymbol{\Sigma T}|^{-\frac{n}{2}},$$

with $c^I_{n,m}(\boldsymbol{\Theta})$ *as in* (4.7) *and* $d_{n,m} = 2^{\frac{nm}{2}}\Gamma_m\left(\frac{n}{2}\right)\sum_{k=0}^{\infty}\sum_{\kappa}\frac{2^k h_1^{(k)}(0)}{k!}\left(\frac{n}{2}\right)_{\kappa}C_{\kappa}\left(\boldsymbol{\Phi}(\boldsymbol{\Sigma}^{-1}-2i\boldsymbol{T})^{-1}\right)$.

Proof. Using (4.5), (4.7) and (4.8), it follows that

$$\begin{aligned}
\psi_{\boldsymbol{X}}(\boldsymbol{T}) &= E(etr(i\boldsymbol{T}\boldsymbol{X})) \\
&= c^{I}_{n,m}(\boldsymbol{\Theta})|\boldsymbol{\Sigma}|^{-\frac{n}{2}}\int_{S_m}|\boldsymbol{X}|^{\frac{n}{2}-\frac{m+1}{2}}etr\left(-\frac{1}{2}\boldsymbol{\Sigma}^{-1}\boldsymbol{X}+i\boldsymbol{T}\boldsymbol{X}\right)h_1(tr[\boldsymbol{X}\boldsymbol{\Phi}])d\boldsymbol{X} \\
&= c^{I}_{n,m}(\boldsymbol{\Theta})|\boldsymbol{\Sigma}|^{-\frac{n}{2}}\sum_{k=0}^{\infty}\sum_{\kappa}\frac{h_1^{(k)}(0)}{k!}\int_{S_m}|\boldsymbol{X}|^{\frac{n}{2}-\frac{m+1}{2}}etr\left(-\frac{1}{2}\boldsymbol{\Sigma}^{-1}\boldsymbol{X}+i\boldsymbol{T}\boldsymbol{X}\right)C_\kappa(\boldsymbol{X}\boldsymbol{\Phi})d\boldsymbol{X} \\
&= c^{I}_{n,m}(\boldsymbol{\Theta})|\boldsymbol{\Sigma}|^{-\frac{n}{2}}\sum_{k=0}^{\infty}\sum_{\kappa}\frac{2^{\frac{nm}{2}+k}h_1^{(k)}(0)(\frac{n}{2})_\kappa\Gamma_m(\frac{n}{2})}{k!} \\
&\quad \times|\boldsymbol{\Sigma}^{-1}-2i\boldsymbol{T}|^{-\frac{n}{2}}C_\kappa(\boldsymbol{\Phi}(\boldsymbol{\Sigma}^{-1}-2i\boldsymbol{T})^{-1}),
\end{aligned}$$

and the proof is complete. □

Theorem 4.3.3. *Let $\boldsymbol{X}\sim W^{I}_{m}(\boldsymbol{\Sigma},\boldsymbol{\Phi},n)$, then the joint density function of the eigenvalues $\boldsymbol{\Lambda}=diag(\lambda_1,\dots,\lambda_m)$, $\lambda_1\geq\dots\geq\lambda_m>0$ of $\boldsymbol{X}$ is given by*

$$\begin{aligned}
g'(\boldsymbol{\Lambda}) &= \frac{c^{I}_{n,m}(\boldsymbol{\Theta})|\boldsymbol{\Sigma}|^{-\frac{n}{2}}\pi^{\frac{m^2}{2}}}{\Gamma_m\left(\frac{m}{2}\right)}\prod_{i=1}^{m}\lambda_i^{\frac{n}{2}-\frac{m+1}{2}}\prod_{i<j}^{m}(\lambda_i-\lambda_j) \\
&\quad\times\sum_{\phi}\frac{h_1^{(k)}(0)C_\phi^{\rho,\kappa}(\boldsymbol{I}_m,\boldsymbol{I}_m)C_\phi^{\rho,\kappa}\left(-\frac{1}{2}\boldsymbol{\Sigma}^{-1},\boldsymbol{\Phi}\right)}{r!k!\left[C_\phi(\boldsymbol{I}_m)\right]^2}C_\phi(\boldsymbol{\Lambda}),
\end{aligned}\tag{4.13}$$

with $c^{I}_{n,m}(\boldsymbol{\Theta})$ as in (4.7) and $\sum_\phi$ given in (3.5).

Proof. From (4.5) and R.53 the joint density function of $\boldsymbol{\Lambda}$ is given by

$$g'(\boldsymbol{\Lambda})=\frac{\pi^{\frac{m^2}{2}}}{\Gamma_m\left(\frac{m}{2}\right)}\prod_{i<j}^{m}(\lambda_i-\lambda_j)\int_{O_m}g(\boldsymbol{H}\boldsymbol{\Lambda}\boldsymbol{H}';\boldsymbol{\Theta})d\boldsymbol{H},$$

with

$$\begin{aligned}
I &= \int_{O_m}g(\boldsymbol{H}\boldsymbol{\Lambda}\boldsymbol{H}';\boldsymbol{\Theta})d\boldsymbol{H} \\
&= c^{I}_{n,m}(\boldsymbol{\Theta})|\boldsymbol{\Sigma}|^{-\frac{n}{2}}\int_{O_m}|\boldsymbol{H}\boldsymbol{\Lambda}\boldsymbol{H}'|^{\frac{n}{2}-\frac{m+1}{2}} \\
&\quad\times etr\left(-\frac{1}{2}\boldsymbol{\Sigma}^{-1}\boldsymbol{H}\boldsymbol{\Lambda}\boldsymbol{H}'\right)h_1(tr[\boldsymbol{H}\boldsymbol{\Lambda}\boldsymbol{H}'\boldsymbol{\Phi}])d\boldsymbol{H}.
\end{aligned}\tag{4.14}$$

Using (4.8) and R.25, follows that

$$\begin{aligned}
I &= c^{I}_{n,m}(\boldsymbol{\Theta})|\boldsymbol{\Sigma}|^{-\frac{n}{2}}|\boldsymbol{\Lambda}|^{\frac{n}{2}-\frac{m+1}{2}}\sum_{r=0}^{\infty}\frac{1}{r!}\sum_{k=0}^{\infty}\frac{h_1^{(k)}(0)}{k!} \\
&\times\sum_{\rho}\sum_{\kappa}\int_{O_m}C_\rho\left(-\frac{1}{2}\boldsymbol{\Sigma}^{-1}\boldsymbol{H}\boldsymbol{\Lambda}\boldsymbol{H}'\right)C_\kappa\left(\boldsymbol{\Phi}\boldsymbol{H}\boldsymbol{\Lambda}\boldsymbol{H}'\right)d\boldsymbol{H},
\end{aligned}$$

where

$$\int_{O_m} C_\rho\left(-\frac{1}{2}\boldsymbol{\Sigma}^{-1}\boldsymbol{H}\boldsymbol{\Lambda}\boldsymbol{H}'\right)C_\kappa\left(\boldsymbol{\Phi}\boldsymbol{H}\boldsymbol{\Lambda}\boldsymbol{H}'\right)d\boldsymbol{H}$$
$$= \sum_{\phi\in\rho,\kappa}\frac{C_\phi^{\rho,\kappa}\left(\boldsymbol{I}_m,\boldsymbol{I}_m\right)C_\phi^{\rho,\kappa}\left(-\frac{1}{2}\boldsymbol{\Sigma}^{-1},\boldsymbol{\Phi}\right)}{\left[C_\phi\left(\boldsymbol{I}_m\right)\right]^2}C_\phi\left(\boldsymbol{\Lambda}\right).$$

Hence (4.13) follows. □

Remark 4.3.2. *For $\boldsymbol{\Sigma} = c_1\boldsymbol{I}$ and $\boldsymbol{\Phi} = c_2\boldsymbol{I}$, (4.13) simplifies as*

$$g'(\boldsymbol{\Lambda}) = \frac{\pi^{\frac{m^2}{2}}c^I_{n,m}(\boldsymbol{\Theta})}{\Gamma_m\left(\frac{m}{2}\right)}c_1^{-\frac{mn}{2}}\prod_{i=1}^m\lambda_i^{\frac{n}{2}-\frac{m+1}{2}}\prod_{i<j}^m\left(\lambda_i-\lambda_j\right)etr\left(-\frac{1}{2c_1}\boldsymbol{\Lambda}\right)h_1\left(c_2tr(\boldsymbol{\Lambda})\right). \quad (4.15)$$

Based on Remark 4.3.2, Tables 4.1, 4.2 and 4.3 illustrate the joint density function of the eigenvalues of $\boldsymbol{X}_{2\times 2}$ for different c_1, c_2, n and weight functions $h_1()$, using (4.15) (see code in Appendix D.2).

Remark 4.3.3. *Note that for $h_1(x) = exp(x)$, (4.5) simplifies to the density function of a Wishart distribution with parameters $\left(\frac{c_1}{1-2c_1c_2}\right)\boldsymbol{I}_m$ and n (see Definition 2.1.1), as in (3.12). Then (4.15) simplifies to (3.13).*

From Tables 4.1, 4.2 and 4.3 it is evident that the functional form of the weight function provides increased flexibility to the user. Negative and positive correlations amongst the eigenvalues can be obtained by using different weight functions. Bimodality can be obtained for the $sin(.)$ weight function for small n. This can be explained by the cyclical behaviour of the sin function.

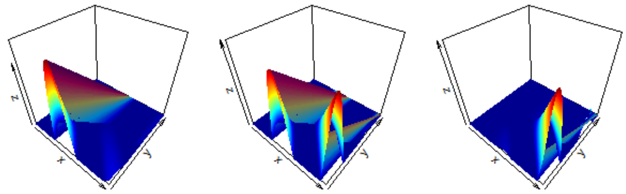
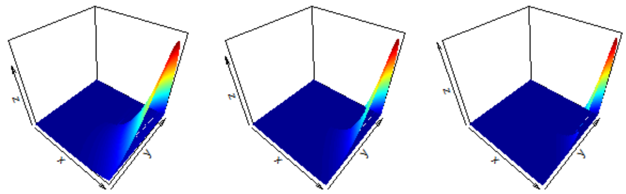
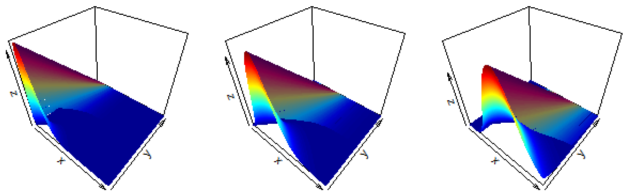
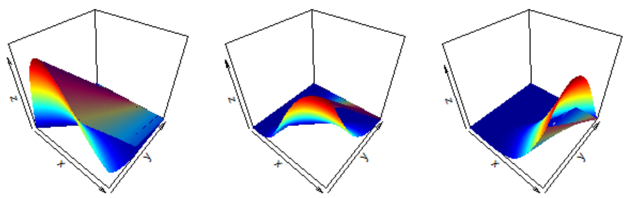
$h_1(c_2x) = sin(c_2x)$	
$h_1(c_2x) = exp(c_2x)$	
$h_1(c_2x) = exp(\frac{1}{c_2x})$	
$h_1(c_2x) = 1 + c_2x$	

TABLE 4.1: Joint density function of the largest (y-axis) and smallest eigenvalue (x-axis) for $c_1 = c_2 = 1$ and $n = 3$ (Left), $n = 6$ (Middle) and $n = 9$ (Right)

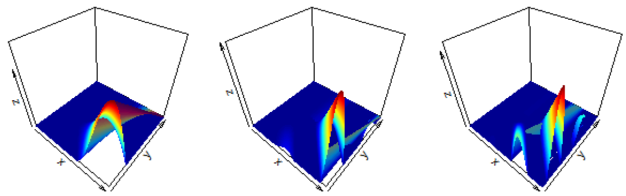
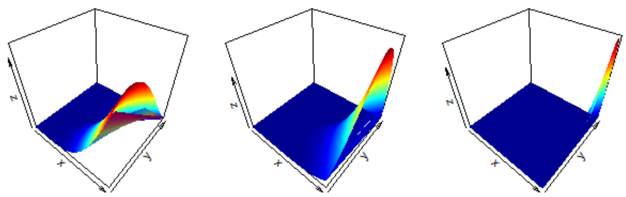
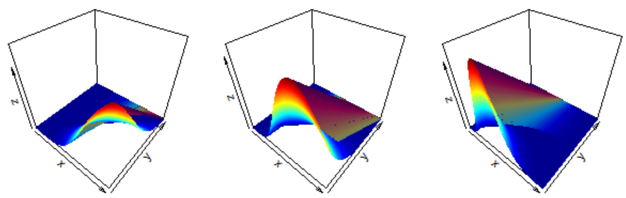
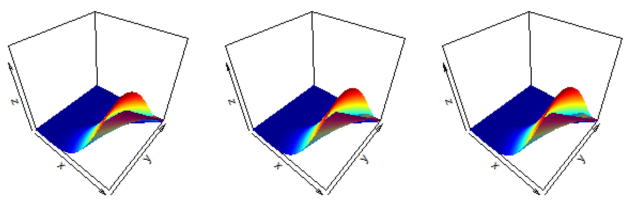
$h_1(c_2x) = sin(c_2x)$	
$h_1(c_2x) = exp(c_2x)$	
$h_1(c_2x) = exp(\frac{1}{c_2x})$	
$h_1(c_2x) = 1 + c_2x$	

TABLE 4.2: Joint density function of the largest (y-axis) and smallest eigenvalue (x-axis) for $n = 9, c_2 = 1$ and $c_1 = 0.1$ (Left), $c_1 = 0.5$ (Middle) and $c_1 = 1.5$ (Right)

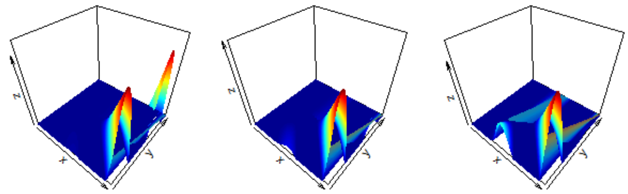
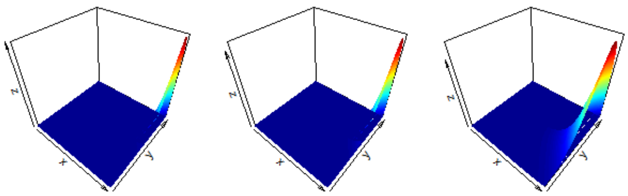
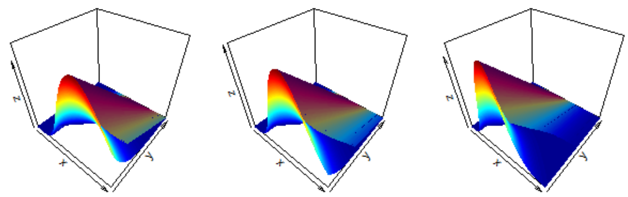
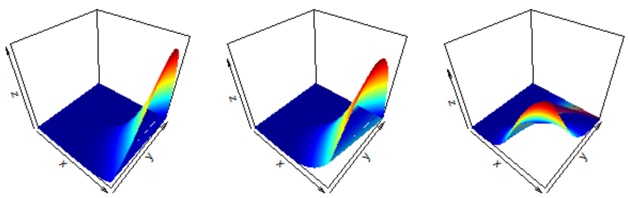
$h_1(c_2x) = sin(c_2x)$	
$h_1(c_2x) = exp(c_2x)$	
$h_1(c_2x) = exp(\frac{1}{c_2x})$	
$h_1(c_2x) = 1 + c_2x$	

TABLE 4.3: Joint density function of the largest (y-axis) and smallest eigenvalue (x-axis) for $n = 9, c_1 = 1$ and $c_2 = 0.1$ (Left), $c_2 = 0.5$ (Middle) and $c_2 = 1.5$ (Right)

4.4 Weighted-type II Wishart distribution

In this section the focus is on the weighted-type II Wishart distribution for which the weight function is of determinant form (see (4.3)).

Definition 4.4.1. *The random matrix $\boldsymbol{X} \in S_m$ is said to have a weighted-type II Wishart distribution (W2WD) with parameters $\boldsymbol{\Psi}$, $\boldsymbol{\Phi} \in S_m$ and the weight function $h_2(.)$, if it has the following density function*

$$\begin{aligned} g(\boldsymbol{X};\boldsymbol{\Theta}) &= \frac{h_2(|\boldsymbol{X\Phi}|)f(\boldsymbol{X};\boldsymbol{\Psi})}{E\left[h_2(|\boldsymbol{X\Phi}|)\right]} \\ &= c^{II}_{n,m}(\boldsymbol{\Theta})|\boldsymbol{\Sigma}|^{-\frac{n}{2}}|\boldsymbol{X}|^{\frac{n}{2}-\frac{m+1}{2}}etr\left(-\frac{1}{2}\boldsymbol{\Sigma}^{-1}\boldsymbol{X}\right)h_2(|\boldsymbol{X\Phi}|), \quad \boldsymbol{\Theta} = (\boldsymbol{\Psi},\boldsymbol{\Phi}) \end{aligned} \tag{4.16}$$

with

$$\begin{aligned} &\left\{c^{II}_{n,m}(\boldsymbol{\Theta})\right\}^{-1} \\ &= \int_{S_m}|\boldsymbol{\Sigma}|^{-\frac{n}{2}}|\boldsymbol{X}|^{\frac{n}{2}-\frac{m+1}{2}}etr\left(-\frac{1}{2}\boldsymbol{\Sigma}^{-1}\boldsymbol{X}\right)h_2(|\boldsymbol{X\Phi}|)d\boldsymbol{X} \qquad (4.17) \\ &= \sum_{k=0}^{\infty}\frac{h_2^{(k)}(0)2^{\frac{(n+2k)m}{2}}\Gamma_m\left(\frac{n+2k}{2}\right)}{k!}|\boldsymbol{\Phi\Sigma}|^k. \qquad (4.18) \end{aligned}$$

denoted as $\boldsymbol{X} \sim \mathbb{W}^{II}_m(\boldsymbol{\Sigma},\boldsymbol{\Phi},n)$, where $f(\boldsymbol{X};\boldsymbol{\Psi})$ is the density function of the Wishart distribution $(W_m(\boldsymbol{\Sigma},n))$ (see Definition 2.1.1) i.e. $\boldsymbol{\Psi} = (\boldsymbol{\Sigma},n)$, $n \geq m$, $\boldsymbol{\Sigma} \in S_m$ and $h_2(.)$ is a Borel measurable function that admits Taylor's series expansion (see R.7). The parameters are restricted to take those values for which the density function is non-negative.

Note that using the Taylor series expansion (see R.7) of $h_2(.)$,

$$h_2[|\boldsymbol{X\Phi}|] = \sum_{k=0}^{\infty}\frac{h_2^{(k)}(0)}{k!}|\boldsymbol{X\Phi}|^k, \tag{4.19}$$

hence

$$E\left[h_2(|\boldsymbol{X\Phi}|)\right] = \sum_{k=0}^{\infty}\frac{h_2^{(k)}(0)}{k!}|\boldsymbol{\Phi}|^k E(|\boldsymbol{X}|^k) = \sum_{k=0}^{\infty}\frac{2^k\Gamma_m\left(\frac{n}{2}+k\right)h_2^{(k)}(0)}{k!\Gamma_m\left(\frac{n}{2}\right)}|\boldsymbol{\Phi\Sigma}|^k$$

from Theorem 2.3.4 and (4.19).

Remark 4.4.1. *It is worthwhile to consider that for the univariate case $m = 1$ in Definition 4.4.1, the weighted-type chi-square distribution is obtained, (both Definitions 4.3.1 and 4.4.1 give the same result for $m = 1$) with density function*

$$g(x;\theta) = \left(\sum_{k=0}^{\infty}\frac{\left(2\phi\sigma^2\right)^k\Gamma\left(\frac{n}{2}+k\right)h^{(k)}(0)}{k!\Gamma\left(\frac{n}{2}\right)}\right)^{-1}\sigma^{-n}x^{\frac{n}{2}-1}e^{-\frac{x}{2\sigma^2}}h(x\phi), \quad \theta = (\sigma^2,\phi),$$

where h can be any of h_i, $i = 1,2$.

4.4.1 Properties

In this section some statistical properties of the W2WD (see Definition 4.4.1) are derived.

Theorem 4.4.1. *Let $\boldsymbol{X} \sim W_m^{II}(\boldsymbol{\Sigma}, \boldsymbol{\Phi}, n)$, then the r^{th} moment of $|\boldsymbol{X}|$ is given by*

$$E\left(|\boldsymbol{X}|^r\right) = \frac{c_{n,m}^{II}(\boldsymbol{\Theta})}{c_{2(\frac{n}{2}+r),m}^{II}(\boldsymbol{\Theta})}|\boldsymbol{\Sigma}|^r,$$

where $c_{n,m}^{II}(\boldsymbol{\Theta})$ and $c_{2(\frac{n}{2}+r),m}^{II}(\boldsymbol{\Theta})$ are as in (4.18).

Proof. Using (4.16) and (4.17) it follows that

$$\begin{aligned} E\left(|\boldsymbol{X}|^r\right) &= c_{n,m}^{II}(\boldsymbol{\Theta})|\boldsymbol{\Sigma}|^{-\frac{n}{2}} \int_{S_m} |\boldsymbol{X}|^{r+\frac{n}{2}-\frac{m+1}{2}} etr\left(-\frac{1}{2}\boldsymbol{\Sigma}^{-1}\boldsymbol{X}\right) h_2(|\boldsymbol{X}\boldsymbol{\Phi}|)d\boldsymbol{X} \\ &= c_{n,m}^{II}(\boldsymbol{\Theta})|\boldsymbol{\Sigma}|^{-\frac{n}{2}} \left(c_{2(\frac{n}{2}+r),m}^{II}(\boldsymbol{\Theta})|\boldsymbol{\Sigma}|^{-\frac{n}{2}-r}\right)^{-1}, \end{aligned}$$

and the proof is complete. □

Theorem 4.4.2. *Let $\boldsymbol{X} \sim W_m^{II}(\boldsymbol{\Sigma}, \boldsymbol{\Phi}, n)$, then the characteristic function of $\boldsymbol{X}$ is given by*

$$\psi_{\boldsymbol{X}}(\boldsymbol{T}) = c_{n,m}^{II}(\boldsymbol{\Theta}) \sum_{k=0}^{\infty} \frac{2^{\frac{(n+2k)m}{2}} h_2^{(k)}(0)\Gamma_m(\frac{n+2k}{2})}{k!} |\boldsymbol{\Phi}\boldsymbol{\Sigma}|^k |\boldsymbol{I} - 2i\boldsymbol{\Sigma}\boldsymbol{T}|^{-\frac{n}{2}-k},$$

with $c_{n,m}^{II}(\boldsymbol{\Theta})$ is as in (4.18).

Proof. Using (4.16), (4.18) and (4.19) it follows that

$$\begin{aligned} \psi_{\boldsymbol{X}}(\boldsymbol{T}) &= E(etr(i\boldsymbol{T}\boldsymbol{X})) \\ &= c_{n,m}^{II}(\boldsymbol{\Theta})|\boldsymbol{\Sigma}|^{-\frac{n}{2}} \int_{S_m} |\boldsymbol{X}|^{\frac{n}{2}-\frac{m+1}{2}} etr\left(-\frac{1}{2}\left(\boldsymbol{\Sigma}^{-1} - 2i\boldsymbol{T}\right)\boldsymbol{X}\right) h_2(|\boldsymbol{X}\boldsymbol{\Phi}|)d\boldsymbol{X} \\ &= c_{n,m}^{II}(\boldsymbol{\Theta})|\boldsymbol{\Sigma}|^{-\frac{n}{2}}|\boldsymbol{\Sigma}^{-1} - 2i\boldsymbol{T}|^{-\frac{n}{2}} \sum_{k=0}^{\infty} \frac{2^{\frac{(n+2k)m}{2}} h_2^{(k)}(0)\Gamma_m(\frac{n+2k}{2})}{k!} |\boldsymbol{\Phi}|^k |\boldsymbol{\Sigma}^{-1} - 2i\boldsymbol{T}|^{-k} \\ &= c_{n,m}^{II}(\boldsymbol{\Theta}) \sum_{k=0}^{\infty} \frac{2^{\frac{(n+2k)m}{2}} h_2^{(k)}(0)\Gamma_m(\frac{n+2k}{2})}{k!} |\boldsymbol{\Phi}\boldsymbol{\Sigma}|^k |\boldsymbol{I} - 2i\boldsymbol{\Sigma}\boldsymbol{T}|^{-\frac{n}{2}-k}. \end{aligned}$$

□

Theorem 4.4.3. *Let $\boldsymbol{X} \sim W_m^{II}(\boldsymbol{\Sigma}, \boldsymbol{\Phi}, n)$, then the joint density function of the eigenvalues $\boldsymbol{\Lambda} = diag(\lambda_1, ..., \lambda_m)$, $\lambda_1 \geq ... \geq \lambda_m > 0$ of $\boldsymbol{X}$ is given by*

$$g'(\boldsymbol{\Lambda}) = c_{n,m}^{II}(\boldsymbol{\Theta})|\boldsymbol{\Sigma}|^{-\frac{n}{2}} \frac{\pi^{\frac{m^2}{2}}}{\Gamma_m\left(\frac{m}{2}\right)} \prod_{i<j}^{m} (\lambda_i - \lambda_j) \sum_{k=0}^{\infty} \frac{h_2^{(k)}(0)}{k!} |\boldsymbol{\Phi}|^k |\boldsymbol{\Lambda}|^{\frac{n}{2}+k-\frac{m+1}{2}} {}_0F_0\left(-\frac{1}{2}\boldsymbol{\Sigma}^{-1}, \boldsymbol{\Lambda}\right), \quad (4.20)$$

with $c_{n,m}^{II}(\boldsymbol{\Theta})$ as in (4.18) and ${}_0F_0(.,.)$ as in R.37.

Proof. From (4.16) and R.53, the density function of $\mathbf{\Lambda}$ follows as

$$g'(\mathbf{\Lambda}) = \frac{\pi^{\frac{m^2}{2}}}{\Gamma_m\left(\frac{m}{2}\right)} \prod_{i<j}^{m} (\lambda_i - \lambda_j) \int_{O_m} g(\mathbf{H\Lambda H}';\mathbf{\Theta})d\mathbf{H},$$

with

$$\begin{aligned} I &= \int_{O_m} g(\mathbf{H\Lambda H}';\mathbf{\Theta})d\mathbf{H} \\ &= c^{II}_{n,m}(\mathbf{\Theta})|\mathbf{\Sigma}|^{-\frac{n}{2}} \int_{O_m} |\mathbf{H\Lambda H}'|^{\frac{n}{2}-\frac{m+1}{2}} etr\left(-\frac{1}{2}\mathbf{\Sigma}^{-1}\mathbf{H\Lambda H}'\right) h_2(|\mathbf{H\Lambda H}'\mathbf{\Phi}|)d\mathbf{H}. \end{aligned}$$

Using (4.19), R.35 and R.38 follows that

$$\begin{aligned} I &= c^{II}_{n,m}(\mathbf{\Theta})|\mathbf{\Sigma}|^{-\frac{n}{2}}|\mathbf{\Lambda}|^{\frac{n}{2}-\frac{m+1}{2}} \sum_{k=0}^{\infty} \frac{h_2^{(k)}(0)}{k!} \int_{O_m} etr\left(-\frac{1}{2}\mathbf{\Sigma}^{-1}\mathbf{H\Lambda H}'\right) |\mathbf{\Phi H\Lambda H}'|^k d\mathbf{H} \\ &= c^{II}_{n,m}(\mathbf{\Theta})|\mathbf{\Sigma}|^{-\frac{n}{2}} \sum_{k=0}^{\infty} \frac{h_2^{(k)}(0)}{k!} |\mathbf{\Phi}|^k |\mathbf{\Lambda}|^{\frac{n}{2}+k-\frac{m+1}{2}} \int_{O_m} etr\left(-\frac{1}{2}\mathbf{\Sigma}^{-1}\mathbf{H\Lambda H}'\right) d\mathbf{H} \\ &= c^{II}_{n,m}(\mathbf{\Theta})|\mathbf{\Sigma}|^{-\frac{n}{2}} \sum_{k=0}^{\infty} \frac{h_2^{(k)}(0)}{k!} |\mathbf{\Phi}|^k |\mathbf{\Lambda}|^{\frac{n}{2}+k-\frac{m+1}{2}} \int_{O_m} {}_0F_0\left(-\frac{1}{2}\mathbf{\Sigma}^{-1}\mathbf{H\Lambda H}'\right) d\mathbf{H} \\ &= c^{II}_{n,m}(\mathbf{\Theta})|\mathbf{\Sigma}|^{-\frac{n}{2}} \sum_{k=0}^{\infty} \frac{h_2^{(k)}(0)}{k!} |\mathbf{\Phi}|^k |\mathbf{\Lambda}|^{\frac{n}{2}+k-\frac{m+1}{2}} {}_0F_0\left(-\frac{1}{2}\mathbf{\Sigma}^{-1}, \mathbf{\Lambda}\right), \end{aligned}$$

hence (4.20) follows. □

4.5 Weighted-type III Wishart distribution

In this section the definition of the weighted-type III Wishart distribution (W3WD) is given. Utilizing a more extended version of (4.4) (allowing more parameters) the following definition follows:

Definition 4.5.1. *The random matrix $\mathbf{X} \in S_m$ is said to have a weighted-type III Wishart distribution (W3WD) with parameters $\mathbf{\Psi}$, $\mathbf{\Phi}_1$ and $\mathbf{\Phi}_2 \in S_m$ and the weight functions $h_1(.)$ and $h_2(.)$, if it has the following density function*

$$\begin{aligned} g(\mathbf{X};\mathbf{\Theta}) &= \frac{h_1(tr[\mathbf{X\Phi}_1])h_2(|\mathbf{X\Phi}_2|)f(\mathbf{X};\mathbf{\Psi})}{E\left[h_1(tr[\mathbf{X\Phi}_1])h_2(|\mathbf{X\Phi}_2|)\right]}, \quad \mathbf{\Theta} = (\mathbf{\Psi}, \mathbf{\Phi}_1, \mathbf{\Phi}_2) \\ &= c^{III}_{n,m}(\mathbf{\Theta})|\mathbf{\Sigma}|^{-\frac{n}{2}}|\mathbf{X}|^{\frac{n}{2}-\frac{m+1}{2}} etr\left(-\frac{1}{2}\mathbf{\Sigma}^{-1}\mathbf{X}\right) h_1(tr[\mathbf{X\Phi}_1])h_2(|\mathbf{X\Phi}_2|) \end{aligned} \tag{4.21}$$

with

$$
\begin{aligned}
&\{c^{III}_{n,m}(\boldsymbol{\Theta})\}^{-1} \\
&= \int_{S_m} |\boldsymbol{\Sigma}|^{-\frac{n}{2}} |\boldsymbol{X}|^{\frac{n}{2}-\frac{m+1}{2}} etr\left(-\frac{1}{2}\boldsymbol{\Sigma}^{-1}\boldsymbol{X}\right) h_1(tr[\boldsymbol{X}\boldsymbol{\Phi}_1]) h_2(|\boldsymbol{X}\boldsymbol{\Phi}_2|) d\boldsymbol{X} \quad (4.22)\\
&= \frac{|\boldsymbol{\Sigma}|^{-\frac{n}{2}}}{2^{\frac{nm}{2}}\Gamma_m\left(\frac{n}{2}\right)} \sum_{k=0}^{\infty} \frac{h_1^{(k)}(0)}{k!} \sum_{\kappa}\sum_{t=0}^{\infty} \frac{h_2^{(t)}(0)|\boldsymbol{\Phi}_2|^t}{t!} \int_{S_m} |\boldsymbol{X}|^{\frac{n}{2}+t-\frac{m+1}{2}} etr\left(-\frac{1}{2}\boldsymbol{\Sigma}^{-1}\boldsymbol{X}\right) C_\kappa(\boldsymbol{X}\boldsymbol{\Phi}_1) d\boldsymbol{X}\\
&= \frac{1}{\Gamma_m\left(\frac{n}{2}\right)} \sum_{k=0}^{\infty}\sum_{t=0}^{\infty} \frac{h_1^{(k)}(0)h_2^{(t)}(0)}{k!t!} 2^{mt+k}\Gamma_m\left(\frac{n}{2}+t\right) |\boldsymbol{\Phi}_2\boldsymbol{\Sigma}|^t \sum_{\kappa}\left(\frac{n}{2}+t\right)_\kappa C_\kappa(\boldsymbol{\Sigma}\boldsymbol{\Phi}_1), \quad (4.23)
\end{aligned}
$$

denoted as $\boldsymbol{X} \sim W_m^{III}(\boldsymbol{\Sigma},\boldsymbol{\Phi}_1,\boldsymbol{\Phi}_2,n)$*, where* $f(\boldsymbol{X};\boldsymbol{\Psi})$ *is the density function of the Wishart distribution* $(W_m(\boldsymbol{\Sigma},n))$ *(see Definition 2.1.1) i.e.* $\boldsymbol{\Psi} = (\boldsymbol{\Sigma},n)$*,* $n \geq m$*,* $\boldsymbol{\Sigma} \in S_m$ *and* $h_1(.)$ *and* $h_2(.)$ *are Borel measurable functions that admit Taylor's series expansion (see R.7). The parameters are restricted to take those values for which the density function is non-negative.*

Note that from (4.8) and (4.19) it follows that

$$
h_1(tr[\boldsymbol{X}\boldsymbol{\Phi}_1])h_2(|\boldsymbol{X}\boldsymbol{\Phi}_2|) = \sum_{k=0}^{\infty}\sum_{\kappa}\sum_{t=0}^{\infty} \frac{h_1^{(k)}(0)h_2^{(t)}(0)}{k!t!} |\boldsymbol{\Phi}_1\boldsymbol{X}|^t \left(\frac{n}{2}+t\right)_\kappa C_\kappa(\boldsymbol{\Phi}_1\boldsymbol{X}). \quad (4.24)
$$

Using (4.21), (4.24) and R.23,

$$
\begin{aligned}
&E\left[h_1(tr[\boldsymbol{X}\boldsymbol{\Phi}_1])h_2(|\boldsymbol{X}\boldsymbol{\Phi}_2|)\right] \\
&= \int_{S_m} \frac{|\boldsymbol{\Sigma}|^{-\frac{n}{2}}}{2^{\frac{nm}{2}}\Gamma_m\left(\frac{n}{2}\right)} |\boldsymbol{X}|^{\frac{n}{2}-\frac{m+1}{2}} h_1(tr[\boldsymbol{X}\boldsymbol{\Phi}_1])h_2(|\boldsymbol{X}\boldsymbol{\Phi}_2|) etr\left(-\frac{1}{2}\boldsymbol{\Sigma}^{-1}\boldsymbol{X}\right) d\boldsymbol{X}\\
&= \frac{|\boldsymbol{\Sigma}|^{-\frac{n}{2}}}{2^{\frac{nm}{2}}\Gamma_m\left(\frac{n}{2}\right)} \sum_{k=0}^{\infty} \frac{h_1^{(k)}(0)}{k!} \sum_{\kappa}\sum_{t=0}^{\infty} \frac{h_2^{(t)}(0)|\boldsymbol{\Phi}_2|^t}{t!} \int_{S_m} |\boldsymbol{X}|^{\frac{n}{2}+t-\frac{m+1}{2}} etr\left(-\frac{1}{2}\boldsymbol{\Sigma}^{-1}\boldsymbol{X}\right) C_\kappa(\boldsymbol{\Phi}_1\boldsymbol{X}) d\boldsymbol{X}\\
&= \frac{|\boldsymbol{\Sigma}|^{-\frac{n}{2}}}{2^{\frac{nm}{2}}\Gamma_m\left(\frac{n}{2}\right)} \sum_{k=0}^{\infty} \frac{h_1^{(k)}(0)}{k!} \sum_{\kappa}\sum_{t=0}^{\infty} \frac{h_2^{(t)}(0)|\boldsymbol{\Phi}_2|^t}{t!} 2^{(\frac{n}{2}+t)m+k}\left(\frac{n}{2}+t\right)_\kappa \Gamma_m\left(\frac{n}{2}+t\right) |\boldsymbol{\Sigma}|^{\frac{n}{2}+t} C_\kappa(\boldsymbol{\Sigma}\boldsymbol{\Phi}_1)\\
&= \frac{1}{\Gamma_m\left(\frac{n}{2}\right)} \sum_{k=0}^{\infty}\sum_{t=0}^{\infty} \frac{h_1^{(k)}(0)h_2^{(t)}(0)}{k!t!} 2^{mt+k}\Gamma_m\left(\frac{n}{2}+t\right) |\boldsymbol{\Phi}_2\boldsymbol{\Sigma}|^t \sum_{\kappa}\left(\frac{n}{2}+t\right)_\kappa C_\kappa(\boldsymbol{\Sigma}\boldsymbol{\Phi}_1).
\end{aligned}
$$

4.5.1 Properties

In this section, some statistical properties of the W3WD (see Definition 4.5.1) are derived.

Theorem 4.5.1. *Let* $\boldsymbol{X} \sim W_m^{III}(\boldsymbol{\Sigma},\boldsymbol{\Phi},n)$*, then the* r^{th} *moment of* $|\boldsymbol{X}|$ *is given by*

$$
E\left(|\boldsymbol{X}|^r\right) = \frac{c^{III}_{n,m}(\boldsymbol{\Theta})}{c^{III}_{2(\frac{n}{2}+r),m}(\boldsymbol{\Theta})} |\boldsymbol{\Sigma}|^r,
$$

where $c^{III}_{n,m}(\boldsymbol{\Theta})$ *and* $c^{III}_{2(\frac{n}{2}+r),m}(\boldsymbol{\Theta})$ *are as in (4.23).*

Proof. Using (4.21) and (4.22) it follows that

$$\begin{aligned} E\left(|\boldsymbol{X}|^r\right) &= c^{III}_{n,m}(\boldsymbol{\Theta})|\boldsymbol{\Sigma}|^{-\frac{n}{2}}\int_{S_m}|\boldsymbol{X}|^{r+\frac{n}{2}-\frac{m+1}{2}}etr\left(-\frac{1}{2}\boldsymbol{\Sigma}^{-1}\boldsymbol{X}\right)h_1(tr[\boldsymbol{X\Phi}_1])h_2(|\boldsymbol{X\Phi}_2|)d\boldsymbol{X} \\ &= c^{III}_{n,m}(\boldsymbol{\Theta})|\boldsymbol{\Sigma}|^{-\frac{n}{2}}\left(c^{III}_{2(\frac{n}{2}+r),m}(\boldsymbol{\Theta})|\boldsymbol{\Sigma}|^{-\frac{n}{2}-r}\right)^{-1}, \end{aligned}$$

and the proof is complete. □

Theorem 4.5.2. *Let $\boldsymbol{X}\sim \boldsymbol{W}^{III}_m(\boldsymbol{\Sigma},\boldsymbol{\Phi},n)$, then characteristic function of $\boldsymbol{X}$ is given by*

$$\psi_{\boldsymbol{X}}(\boldsymbol{T}) = c^{III}_{n,m}(\boldsymbol{\Theta})\sum_{k=0}^{\infty}\sum_{t=0}^{\infty}\sum_{\kappa}\frac{1}{d^*_{n,m}}C_\kappa((\boldsymbol{\Sigma}^{-1}-2i\boldsymbol{T})\boldsymbol{\Phi}_1)|\boldsymbol{I}-2i\boldsymbol{\Sigma T}|^{-\frac{n}{2}-t},$$

with $c^{III}_{n,m}(\boldsymbol{\Theta})$ as in (4.23) and $d^*_{n,m} = \frac{h_1^{(k)}(0)h_2^{(t)}(0)2^{mt+k}\Gamma_m\left(\frac{n}{2}+t\right)}{2^{\frac{nm}{2}}k!t!}|\boldsymbol{\Phi}_2|^t\left(\frac{n}{2}+t\right)_\kappa$.

Proof. Using (4.21), (4.22) and (4.23),

$$\begin{aligned} \psi_{\boldsymbol{X}}(\boldsymbol{T}) &= E(etr(i\boldsymbol{TX})) \\ &= c^{III}_{n,m}(\boldsymbol{\Theta})|\boldsymbol{\Sigma}|^{-\frac{n}{2}}\int_{S_m}|\boldsymbol{X}|^{\frac{n}{2}-\frac{m+1}{2}}etr\left(-\frac{1}{2}\boldsymbol{\Sigma}^{-1}\boldsymbol{X}+i\boldsymbol{TX}\right)h_1(tr[\boldsymbol{X\Phi}_1])h_2(|\boldsymbol{X\Phi}_2|)d\boldsymbol{X} \\ &= c^{III}_{n,m}(\boldsymbol{\Theta})\sum_{k=0}^{\infty}\sum_{t=0}^{\infty}\sum_{\kappa}\frac{h_1^{(k)}(0)h_2^{(t)}(0)2^{mt+k}\Gamma_m\left(\frac{n}{2}+t\right)}{2^{\frac{nm}{2}}k!t!}|\boldsymbol{\Phi}_2|^t\left(\frac{n}{2}+t\right)_\kappa \\ &\quad \times C_\kappa((\boldsymbol{\Sigma}^{-1}-2i\boldsymbol{T})\boldsymbol{\Phi}_1)|\boldsymbol{I}-2i\boldsymbol{\Sigma T}|^{-\frac{n}{2}-t}. \end{aligned}$$

□

Theorem 4.5.3. *Let $\boldsymbol{X}\sim W^{III}_m(\boldsymbol{\Sigma},\boldsymbol{\Phi},n)$, then the joint density function of the eigenvalues $\boldsymbol{\Lambda} = diag(\lambda_1,\dots,\lambda_m)$, $\lambda_1\geq\dots\geq\lambda_m>0$ of $\boldsymbol{X}$ is given by*

$$\begin{aligned} g'(\boldsymbol{\Lambda}) &= \frac{\pi^{\frac{m^2}{2}}}{\Gamma_m\left(\frac{m}{2}\right)}c^{III}_{n,m}(\boldsymbol{\Theta})|\boldsymbol{\Sigma}|^{-\frac{n}{2}}\prod_{i<j}^{m}(\lambda_i-\lambda_j)\sum_{t=0}^{\infty}\sum_{\phi}\frac{h_1^{(k)}(0)h_2^{(t)}(0)}{k!t!r!}|\boldsymbol{\Phi}_2|^t \\ &\quad\times|\boldsymbol{\Lambda}|^{\frac{n}{2}+t-\frac{m+1}{2}}\frac{C_\phi(\boldsymbol{\Lambda})C_\phi^{\rho,\kappa}(\boldsymbol{I}_m,\boldsymbol{I}_m)C_\phi^{\rho,\kappa}\left(-\frac{1}{2}\boldsymbol{\Sigma}^{-1},\boldsymbol{\Phi}\right)}{\left[C_\phi(\boldsymbol{I}_m)\right]^2}, \end{aligned} \tag{4.25}$$

with $c^{III}_{n,m}(\boldsymbol{\Theta})$ as in (4.23) and $\sum_\phi$ as in (3.5).

Proof. From R.53 the density function of $\boldsymbol{\Lambda}$ follows as

$$g'(\boldsymbol{\Lambda}) = \frac{\pi^{\frac{m^2}{2}}}{\Gamma_m\left(\frac{m}{2}\right)}\prod_{i<j}^{m}(\lambda_i-\lambda_j)\int_{O_m}g(\boldsymbol{H\Lambda H}';\boldsymbol{\Theta})d\boldsymbol{H},$$

with

$$\begin{aligned} I &= \int_{O_m}g(\boldsymbol{H\Lambda H}';\boldsymbol{\Theta})d\boldsymbol{H} \\ &= c^{III}_{n,m}(\boldsymbol{\Theta})|\boldsymbol{\Sigma}|^{-\frac{n}{2}}\int_{O_m}|\boldsymbol{H\Lambda H}'|^{\frac{n}{2}-\frac{m+1}{2}}etr\left(-\frac{1}{2}\boldsymbol{\Sigma}^{-1}\boldsymbol{H\Lambda H}'\right)h_1(tr[\boldsymbol{H\Lambda H}'\boldsymbol{\Phi}_1])h_2(|\boldsymbol{H\Lambda H}'\boldsymbol{\Phi}_2|)d\boldsymbol{H}. \end{aligned}$$

Using (4.24) and R.20, the following is obtained

$$
\begin{aligned}
I &= c^{III}_{n,m}(\boldsymbol{\Theta})|\boldsymbol{\Sigma}|^{-\frac{n}{2}}|\boldsymbol{\Lambda}|^{\frac{n}{2}-\frac{m+1}{2}}\sum_{k=0}^{\infty}\sum_{t=0}^{\infty}\frac{h_1^{(k)}(0)h_2^{(t)}}{k!t!}|\boldsymbol{\Phi}_2\boldsymbol{\Lambda}|^t \\
&\quad\times\sum_{\kappa}\int_{O_m} etr\left(-\frac{1}{2}\boldsymbol{\Sigma}^{-1}\boldsymbol{H}\boldsymbol{\Lambda}\boldsymbol{H}'\right)C_\kappa\left(\boldsymbol{\Phi}_1\boldsymbol{H}\boldsymbol{\Lambda}\boldsymbol{H}'\right)d\boldsymbol{H} \\
&= c^{III}_{n,m}(\boldsymbol{\Theta})|\boldsymbol{\Sigma}|^{-\frac{n}{2}}|\boldsymbol{\Lambda}|^{\frac{n}{2}-\frac{m+1}{2}}\sum_{r=0}^{\infty}\frac{1}{r!}\sum_{k=0}^{\infty}\sum_{t=0}^{\infty}\frac{h_1^{(k)}(0)h_2^{(t)}}{k!t!}|\boldsymbol{\Phi}_2\boldsymbol{\Lambda}|^t \\
&\quad\times\sum_{\rho}\sum_{\kappa}\int_{O_m} C_\rho\left(-\frac{1}{2}\boldsymbol{\Sigma}^{-1}\boldsymbol{H}\boldsymbol{\Lambda}\boldsymbol{H}'\right)C_\kappa\left(\boldsymbol{\Phi}_1\boldsymbol{H}\boldsymbol{\Lambda}\boldsymbol{H}'\right)d\boldsymbol{H} \\
&= c^{III}_{n,m}(\boldsymbol{\Theta})|\boldsymbol{\Sigma}|^{-\frac{n}{2}}\sum_{r=0}^{\infty}\sum_{k=0}^{\infty}\sum_{t=0}^{\infty}\sum_{\rho}\sum_{\kappa}\frac{h_1^{(k)}(0)h_2^{(t)}(0)}{k!t!r!}|\boldsymbol{\Phi}_2|^t|\boldsymbol{\Lambda}|^{\frac{n}{2}+t-\frac{m+1}{2}} \\
&\times\int_{O_m} C_\rho\left(-\frac{1}{2}\boldsymbol{\Sigma}^{-1}\boldsymbol{H}\boldsymbol{\Lambda}\boldsymbol{H}'\right)C_\kappa\left(\boldsymbol{\Phi}_1\boldsymbol{H}\boldsymbol{\Lambda}\boldsymbol{H}'\right)d\boldsymbol{H}.
\end{aligned}
$$

From R.25 follows that

$$
\begin{aligned}
&\int_{O_m} C_\rho\left(-\frac{1}{2}\boldsymbol{\Sigma}^{-1}\boldsymbol{H}\boldsymbol{\Lambda}\boldsymbol{H}'\right)C_\kappa\left(\boldsymbol{\Phi}\boldsymbol{H}\boldsymbol{\Lambda}\boldsymbol{H}'\right)d\boldsymbol{H} \\
&= \sum_{\phi\in\rho,\kappa}\frac{C_\phi\left(\boldsymbol{\Lambda}\right)C_\phi^{\rho,\kappa}\left(\boldsymbol{I}_m,\boldsymbol{I}_m\right)C_\phi^{\rho,\kappa}\left(-\frac{1}{2}\boldsymbol{\Sigma}^{-1},\boldsymbol{\Phi}\right)}{\left[C_\phi\left(\boldsymbol{I}_m\right)\right]^2},
\end{aligned}
$$

and hence (4.25) follows. □

4.6 Bayesian applications

In this section a special case of the W1WD is applied as the prior for the normal model which is outlayed in Figure 4.2. Firstly, the Kummer gamma distribution (4.12 with $\phi = 1$) is considered as a new prior for the univariate normal model. In Chapters 2 and 3, the priors under consideration for the underlying matrix variate model were the inverse Wishart, Wishart distributions and the hypergeometric Wishart distributions respectively. Now the Kummer Wishart distribution (4.11) is considered as the prior instead, since it has not been studied as a prior in literature. Note that the SEL function (see R.60) is used throughout this chapter.

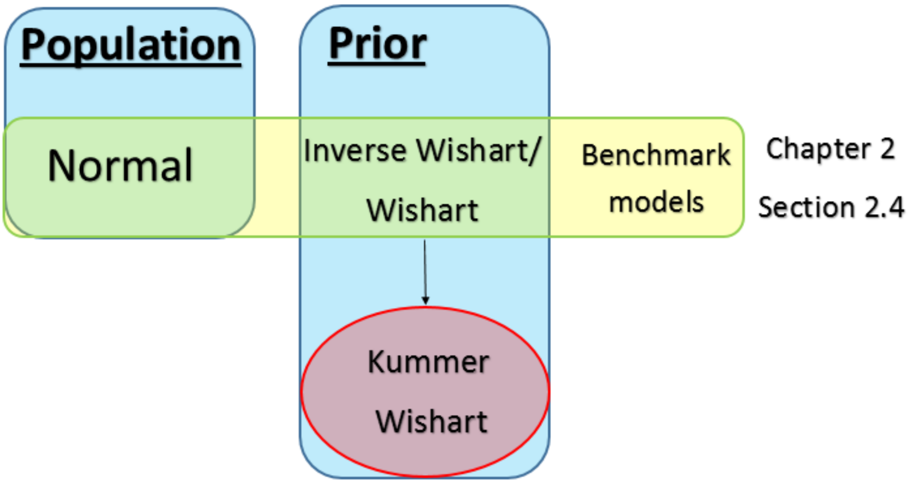


FIGURE 4.2: Outlay of Section 4.6

4.6.1 Univariate Bayesian analysis

Consider a sample of n_1 observations, $\boldsymbol{X} = \{X_1, ..., X_{n_1}\}$, from the normal model, i.e. $X \sim N(\mu, \sigma^2)$, where both parameters are unknown, with likelihood function (3.21). Assume an improper objective prior (see R.55) for $\mu, \pi(\mu) = 1$, and independently the Kummer gamma prior (4.12 with $\phi = 1$) with parameters $\frac{n}{2}, \frac{1}{2\varphi}$ and ω as the subjective prior for σ^2 such that the joint prior density function is given by

$$\pi\left(\mu, \sigma^2\right) \quad \propto \quad (\sigma^2)^{\frac{n}{2}-1} exp\left(-\frac{1}{2\varphi}\sigma^2\right)(1+\sigma^2)^{\omega}, \tag{4.26}$$

where $\pi\left(\mu, \sigma^2\right) = \pi(\mu)\,\pi\left(\sigma^2\right)$.

Then, the joint posterior density function is obtained from (3.21) and (4.26) as follows

$$\begin{aligned} & q\left(\mu, \sigma^2 | \boldsymbol{X}\right) \\ \propto \quad & (\sigma^2)^{\frac{n}{2}-\frac{n_1}{2}-1} exp\left(-\frac{1}{2\varphi}\sigma^2\right)(1+\sigma^2)^{\omega} exp\left[-\frac{1}{2\sigma^2}\sum_{i=1}^{n_1}(X_i-\mu)^2\right] \\ = \quad & (\sigma^2)^{\frac{n}{2}-\frac{n_1}{2}-1} exp\left(-\frac{1}{2\varphi}\sigma^2\right)(1+\sigma^2)^{\omega} exp\left(-\frac{1}{2\frac{\sigma^2}{n_1}}\left[\left(\mu-\overline{X}\right)^2+\frac{1}{n_1}\sum_{i=1}^{n_1}X_i^2-\overline{X}^2\right]\right). \end{aligned} \tag{4.27}$$

Note that

$$q\left(\mu | \boldsymbol{X}, \sigma^2\right) \propto exp\left(-\frac{1}{2\frac{\sigma^2}{n_1}}\left[\mu-\overline{X}\right]^2\right),$$

hence $\mu|\boldsymbol{X},\sigma^2 \sim N(\overline{X},\frac{\sigma^2}{n_1})$, and

$$q\left(\sigma^2|\boldsymbol{X},\mu\right) \propto \left(\sigma^2\right)^{\frac{n}{2}-\frac{n_1}{2}-1} exp\left(-\frac{1}{2\varphi}\sigma^2\right)\left(1+\sigma^2\right)^{\omega} exp\left[-\frac{1}{2\sigma^2}\sum_{i=1}^{n_1}(X_i-\mu)^2\right] \quad (4.28)$$

Theorem 4.6.1. *The marginal posterior density function of μ for the normal model with prior (4.26) is*

$$q(\mu|\boldsymbol{X}) = \frac{\Gamma\left(\frac{n}{2}\right)E_{\sigma_3^2}\left[(\sigma_3^2)^{-\frac{n_1}{2}}(1+\sigma_3^2)^{\omega} exp\left(-\frac{1}{2\sigma_3^2}\sum_{i=1}^{n_1}(X_i-\mu)^2\right)\right]}{\left(2\pi n_1^{-1}\right)^{\frac{1}{2}}(2\varphi)^{\frac{1}{2}}\Gamma\left(\frac{n}{2}+\frac{1}{2}\right)E_{\sigma_4^2}\left[(\sigma_4^2)^{-\frac{n_1}{2}}(1+\sigma_4^2)^{\omega} exp\left(-\frac{1}{2\sigma_4^2}\left(\sum_{i=1}^{n_1}X_i^2-n_1\overline{X}^2\right)\right)\right]}, \quad (4.29)$$

with $\sigma_3^2 \sim Gamma\left(\frac{n}{2},\frac{1}{2\varphi}\right)$ and $\sigma_4^2 \sim Gamma\left(\frac{n}{2}+\frac{1}{2},\frac{1}{2\varphi}\right)$.

Proof. From (4.27), using Remark 3.5.2 and R.41 follows that

$$\begin{aligned} q(\mu|\boldsymbol{X}) &\propto \int_0^{\infty}\left(\sigma^2\right)^{\frac{n}{2}-\frac{n_1}{2}-1} exp\left(-\frac{1}{2\varphi}\sigma^2\right)\left(1+\sigma^2\right)^{\omega} exp\left[-\frac{1}{2\sigma^2}\sum_{i=1}^{n_1}(X_i-\mu)^2\right]d\sigma^2 \\ &\propto E_{\sigma_3^2}\left[(\sigma_3^2)^{-\frac{n_1}{2}}(1+\sigma_3^2)^{\omega} exp\left(-\frac{1}{2\sigma_3^2}\sum_{i=1}^{n_1}(X_i-\mu)^2\right)\right], \end{aligned}$$

with $\sigma_3^2 \sim Gamma\left(\frac{n}{2},\frac{1}{2\varphi}\right)$. Hence

$$q(\mu|\boldsymbol{X}) = c_\mu E_{\sigma_3^2}\left[(\sigma_3^2)^{-\frac{n_1}{2}}(1+\sigma_3^2)^{\omega} exp\left(-\frac{1}{2\sigma_3^2}\sum_{i=1}^{n_1}(X_i-\mu)^2\right)\right], \quad (4.30)$$

with

$$\begin{aligned} &c_\mu^{-1} \\ &= \int_{-\infty}^{\infty} E_{\sigma_3^2}\left[(\sigma_3^2)^{-\frac{n_1}{2}}(1+\sigma_3^2)^{\omega} exp\left(-\frac{1}{2\sigma_3^2}\sum_{i=1}^{n_1}(X_i-\mu)^2\right)\right]d\mu \\ &= \int_{-\infty}^{\infty}\int_0^{\infty}\frac{(2\varphi)^{-\frac{n}{2}}}{\Gamma\left(\frac{n}{2}\right)}\left(\sigma^2\right)^{\frac{n}{2}-\frac{n_1}{2}-1} exp\left(-\frac{1}{2\varphi}\sigma^2\right)\left(1+\sigma^2\right)^{\omega} exp\left(-\frac{1}{2\sigma^2}\sum_{i=1}^{n_1}(X_i-\mu)^2\right)d\sigma^2 d\mu \\ &= \int_0^{\infty}\frac{(2\varphi)^{-\frac{n}{2}}}{\Gamma\left(\frac{n}{2}\right)}\left(\sigma^2\right)^{\frac{n}{2}-\frac{n_1}{2}-1} exp\left(-\frac{1}{2\varphi}\sigma^2\right)\left(1+\sigma^2\right)^{\omega}\int_{-\infty}^{\infty} exp\left(-\frac{1}{2\sigma^2}\sum_{i=1}^{n_1}(X_i-\mu)^2\right)d\mu d\sigma^2 \\ &= \int_0^{\infty}\frac{(2\varphi)^{-\frac{n}{2}}}{\Gamma\left(\frac{n}{2}\right)}\left(\sigma^2\right)^{\frac{n}{2}-\frac{n_1}{2}-1} exp\left(-\frac{1}{2\varphi}\sigma^2\right) \\ &\quad\times\left(1+\sigma^2\right)^{\omega}\int_{-\infty}^{\infty} exp\left(-\frac{1}{2\frac{\sigma^2}{n_1}}\left[(\mu-\overline{X})^2+\frac{1}{n_1}\sum_{i=1}^{n_1}X_i^2-\overline{X}^2\right]\right)d\mu d\sigma^2 \end{aligned}$$

$$
\begin{aligned}
&= \int_0^\infty \left(2\pi\frac{\sigma^2}{n_1}\right)^{\frac{1}{2}} \frac{(2\varphi)^{-\frac{n}{2}}}{\Gamma\left(\frac{n}{2}\right)} (\sigma^2)^{\frac{n}{2}-\frac{n_1}{2}-1} exp\left(-\frac{1}{2\varphi}\sigma^2\right) \\
&\quad \times (1+\sigma^2)^\omega exp\left(-\frac{1}{2\sigma^2}\left(\sum_{i=1}^{n_1} X_i^2 - n_1\overline{X}^2\right)\right) d\sigma^2 \\
&= (2\pi n_1^{-1})^{\frac{1}{2}} \int_0^\infty \frac{(2\varphi)^{-\frac{n}{2}}}{\Gamma\left(\frac{n}{2}\right)} (\sigma^2)^{\frac{n}{2}-\frac{n_1}{2}+\frac{1}{2}-1} \\
&\quad \times exp\left(-\frac{1}{2\varphi}\sigma^2\right)(1+\sigma^2)^\omega exp\left(-\frac{1}{2\sigma^2}\left(\sum_{i=1}^{n_1} X_i^2 - n_1\overline{X}^2\right)\right) d\sigma^2 \\
&= (2\pi n_1^{-1})^{\frac{1}{2}} \frac{(2\varphi)^{-\frac{n}{2}}}{\Gamma\left(\frac{n}{2}\right)} \frac{\Gamma\left(\frac{n}{2}+\frac{1}{2}\right)}{(2\varphi)^{-\frac{n}{2}-\frac{1}{2}}} E_{\sigma_4^2}\left[(\sigma_4^2)^{-\frac{n_1}{2}}(1+\sigma_4^2)^\omega exp\left(-\frac{1}{2\sigma_4^2}\left(\sum_{i=1}^{n_1} X_i^2 - n_1\overline{X}^2\right)\right)\right] \\
&= \frac{(2\pi n_1^{-1})^{\frac{1}{2}}(2\varphi)^{\frac{1}{2}}\Gamma\left(\frac{n}{2}+\frac{1}{2}\right)}{\Gamma\left(\frac{n}{2}\right)} \\
&\quad \times E_{\sigma_4^2}\left[(\sigma_4^2)^{-\frac{n_1}{2}}(1+\sigma^2)^\omega exp\left(-\frac{1}{2\sigma^2}\left(\sum_{i=1}^{n_1} X_i^2 - n_1\overline{X}^2\right)\right)\right], \qquad (4.31)
\end{aligned}
$$

from (3.21), R.39, Remark 3.5.2 and R.41, with $\sigma_4^2 \sim Gamma\left(\frac{n}{2}+\frac{1}{2}, \frac{1}{2\varphi}\right)$. Therefore (4.29) follows from (4.30) and (4.31). □

Theorem 4.6.2. *The marginal posterior density of σ^2 for the normal model with prior (4.26) is*

$$
q(\sigma^2|\boldsymbol{X}) = \frac{(\sigma^2)^{\frac{n}{2}-\frac{n_1}{2}-\frac{1}{2}} exp\left(-\frac{1}{2\varphi}\sigma^2\right)(1+\sigma^2)^\omega exp\left(-\frac{1}{2\sigma^2}\left[\sum_{i=1}^{n_1} X_i^2 - \overline{X}^2\right]\right)}{\Gamma\left(\frac{n}{2}+\frac{1}{2}\right)(2\varphi)^{-\frac{n}{2}-\frac{1}{2}} E_{\sigma_4^2}\left[(\sigma_4^2)^{-\frac{n_1}{2}}(1+\sigma_4^2)^\omega exp\left(-\frac{1}{2\sigma_4^2}\left[\sum_{i=1}^{n_1} X_i^2 - \overline{X}^2\right]\right)\right]}, \qquad (4.32)
$$

with $\sigma_4^2 \sim Gamma\left(\frac{n}{2}+\frac{1}{2}, \frac{1}{2\varphi}\right)$.

Proof. From (4.27) and R.39

$$
\begin{aligned}
q(\sigma^2|\boldsymbol{X}) &\propto (\sigma^2)^{\frac{n}{2}-\frac{n_1}{2}-1} exp\left(-\frac{1}{2\varphi}\sigma^2\right)(1+\sigma^2)^\omega exp\left(-\frac{1}{2\sigma^2}\left[\sum_{i=1}^{n_1} X_i^2 - \overline{X}^2\right]\right) \\
&\quad \times \int_{-\infty}^{\infty} exp\left(-\frac{1}{2\frac{\sigma^2}{n_1}}\left[(\mu-\overline{X})^2\right]\right) d\mu \\
&= (\sigma^2)^{\frac{n}{2}-\frac{n_1}{2}-1} exp\left(-\frac{1}{2\varphi}\sigma^2\right)(1+\sigma^2)^\omega exp\left(-\frac{1}{2\sigma^2}\left[\sum_{i=1}^{n_1} X_i^2 - \overline{X}^2\right]\right)\left(2\pi\frac{\sigma^2}{n_1}\right)^{\frac{1}{2}}.
\end{aligned}
$$

Hence

$$
q(\sigma^2|\boldsymbol{X}) = c_{\sigma^2}(\sigma^2)^{\frac{n}{2}-\frac{n_1}{2}-\frac{1}{2}} exp\left(-\frac{1}{2\varphi}\sigma^2\right)(1+\sigma^2)^\omega exp\left(-\frac{1}{2\sigma^2}\left[\sum_{i=1}^{n_1} X_i^2 - \overline{X}^2\right]\right), \qquad (4.33)
$$

with

$$\begin{aligned} c_{\sigma^2}^{-1} &= \int_0^\infty (\sigma^2)^{\frac{n}{2}-\frac{n_1}{2}-\frac{1}{2}} exp\left(-\frac{1}{2\varphi}\sigma^2\right)(1+\sigma^2)^\omega exp\left(-\frac{1}{2\sigma^2}\left[\sum_{i=1}^{n_1} X_i^2 - \overline{X}^2\right]\right) d\sigma^2 \\ &= \Gamma\left(\frac{n}{2}+\frac{1}{2}\right)(2\varphi)^{-\frac{n}{2}-\frac{1}{2}} E_{\sigma_4^2}\left[(\sigma_4^2)^{-\frac{n_1}{2}}(1+\sigma_4^2)^\omega exp\left(-\frac{1}{2\sigma_4^2}\left[\sum_{i=1}^{n_1} X_i^2 - \overline{X}^2\right]\right)\right], \end{aligned} \tag{4.34}$$

using Remark 3.5.2 and R.41, where $\sigma_4^2 \sim Gamma\left(\frac{n}{2}+\frac{1}{2}, \frac{1}{2\varphi}\right)$. Therefore (4.32) follows from (4.33) and (4.34). □

Theorem 4.6.3. *The Bayes estimator of μ for the normal model with prior (4.26) is*

$$\widehat{\mu}_{B,KG} = \overline{X} \tag{4.35}$$

Proof. From (4.29) the posterior expected value of $\mu - \overline{X}$ is

$$\begin{aligned} & E(\mu - \overline{X}|\boldsymbol{X}) \\ = & \frac{\int_{-\infty}^{\infty}(\mu-\overline{X})\Gamma\left(\frac{n}{2}\right)E_{\sigma_3^2}\left[(\sigma_3^2)^{-\frac{n_1}{2}}(1+\sigma_3^2)^\omega exp\left(-\frac{1}{2\sigma_3^2}\sum_{i=1}^{n_1}(X_i-\mu)^2\right)\right]d\mu}{(2\pi n_1^{-1})^{\frac{1}{2}}(2\varphi)^{\frac{1}{2}}\Gamma\left(\frac{n}{2}+\frac{1}{2}\right)E_{\sigma_4^2}\left[(\sigma_4^2)^{-\frac{n_1}{2}}(1+\sigma_4^2)^\omega exp\left(-\frac{1}{2\sigma_4^2}\left(\sum_{i=1}^{n_1}X_i^2 - n_1\overline{X}^2\right)\right)\right]} \end{aligned}$$

where $\sigma_3^2 \sim Gamma\left(\frac{n}{2}, \frac{1}{2\varphi}\right)$ and $\sigma_4^2 \sim Gamma\left(\frac{n}{2}+\frac{1}{2}, \frac{1}{2\varphi}\right)$. Note that $E(\mu - \overline{X}) = 0$ since the integrand is an odd function, hence $E(\mu) = \widehat{\mu}_{B,KG} = \overline{X}$. □

Theorem 4.6.4. *The Bayes estimator of σ^2 for the normal model with prior (4.26) is*

$$\widehat{\sigma^2}_{B,KG} = \frac{\Gamma\left(\frac{n}{2}+\frac{3}{2}\right)E_{\sigma_5^2}\left[(\sigma_5^2)^{-\frac{n_1}{2}}(1+\sigma_5^2)^\omega exp\left(-\frac{1}{2\sigma_5^2}\left[\sum_{i=1}^{n_1}X_i^2-\overline{X}^2\right]\right)\right]}{\Gamma\left(\frac{n}{2}+\frac{1}{2}\right)(2\varphi)E_{\sigma_4^2}\left[(\sigma_4^2)^{-\frac{n_1}{2}}(1+\sigma_4^2)^\omega exp\left(-\frac{1}{2\sigma_4^2}\left[\sum_{i=1}^{n_1}X_i^2-\overline{X}^2\right]\right)\right]}, \tag{4.36}$$

where $\sigma_4^2 \sim Gamma\left(\frac{n}{2}+\frac{1}{2}, \frac{1}{2\varphi}\right)$ and $\sigma_5^2 \sim Gamma\left(\frac{n}{2}+\frac{3}{2}, \frac{1}{2\varphi}\right)$.

Proof. From (4.32) the posterior expected value of σ^2 is

$$\begin{aligned} E(\sigma^2|\boldsymbol{X}) &= \frac{\int_0^\infty (\sigma^2)^{\frac{n}{2}-\frac{n_1}{2}+\frac{1}{2}} exp\left(-\frac{1}{2\varphi}\sigma^2\right)(1+\sigma^2)^\omega exp\left(-\frac{1}{2\sigma^2}\left[\sum_{i=1}^{n_1}X_i^2-\overline{X}^2\right]\right)d\sigma^2}{\Gamma\left(\frac{n}{2}+\frac{1}{2}\right)(2\varphi)^{-\frac{n}{2}-\frac{1}{2}}E_{\sigma_4^2}\left[(\sigma_4^2)^{-\frac{n_1}{2}}(1+\sigma_4^2)^\omega exp\left(-\frac{1}{2\sigma_4^2}\left[\sum_{i=1}^{n_1}X_i^2-\overline{X}^2\right]\right)\right]} \\ &= \frac{\Gamma\left(\frac{n}{2}+\frac{3}{2}\right)E_{\sigma_5^2}\left[(\sigma_5^2)^{-\frac{n_1}{2}}(1+\sigma_5^2)^\omega exp\left(-\frac{1}{2\sigma_5^2}\left[\sum_{i=1}^{n_1}X_i^2-\overline{X}^2\right]\right)\right]}{\Gamma\left(\frac{n}{2}+\frac{1}{2}\right)(2\varphi)E_{\sigma_4^2}\left[(\sigma_4^2)^{-\frac{n_1}{2}}(1+\sigma_4^2)^\omega exp\left(-\frac{1}{2\sigma_4^2}\left[\sum_{i=1}^{n_1}X_i^2-\overline{X}^2\right]\right)\right]}, \end{aligned}$$

from Remark 3.5.2 and R.41, where $\sigma_4^2 \sim Gamma\left(\frac{n}{2}+\frac{1}{2}, \frac{1}{2\varphi}\right)$ and $\sigma_5^2 \sim Gamma\left(\frac{n}{2}+\frac{3}{2}, \frac{1}{2\varphi}\right)$. □

4.6.2 Multivariate Bayesian analysis

Consider the matrix variate normal model, $\boldsymbol{X} \sim N_{m,p}(\boldsymbol{\mu}, \boldsymbol{\Sigma} \otimes \boldsymbol{\Omega})$ as in R.49 with likelihood function (3.33) based on a sample, $\boldsymbol{X} = \{\boldsymbol{X}_1, ..., \boldsymbol{X}_{n_1}\}$, of size n_1. Consider an improper objective prior (see R.55) for $\boldsymbol{\mu}, \pi(\boldsymbol{\mu}) = 1$, and the Kummer Wishart distribution (4.11) with parameters $\boldsymbol{\Phi}, \boldsymbol{\Theta}$ and n as the subjective prior for $\boldsymbol{\Sigma}$, i.e. $\boldsymbol{\Sigma} \sim KW_m(\boldsymbol{\Phi}, \boldsymbol{\Theta}, n)$, such that the joint prior density function is

$$\pi(\boldsymbol{\mu}, \boldsymbol{\Sigma}) \propto |\boldsymbol{\Phi}|^{-\frac{n}{2}} |\boldsymbol{\Sigma}|^{\frac{n}{2} - \frac{m+1}{2}} etr\left(-\frac{1}{2}\boldsymbol{\Phi}^{-1}\boldsymbol{\Sigma}\right)(1 + tr(\boldsymbol{\Sigma\Theta})). \tag{4.37}$$

From (3.33) and (4.37) the joint posterior density function is

$$\begin{aligned} q(\boldsymbol{\mu}, \boldsymbol{\Sigma}|\boldsymbol{X}) \quad &\propto \quad |\boldsymbol{\Sigma}|^{-\frac{n_1 p}{2} + \frac{n}{2} - \frac{m+1}{2}} etr\left[-\frac{1}{2}\boldsymbol{\Sigma}^{-1}\left[\boldsymbol{V} + n_1(\overline{\boldsymbol{X}} - \boldsymbol{\mu})\boldsymbol{\Omega}^{-1}(\overline{\boldsymbol{X}} - \boldsymbol{\mu})'\right]\right] \\ &\quad \times etr\left(-\frac{1}{2}\boldsymbol{\Phi}^{-1}\boldsymbol{\Sigma}\right)(1 + tr(\boldsymbol{\Sigma\Theta})), \end{aligned} \tag{4.38}$$

with $\boldsymbol{V}$ as in (2.7). Note that

$$q(\boldsymbol{\mu}|\boldsymbol{X}, \boldsymbol{\Sigma}) \propto etr\left[-\frac{n_1}{2}\boldsymbol{\Sigma}^{-1}(\overline{\boldsymbol{X}} - \boldsymbol{\mu})\boldsymbol{\Omega}^{-1}(\overline{\boldsymbol{X}} - \boldsymbol{\mu})'\right],$$

hence $\boldsymbol{\mu}|\boldsymbol{X}, \boldsymbol{\Sigma} \sim N_{m,p}(\overline{\boldsymbol{X}}, \frac{1}{n_1}\boldsymbol{\Sigma} \otimes \boldsymbol{\Omega})$, and

$$\begin{aligned} q(\boldsymbol{\Sigma}|\boldsymbol{X}, \boldsymbol{\mu}) \quad &\propto \quad |\boldsymbol{\Sigma}|^{-\frac{n_1 p}{2} + \frac{n}{2} - \frac{m+1}{2}} etr\left[-\frac{1}{2}\boldsymbol{\Sigma}^{-1}\left[\boldsymbol{V} + n_1(\overline{\boldsymbol{X}} - \boldsymbol{\mu})\boldsymbol{\Omega}^{-1}(\overline{\boldsymbol{X}} - \boldsymbol{\mu})'\right]\right] \\ &\quad \times etr\left(-\frac{1}{2}\boldsymbol{\Phi}^{-1}\boldsymbol{\Sigma}\right)(1 + tr(\boldsymbol{\Sigma\Theta})). \end{aligned} \tag{4.39}$$

Theorem 4.6.5. *The marginal posterior density function of $\boldsymbol{\mu}$ for the normal model with prior (4.37) is*

$$\begin{aligned} & q(\boldsymbol{\mu}|\boldsymbol{X}) \\ = & \frac{\left(\frac{\pi}{n_1}\right)^{-\frac{mp}{2}} \Gamma_m\left(\frac{n}{2}\right) |\boldsymbol{\Omega}|^{-\frac{m}{2}} |\boldsymbol{\Phi}|^{-\frac{p}{2}}}{\Gamma_m\left(\frac{n+p}{2}\right) E_{\boldsymbol{\Sigma}_4}\left[|\boldsymbol{\Sigma}_4|^{-\frac{n_1 p}{2}} etr\left[-\frac{1}{2}\boldsymbol{\Sigma}_4^{-1}\boldsymbol{V}\right](1 + tr(\boldsymbol{\Sigma}_4\boldsymbol{\Theta}))\right]} \\ & \times E_{\boldsymbol{\Sigma}_3}\left[|\boldsymbol{\Sigma}_3|^{-\frac{n_1 p}{2}} etr\left[-\frac{1}{2}\boldsymbol{\Sigma}_3^{-1}\left[\boldsymbol{V} + n_1(\overline{\boldsymbol{X}} - \boldsymbol{\mu})\boldsymbol{\Omega}^{-1}(\overline{\boldsymbol{X}} - \boldsymbol{\mu})'\right]\right](1 + tr(\boldsymbol{\Sigma}_3\boldsymbol{\Theta}))\right], \end{aligned} \tag{4.40}$$

with $\boldsymbol{\Sigma}_3 \sim W_m(\boldsymbol{\Phi}, n)$, $\boldsymbol{\Sigma}_4 \sim W_m(\boldsymbol{\Phi}, n+p)$ and $\boldsymbol{V}$ as defined in (2.7).

Proof. From (4.38), using Remark 3.5.2 and Definition 2.1.1,

$$\begin{aligned} q(\boldsymbol{\mu}|\boldsymbol{X}) \quad &\propto \quad \int_{S_m} |\boldsymbol{\Sigma}|^{\frac{n}{2} - \frac{n_1 p}{2} - \frac{m+1}{2}} etr\left[-\frac{1}{2}\boldsymbol{\Sigma}^{-1}\left[\boldsymbol{V} + n_1(\overline{\boldsymbol{X}} - \boldsymbol{\mu})\boldsymbol{\Omega}^{-1}(\overline{\boldsymbol{X}} - \boldsymbol{\mu})'\right]\right] \\ &\quad \times etr\left(-\frac{1}{2}\boldsymbol{\Phi}^{-1}\boldsymbol{\Sigma}\right)(1 + tr(\boldsymbol{\Sigma\Theta}))d\boldsymbol{\Sigma} \\ &\propto \quad E_{\boldsymbol{\Sigma}_3}\left[|\boldsymbol{\Sigma}_3|^{-\frac{n_1 p}{2}} etr\left[-\frac{1}{2}\boldsymbol{\Sigma}_3^{-1}\left[\boldsymbol{V} + n_1(\overline{\boldsymbol{X}} - \boldsymbol{\mu})\boldsymbol{\Omega}^{-1}(\overline{\boldsymbol{X}} - \boldsymbol{\mu})'\right]\right](1 + tr(\boldsymbol{\Sigma}_3\boldsymbol{\Theta}))\right], \end{aligned}$$

with $\boldsymbol{\Sigma}_3 \sim W_m(\boldsymbol{\Phi}, n)$. Hence

$$q(\boldsymbol{\mu}|\boldsymbol{X}) = c_{\boldsymbol{\mu}} E_{\boldsymbol{\Sigma}_3}\left[|\boldsymbol{\Sigma}_3|^{-\frac{n_1 p}{2}} etr\left[-\frac{1}{2}\boldsymbol{\Sigma}_3^{-1}\left[\boldsymbol{V} + n_1(\overline{\boldsymbol{X}} - \boldsymbol{\mu})\boldsymbol{\Omega}^{-1}(\overline{\boldsymbol{X}} - \boldsymbol{\mu})'\right]\right](1 + tr(\boldsymbol{\Sigma}_3\boldsymbol{\Theta}))\right], \quad (4.41)$$

with

$$\begin{aligned}
& c_{\boldsymbol{\mu}}^{-1} \\
= & \int_{\mathbb{R}^{m\times p}} E_{\boldsymbol{\Sigma}_3}\left[|\boldsymbol{\Sigma}_3|^{-\frac{n_1 p}{2}} etr\left[-\frac{1}{2}\boldsymbol{\Sigma}_3^{-1}\left[\boldsymbol{V} + n_1(\overline{\boldsymbol{X}} - \boldsymbol{\mu})\boldsymbol{\Omega}^{-1}(\overline{\boldsymbol{X}} - \boldsymbol{\mu})'\right]\right](1 + tr(\boldsymbol{\Sigma}\boldsymbol{\Theta}))\right] d\boldsymbol{\mu} \\
= & \int_{\mathbb{R}^{m\times p}} \left[2^{\frac{mp}{2}}\Gamma_m\left(\frac{n}{2}\right)\right]^{-1} |\boldsymbol{\Phi}|^{-\frac{n}{2}} \int_{S_m} |\boldsymbol{\Sigma}|^{\frac{n}{2} - \frac{n_1 p}{2} - \frac{m+1}{2}} etr\left(-\frac{1}{2}\boldsymbol{\Phi}^{-1}\boldsymbol{\Sigma}\right) \\
& \times etr\left[-\frac{1}{2}\boldsymbol{\Sigma}^{-1}\left[\boldsymbol{V} + n_1(\overline{\boldsymbol{X}} - \boldsymbol{\mu})\boldsymbol{\Omega}^{-1}(\overline{\boldsymbol{X}} - \boldsymbol{\mu})'\right]\right](1 + tr(\boldsymbol{\Sigma}\boldsymbol{\Theta}))d\boldsymbol{\Sigma}d\boldsymbol{\mu} \\
= & \left[2^{\frac{nm}{2}}\Gamma_m\left(\frac{n}{2}\right)\right]^{-1} |\boldsymbol{\Phi}|^{-\frac{n}{2}} \int_{S_m} |\boldsymbol{\Sigma}|^{\frac{n}{2} - \frac{n_1 p}{2} - \frac{m+1}{2}} etr\left[-\frac{1}{2}\boldsymbol{\Sigma}^{-1}\boldsymbol{V}\right] etr\left(-\frac{1}{2}\boldsymbol{\Phi}^{-1}\boldsymbol{\Sigma}\right) \\
& \times(1 + tr(\boldsymbol{\Sigma}\boldsymbol{\Theta}))\int_{\mathbb{R}^{m\times p}} etr\left[-\frac{1}{2}\boldsymbol{\Sigma}^{-1}n_1(\overline{\boldsymbol{X}} - \boldsymbol{\mu})\boldsymbol{\Omega}^{-1}(\overline{\boldsymbol{X}} - \boldsymbol{\mu})'\right] d\boldsymbol{\mu}d\boldsymbol{\Sigma} \\
= & \frac{2^{-\frac{nm}{2}}\left(\frac{2\pi}{n_1}\right)^{\frac{mp}{2}}}{\Gamma_m\left(\frac{n}{2}\right)} |\boldsymbol{\Omega}|^{\frac{m}{2}} |\boldsymbol{\Phi}|^{-\frac{n}{2}} \\
& \times \int_{S_m} |\boldsymbol{\Sigma}|^{\frac{n+p}{2} - \frac{n_1 p}{2} - \frac{m+1}{2}} etr\left[-\frac{1}{2}\boldsymbol{\Sigma}^{-1}\boldsymbol{V}\right] etr\left(-\frac{1}{2}\boldsymbol{\Phi}^{-1}\boldsymbol{\Sigma}\right)(1 + tr(\boldsymbol{\Sigma}\boldsymbol{\Theta}))d\boldsymbol{\Sigma} \\
= & \frac{2^{-\frac{nm}{2}}\left(\frac{2\pi}{n_1}\right)^{\frac{mp}{2}}}{\Gamma_m\left(\frac{n}{2}\right)} |\boldsymbol{\Omega}|^{\frac{m}{2}} |\boldsymbol{\Phi}|^{-\frac{n}{2}} \\
& \times 2^{\frac{(n+p)m}{2}}\Gamma_m\left(\frac{n+p}{2}\right)|\boldsymbol{\Phi}|^{\frac{n+p}{2}} E_{\boldsymbol{\Sigma}_4}\left[|\boldsymbol{\Sigma}_4|^{-\frac{n_1 p}{2}} etr\left[-\frac{1}{2}\boldsymbol{\Sigma}_4^{-1}\boldsymbol{V}\right](1 + tr(\boldsymbol{\Sigma}_4\boldsymbol{\Theta}))\right] \\
= & \frac{2^{mp}\Gamma_m\left(\frac{n+p}{2}\right)\left(\frac{\pi}{n_1}\right)^{\frac{mp}{2}}}{\Gamma_m\left(\frac{n}{2}\right)} |\boldsymbol{\Omega}|^{\frac{m}{2}} |\boldsymbol{\Phi}|^{\frac{p}{2}} E_{\boldsymbol{\Sigma}_4}\left[|\boldsymbol{\Sigma}_4|^{-\frac{n_1 p}{2}} etr\left[-\frac{1}{2}\boldsymbol{\Sigma}_4^{-1}\boldsymbol{V}\right](1 + tr(\boldsymbol{\Sigma}_4\boldsymbol{\Theta}))\right], \quad (4.42)
\end{aligned}$$

from Remark 3.5.2, Definition 2.1.1 and R.49, where $\boldsymbol{\Sigma}_4 \sim W_m(\boldsymbol{\Phi}, n+p)$. From (4.41) and (4.42), (4.40) follows. □

Theorem 4.6.6. *The marginal posterior density function of $\boldsymbol{\Sigma}$ for the normal model with prior (4.37) is*

$$\begin{aligned}
q(\boldsymbol{\Sigma}|\boldsymbol{X}) = & \frac{2^{-\frac{(n+p)m}{2}}|\boldsymbol{\Phi}|^{-\frac{n+p}{2}}}{\Gamma_m\left(\frac{n+p}{2}\right)E_{\boldsymbol{\Sigma}_4}\left[|\boldsymbol{\Sigma}_4|^{-\frac{n_1 p}{2}} etr\left[-\frac{1}{2}\boldsymbol{\Sigma}_4^{-1}\boldsymbol{V}\right](1 + tr(\boldsymbol{\Sigma}_4\boldsymbol{\Theta}))\right]} \\
& \times|\boldsymbol{\Sigma}|^{\frac{n+p}{2} - \frac{n_1 p}{2} - \frac{m+1}{2}} etr\left[-\frac{1}{2}\boldsymbol{\Sigma}^{-1}\boldsymbol{V}\right] etr\left(-\frac{1}{2}\boldsymbol{\Phi}^{-1}\boldsymbol{\Sigma}\right)(1 + tr(\boldsymbol{\Sigma}\boldsymbol{\Theta})), \quad (4.43)
\end{aligned}$$

with $\boldsymbol{\Sigma}_4 \sim W_m(\boldsymbol{\Phi}, n+p)$ and $\boldsymbol{V}$ as defined in (2.7).

Proof. From (4.38) and R.49 follows that

$$\begin{aligned} q(\boldsymbol{\Sigma}|\boldsymbol{X}) &\propto |\boldsymbol{\Sigma}|^{\frac{n}{2}-\frac{n_1p}{2}-\frac{m+1}{2}} etr\left[-\frac{1}{2}\boldsymbol{\Sigma}^{-1}\boldsymbol{V}\right] etr\left(-\frac{1}{2}\boldsymbol{\Phi}^{-1}\boldsymbol{\Sigma}\right)(1+tr(\boldsymbol{\Sigma\Theta})) \\ &\quad \times \int_{\mathbb{R}^{m\times p}} etr\left[-\frac{1}{2}\boldsymbol{\Sigma}^{-1}n_1(\overline{\boldsymbol{X}}-\boldsymbol{\mu})\boldsymbol{\Omega}^{-1}(\overline{\boldsymbol{X}}-\boldsymbol{\mu})'\right] d\boldsymbol{\mu} \\ &\propto |\boldsymbol{\Sigma}|^{\frac{n+p}{2}-\frac{n_1p}{2}-\frac{m+1}{2}} etr\left[-\frac{1}{2}\boldsymbol{\Sigma}^{-1}\boldsymbol{V}\right] etr\left(-\frac{1}{2}\boldsymbol{\Phi}^{-1}\boldsymbol{\Sigma}\right)(1+tr(\boldsymbol{\Sigma\Theta})). \end{aligned}$$

Hence

$$q(\boldsymbol{\Sigma}|\boldsymbol{X}) = c_{\boldsymbol{\Sigma}}|\boldsymbol{\Sigma}|^{\frac{n+p}{2}-\frac{n_1p}{2}-\frac{m+1}{2}} etr\left[-\frac{1}{2}\boldsymbol{\Sigma}^{-1}\boldsymbol{V}\right] etr\left(-\frac{1}{2}\boldsymbol{\Phi}^{-1}\boldsymbol{\Sigma}\right)(1+tr(\boldsymbol{\Sigma\Theta})), \quad (4.44)$$

with

$$\begin{aligned} c_{\boldsymbol{\Sigma}}^{-1} &= \int_{S_m} |\boldsymbol{\Sigma}|^{\frac{n+p}{2}-\frac{n_1p}{2}-\frac{m+1}{2}} etr\left[-\frac{1}{2}\boldsymbol{\Sigma}^{-1}\boldsymbol{V}\right] etr\left(-\frac{1}{2}\boldsymbol{\Phi}^{-1}\boldsymbol{\Sigma}\right)(1+tr(\boldsymbol{\Sigma\Theta}))d\boldsymbol{\Sigma} \\ &= 2^{\frac{(n+p)m}{2}}\Gamma_m\left(\frac{n+p}{2}\right)|\boldsymbol{\Phi}|^{\frac{n+p}{2}} E_{\boldsymbol{\Sigma}_4}\left[|\boldsymbol{\Sigma}|^{-\frac{n_1p}{2}} etr\left[-\frac{1}{2}\boldsymbol{\Sigma}^{-1}\boldsymbol{V}\right](1+tr(\boldsymbol{\Sigma\Theta}))\right], \quad (4.45) \end{aligned}$$

from Remark 3.5.2 and Definition 2.1.1, with $\boldsymbol{\Sigma}_4 \sim W_m(\boldsymbol{\Phi}, n+p)$. Hence (4.43) follows from (4.44) and (4.45). □

Theorem 4.6.7. *The Bayes estimator of $\boldsymbol{\mu}$ for the normal model with prior (4.37) is*

$$\widehat{\boldsymbol{\mu}}_{B,KW} = \overline{\boldsymbol{X}}$$

Proof. From (4.40) note that

$$\begin{aligned} &E(\boldsymbol{\mu}-\overline{\boldsymbol{X}}) \\ &= \int_{\mathbb{R}^{m\times p}} (\boldsymbol{\mu}-\overline{\boldsymbol{X}}) \frac{2^{-mp}\left(\frac{\pi}{n_1}\right)^{-\frac{mp}{2}}\Gamma_m\left(\frac{n}{2}\right)|\boldsymbol{\Omega}|^{-\frac{m}{2}}|\boldsymbol{\Phi}|^{-\frac{p}{2}}}{\Gamma_m\left(\frac{n+p}{2}\right)E_{\boldsymbol{\Sigma}_4}\left[|\boldsymbol{\Sigma}_4|^{-\frac{n_1p}{2}} etr\left[-\frac{1}{2}\boldsymbol{\Sigma}_4^{-1}\boldsymbol{V}\right](1+tr(\boldsymbol{\Sigma}_4\boldsymbol{\Theta}))\right]} \\ &\quad \times E_{\boldsymbol{\Sigma}_3}\left[|\boldsymbol{\Sigma}_3|^{-\frac{n_1p}{2}} etr\left[-\frac{1}{2}\boldsymbol{\Sigma}_3^{-1}\left[\boldsymbol{V}+n_1(\overline{\boldsymbol{X}}-\boldsymbol{\mu})\boldsymbol{\Omega}^{-1}(\overline{\boldsymbol{X}}-\boldsymbol{\mu})'\right]\right](1+tr(\boldsymbol{\Sigma}_3\boldsymbol{\Theta}))\right]d\boldsymbol{\mu}. \end{aligned}$$

It is clear that this is the integral of an odd function, hence $E(\boldsymbol{\mu}-\overline{\boldsymbol{X}}) = \mathbf{0}$. Therefore, the Bayes estimator of $\boldsymbol{\mu}$ is

$$\widehat{\boldsymbol{\mu}}_{B,KW} = E(\boldsymbol{\mu}|\boldsymbol{X}) = \overline{\boldsymbol{X}}.$$

□

Theorem 4.6.8. *The Bayes estimator of $|\boldsymbol{\Sigma}|^r$ for the normal model with prior (4.37) is*

$$\widehat{|\boldsymbol{\Sigma}|^r}_{B,KW} = \frac{2^{\frac{mr}{2}}\Gamma_m\left(\frac{n+p+r}{2}\right)|\boldsymbol{\Phi}|^{\frac{r}{2}}E_{\boldsymbol{\Sigma}_5}\left[|\boldsymbol{\Sigma}_5|^{-\frac{n_1p}{2}} etr\left[-\frac{1}{2}\boldsymbol{\Sigma}_5^{-1}\boldsymbol{V}\right](1+tr(\boldsymbol{\Sigma}_5\boldsymbol{\Theta}))\right]}{\Gamma_m\left(\frac{n+p}{2}\right)E_{\boldsymbol{\Sigma}_4}\left[|\boldsymbol{\Sigma}_4|^{-\frac{n_1p}{2}} etr\left[-\frac{1}{2}\boldsymbol{\Sigma}_4^{-1}\boldsymbol{V}\right](1+tr(\boldsymbol{\Sigma}_4\boldsymbol{\Theta}))\right]}, \quad (4.46)$$

where $\mathbf{\Sigma}_4 \sim W_m(\mathbf{\Phi}, n+p)$, $\mathbf{\Sigma}_5 \sim W_m(\mathbf{\Phi}, n+p+r)$ *and* $\mathbf{V}$ *as defined in* (2.7).

Proof. From (4.43), Remark 3.5.2 and Definition 2.1.1 the posterior mean of $|\mathbf{\Sigma}|^r$ is

$$\begin{aligned}
& E\left(|\mathbf{\Sigma}|^r|\mathbf{X}\right) \\
= & \frac{2^{-\frac{(n+p)m}{2}}|\mathbf{\Phi}|^{-\frac{n+p}{2}}}{\Gamma_m\left(\frac{n+p}{2}\right)E_{\mathbf{\Sigma}_4}\left[|\mathbf{\Sigma}_4|^{-\frac{n_1p}{2}}etr\left[-\frac{1}{2}\mathbf{\Sigma}_4^{-1}\mathbf{V}\right]\left(1+tr(\mathbf{\Sigma}_4\mathbf{\Theta})\right)\right]} \\
& \times\int_{S_m}|\mathbf{\Sigma}|^{\frac{n+p+r}{2}-\frac{n_1p}{2}-\frac{m+1}{2}}etr\left[-\frac{1}{2}\mathbf{\Sigma}^{-1}\mathbf{V}\right]etr\left(-\frac{1}{2}\mathbf{\Phi}^{-1}\mathbf{\Sigma}\right)\left(1+tr(\mathbf{\Sigma}\mathbf{\Theta})\right)d\mathbf{\Sigma} \\
= & \frac{2^{-\frac{(n+p)m}{2}}|\mathbf{\Phi}|^{-\frac{n+p}{2}}}{\Gamma_m\left(\frac{n+p}{2}\right)E_{\mathbf{\Sigma}_4}\left[|\mathbf{\Sigma}_4|^{-\frac{n_1p}{2}}etr\left[-\frac{1}{2}\mathbf{\Sigma}_4^{-1}\mathbf{V}\right]\left(1+tr(\mathbf{\Sigma}_4\mathbf{\Theta})\right)\right]} \\
& \times\frac{\Gamma_m\left(\frac{n+p+1}{2}\right)}{2^{-\frac{(n+p+r)m}{2}}}|\mathbf{\Phi}|^{\frac{n+p+1}{2}}E_{\mathbf{\Sigma}_5}\left[|\mathbf{\Sigma}_5|^{-\frac{n_1p}{2}}etr\left[-\frac{1}{2}\mathbf{\Sigma}_5^{-1}\mathbf{V}\right]\left(1+tr(\mathbf{\Sigma}_5\mathbf{\Theta})\right)\right],
\end{aligned}$$

and the proof is complete. □

4.7 Numerical study

4.7.1 Univariate case

In this section, a special univariate case of the weighted-type I Wishart distribution, the Kummer gamma distribution (4.12 with $\phi = 1$), is applied as the prior for the variance of the univariate normal model.

4.7.1.1. Algorithm

In this section a Gibbs sampling algorithm is proposed to simulate samples from (4.29) and (4.32) as follows:

Algorithm 7

1. Initialize μ_0 and σ_0^2
2. Repeat the following steps for $t = 1,\dots,100\,000$ times:
   (a) Generate $\mu_t \sim N(\overline{X}, \frac{1}{n_1}\sigma_{t-1}^2)$
   (b) Calculate $A_t = \sum_{i=1}^{n_1}(X_i - \mu_t)^2$
   (c) Metropolis-Hastings algorithm:
      i. Generate the random numbers $\sigma^{2*} \sim G(\frac{n}{2}, \varphi^{-1})$
      ii. If $min\left(\frac{f^*[\sigma^{2*}|\mu_t]}{f^*[\sigma_{t-1}^2|\mu_t]}, 1\right) > u$ where u is a random uniform$(0,1)$ variate, then $\sigma_t^2 = \sigma^{2*}$ else $\sigma_t^2 = \sigma_{t-1}^2$, with
      $f^*[\sigma^2|\mu_t] \propto (\sigma^2)^{\frac{n}{2}-\frac{n_1}{2}-1}exp\left[-\frac{1}{2\varphi}\sigma^2\right](1+\sigma^2)^{\omega}exp\left[-\frac{1}{2\sigma^2}A_t\right]$ from (4.28).

3. Burn-in the first couple of observations, i.e. the posterior observations are $\mu_{1000},\ldots,\mu_{100\,000}$ and $\sigma_{1000},\ldots,\sigma_{100\,000}$.

4.7.1.2. Simulation study

A sample of size $n_1 = 18$ is simulated from a normal model with zero mean and unit variance, i.e. $X \sim N(0,1)$. For this study, prior (4.26) is assumed with $\omega = 1$. To evaluate the performance of this prior, posterior samples of μ and σ^2 can be simulated using Algorithm 7, and the Bayes estimates calculated or (4.35) and (4.36) can be used. Note that the prior expected value of σ^2 is

$$\begin{aligned} E(\sigma^2) &= \frac{(2\varphi)^{-\frac{n}{2}}}{\Gamma\left(\frac{n}{2}\right)(n\varphi+1)}\int_0^\infty (\sigma^2)^{\frac{n}{2}+1-1} exp\left(-\frac{1}{2\varphi}\sigma^2\right)(1+\sigma^2)d\sigma^2 \\ &= \frac{(2\varphi)^{-\frac{n}{2}}}{\Gamma\left(\frac{n}{2}\right)(n\varphi+1)}(2\varphi)^{\frac{n}{2}+1}\Gamma\left(\frac{n}{2}+1\right)(n\varphi+2\varphi+1) \\ &= \frac{n\varphi(1+(n+2)\varphi)}{1+n\varphi}, \end{aligned} \tag{4.47}$$

from (4.12) with $\omega = 1$, which is set equal to the prior belief $E(\sigma^2) = 0.9$.

Therefore, the hyperparameters are chosen as $n = 4$ and $\varphi = 0.185$ with $\omega = 1$. The MLE's of the parameters are given by

$$\widehat{\mu} = \overline{X} = 0.2405, \qquad \widehat{\sigma^2} = s^2 = 0.7604.$$

The Bayes estimates of the parameters calculated using the sample, from (4.35) and (4.36), respectively, are

$$\widehat{\mu}_{B,KG} = \overline{X} = 0.2405, \qquad \widehat{\sigma^2}_{B,KG} = 0.7939$$

It is evident from the comparison of the estimates that the Bayes estimate for σ^2 is closer to the target parameter value than the MLE. To calculate the coverage probability for this prior, 1000 samples are simulated from the underlying model and for each the 95% credible interval is calculated using (4.32). The median credible interval width of the 1000 intervals are calculated as 0.9 and the coverage probability as 92.1%. The marginal posterior density functions of μ and σ^2, (4.29) and (4.32), are given in Figure 4.3 (see code in Appendix D.3).

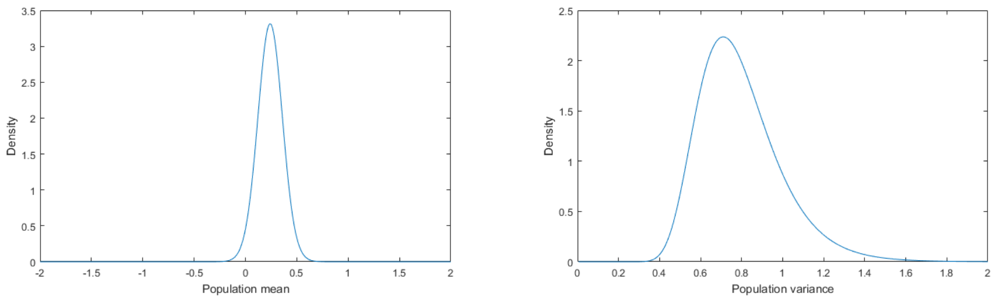


FIGURE 4.3: Marginal posterior density functions of μ (left) (4.29) and σ^2 (right) (4.32)

4.7.1.3. Forestry dataset

The forestry dataset used in Section 3.6.1 is revisited. Recall that the volume should be estimated based on a sample of 31 trees. The data is found to be non-normal and a log-transformation corrected this.

Let X be the volume of the tree and $Y = log(X)$ be the log-transformed variable, then $Y \sim N(\mu, \sigma^2)$. The prior (4.26) is applied with $\omega = 1$ and the prior information is specified as $E[\sigma^2] = 0.3$ based on expert information. The hyperparameters are thus chosen as $n = 4$ and $\varphi = 0.625$ such that $E[\sigma^2] = 0.3$.

The resulting MLE of σ^2 is 0.277 and the Bayes estimate is 0.2798 for the Kummer gamma prior with credible interval $(0.18; 0.40)$. The likelihood, prior and posterior density function of σ^2 for this dataset is displayed in Figure 4.4 (see code in Appendix D.4). Since the posterior is more concentrated than the prior, this prior is suitable for the dataset (see Lesaffre and Lawson (2012)).

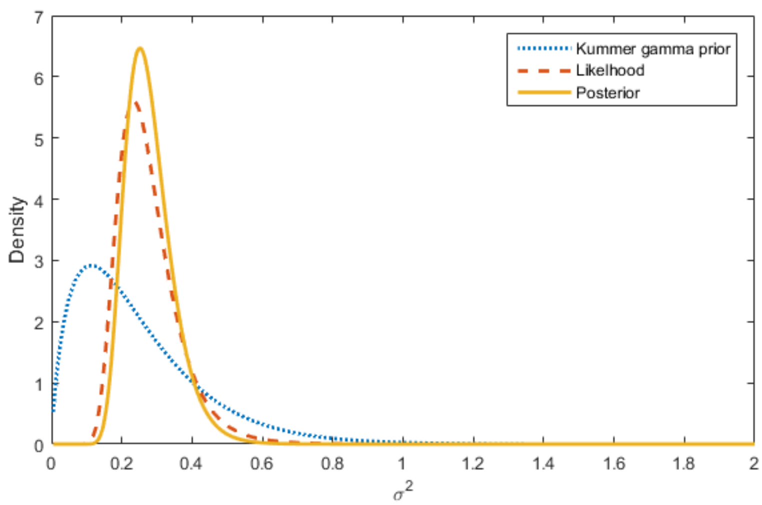


FIGURE 4.4: Likelihood, prior and analytical posterior density functions of σ^2

A comparative study using different priors for this dataset, is done in Section 5.2.2.

4.7.2 Multivariate case

In this section, the Kummer Wishart distribution (4.11) is applied as the prior for the covariance matrix of the matrix variate normal model.

4.7.2.1. Algorithm

The algorithms proposed in Section 2.5 are adjusted to develop a Gibbs sampling algorithm for the simulation of posterior samples of $\boldsymbol{\mu}$ and $\boldsymbol{\Sigma}$ under the prior in (4.26), i.e. $\boldsymbol{\Sigma} \sim KW_m(\boldsymbol{\Phi}, \boldsymbol{\Theta}, n)$ (see (4.11)) and $\pi(\boldsymbol{\mu}) = 1$. The algorithm for the simulation of samples from (4.40) and (4.43) is as follows:

Algorithm 8

1. Initialize $\boldsymbol{\mu}_0$ and $\boldsymbol{\Sigma}_0$
2. Repeat the following steps for $t = 1, ..., 100\,000$ times:
   (a) Generate $\boldsymbol{\mu}_t \sim N_{m,p}(\overline{\boldsymbol{X}}, \frac{1}{n_1}\boldsymbol{\Sigma}_{t-1} \otimes \boldsymbol{\Omega})$
   (b) Calculate $\boldsymbol{A}_t = \boldsymbol{V} + n_1(\overline{\boldsymbol{X}} - \boldsymbol{\mu}_t)(\overline{\boldsymbol{X}} - \boldsymbol{\mu}_t)'$
   (c) Metropolis-Hastings algorithm:
      i. Generate the random matrices $\boldsymbol{\Sigma}_1^* \sim W_m(\boldsymbol{\Phi}_1, m_5)$ and $\boldsymbol{\Sigma}_2^* \sim W_m^{-1}(\boldsymbol{\Phi}_2, m_5^*)$ such that $E[\boldsymbol{\Sigma}_1^*] = cE[\boldsymbol{\Sigma}_2^*]$.
      ii. Calculate $\boldsymbol{\Sigma}^* = w\boldsymbol{\Sigma}_1^* + (1-w)\boldsymbol{\Sigma}_2^*$ for some $0 < w < 1$
      iii. If $min\left(\frac{f^*[\boldsymbol{\Sigma}^*|\boldsymbol{\mu}_t]}{f^*[\boldsymbol{\Sigma}_{t-1}|\boldsymbol{\mu}_t]}, 1\right) > u$ where u is a random uniform$(0,1)$ variate, then $\boldsymbol{\Sigma}_t = \boldsymbol{\Sigma}^*$ else $\boldsymbol{\Sigma}_t = \boldsymbol{\Sigma}_{t-1}$, with $f^*[\boldsymbol{\Sigma}|\boldsymbol{\mu}_t] \propto |\boldsymbol{\Sigma}|^{-0.5(n-n_1p-m-1)}(1 + tr(\boldsymbol{\Sigma\Theta}))etr\left[-\frac{1}{2}\boldsymbol{\Sigma}^{-1}\boldsymbol{A}_t\right] etr\left[-\frac{1}{2}\boldsymbol{\Phi}^{-1}\boldsymbol{\Sigma}\right]$ from (4.39).
3. Burn-in the first couple of observations, i.e. the posterior observations are $\boldsymbol{\mu}_{1000}, ..., \boldsymbol{\mu}_{100\,000}$ and $\boldsymbol{\Sigma}_{1000}, ..., \boldsymbol{\Sigma}_{100\,000}$.

4.7.2.2. Simulation study

A sample of size 10 is simulated from a multivariate normal distribution with $m = 3$, a zero mean vector and identity covariance matrix i.e. $\boldsymbol{\mu} = \boldsymbol{0}$ and $\boldsymbol{\Sigma}_{m\times m} = \boldsymbol{I}_m$. The Kummer Wishart prior (4.11) is assumed in a Bayesian analysis to estimate the covariance matrix, i.e. $\boldsymbol{\Sigma} \sim KW_m(\boldsymbol{\Phi}, \boldsymbol{\Theta}, n)$ such that the joint prior is (4.26). The hyperparameters are chosen as $\boldsymbol{\Theta} = \boldsymbol{I}_m, \boldsymbol{\Phi} = \boldsymbol{I}_m$ and $n = m_5 = m_5^* = 3$. The initial values are chosen as $\boldsymbol{\mu}_0 = [0.5 \;\; 0.5 \;\; 0.5]$ and $\boldsymbol{\Sigma}_0 = 1.1\boldsymbol{I}_m$.

A posterior sample of size 5000 is simulated using Algorithm 8 (see code in Appendix D.5). The Bayes estimate is then calculated as the posterior mean of the sample. The MLE and the Bayes estimate for $\boldsymbol{\Sigma}$ is

$$\widehat{\boldsymbol{\Sigma}}_{MLE} = \begin{bmatrix} 1.8719 & 0.2168 & 0.9523 \\ 0.2168 & 2.9553 & -0.2471 \\ 0.9523 & -0.2471 & 1.0715 \end{bmatrix}$$

$$\widehat{\boldsymbol{\Sigma}}_{B,KW} = \begin{bmatrix} 1.1389 & 0.0115 & -0.0098 \\ 0.0115 & 1.0401 & -0.0132 \\ -0.0098 & -0.0132 & 1.0763 \end{bmatrix}$$

The Frobenius norm of the errors (see Remark 2.5.4) are calculated for each estimate and given in Table 4.4.

Frobenius norm	Value
$\lvert\lvert\widehat{\boldsymbol{\Sigma}}_{MLE} - \boldsymbol{\Sigma}\rvert\rvert_F$	1.2336
$\lvert\lvert\widehat{\boldsymbol{\Sigma}}_{KW} - \boldsymbol{\Sigma}\rvert\rvert_F$	0.1468

TABLE 4.4: Frobenius norm of the error of the estimates calculated from the simulated sample

The Kummer Wishart prior delivers an estimate with less error than the MLE and is hence favorable.

4.7.2.3. Abalone dataset

The abalone dataset, used in Section 3.6.2 is revisited. The aim is to estimate the age of abalone based on the number of rings derived from various other variables. A multivariate normal sample of size $n_1 = 20$ and dimension 6 is used. Hence, $\boldsymbol{X}_{1\times 6} \sim N(\boldsymbol{\mu}_{1\times 6}, \boldsymbol{\Sigma}_{6\times 6})$. The aim is to estimate $\boldsymbol{\Sigma}$.

Assume the prior (4.26), with $\boldsymbol{\Phi} = \boldsymbol{S}_1, \boldsymbol{\Theta} = 0.5\boldsymbol{S}_1, p = 6, n = 6$ where $\boldsymbol{S}_1$ is the sample covariance matrix of the available data excluding the 20 observations in the chosen subsample.

The Bayes estimate calculated for $\boldsymbol{\Sigma}$ under this prior using Algorithm 8 (see code in Appendix D.6), as well as the MLE are

$$\widehat{\boldsymbol{\Sigma}}_{MLE} = \begin{bmatrix} 0.014 & 0.012 & 0.004 & 0.054 & 0.015 & 0.012 \\ 0.012 & 0.01 & 0.003 & 0.045 & 0.013 & 0.01 \\ 0.004 & 0.003 & 0.002 & 0.017 & 0.005 & 0.004 \\ 0.054 & 0.045 & 0.017 & 0.24 & 0.065 & 0.052 \\ 0.015 & 0.013 & 0.005 & 0.065 & 0.019 & 0.014 \\ 0.012 & 0.01 & 0.004 & 0.052 & 0.014 & 0.012 \end{bmatrix}$$

$$\widehat{\boldsymbol{\Sigma}}_{B,KW} = \begin{bmatrix} 0.016 & 0.013 & 0.005 & 0.067 & 0.018 & 0.015 \\ 0.013 & 0.011 & 0.004 & 0.058 & 0.016 & 0.013 \\ 0.005 & 0.004 & 0.002 & 0.019 & 0.005 & 0.004 \\ 0.067 & 0.058 & 0.019 & 0.315 & 0.084 & 0.068 \\ 0.018 & 0.016 & 0.005 & 0.084 & 0.023 & 0.018 \\ 0.015 & 0.013 & 0.004 & 0.068 & 0.018 & 0.016 \end{bmatrix}$$

The convergence of Algorithm 8 is clear from Figure 4.5.

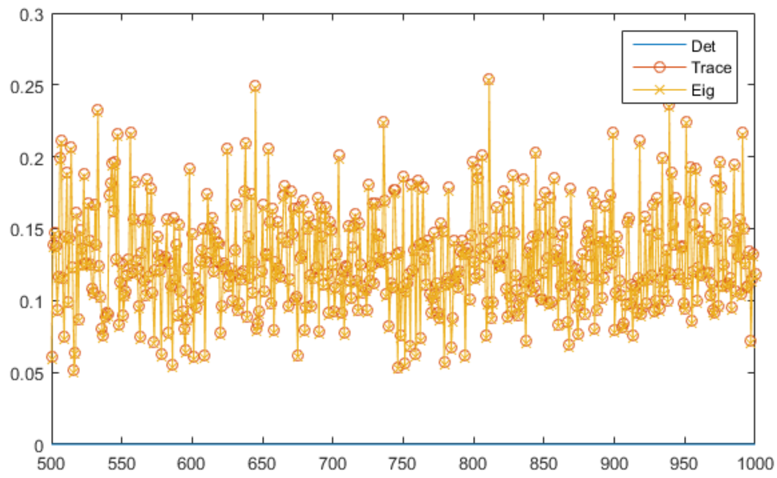


FIGURE 4.5: Convergence measures of Algorithm 8 for updates 500-1000 in the posterior sample

A comparative study using different priors for this dataset, is done in Section 5.3.2.

4.8 Conclusion

Due to the lack of matrix variate developments of weighted distributions, this chapter originated from the idea of coupling the Wishart density function with some weight functions. The main contributions in this chapter are summarized as follows:

Theoretical contributions

- The Weighted-type I, II and III Wishart distributions were defined.
- Various properties of the proposed distributions were derived and illustrated.
- Univariate and multivariate Bayesian analysis of the normal model were done with a special case of the W1WD as the prior, using an innovative approach circumventing the use of sampling algorithms for the univariate case.

Computational contributions

- New algorithms were developed for the application of a special case of the W1WD as a prior for the univariate and multivariate normal models.

Chapter 5

Comparative prior evaluation

In the preceding chapters, new matrix variate models within the Wishart ensemble were developed and special cases of these proposed distributions were applied as priors for the covariance matrix of the matrix variate normal model. The univariate counterparts of the new models were also considered as priors for the variance of the univariate normal model. In this chapter, the application of the new models in Bayesian analysis is elevated with a thorough comparative study, using simulated and real datasets. The algorithms proposed in the previous chapters for the Bayesian analysis of the normal model, are assessed using various convergence measures and different initial values.

5.1 Algorithms evaluation

5.1.1 Convergence of the simulated chains

The algorithms proposed throughout the study can be evaluated for convergence in two ways. Firstly single convergence measures, as used in the various chapters already and then secondly composite convergence measures (see code in Appendix E.1).

For a simulated sample of $\boldsymbol{\Sigma}$, the determinant $|\boldsymbol{\Sigma}_i|$, trace $tr(\boldsymbol{\Sigma}_i)$, and largest eigenvalues $\lambda_{1,i}$, of each simulated matrix are used as single convergence measures to evaluate the convergence of the chains. Following Gelman, Carlin, Stern, and Rubin 1995, the composite convergence measures consider the change in the average single convergence measures after each addition to the chain, using the average determinant,

$$\overline{|\boldsymbol{\Sigma}|}_j = \frac{1}{j}\sum_{i=1}^{j}|\boldsymbol{\Sigma}_i|,$$

average trace,

$$\overline{tr(\boldsymbol{\Sigma})}_j = \frac{1}{j}\sum_{i=1}^{j}tr(\boldsymbol{\Sigma}_i),$$

and the average largest eigenvalue

$$\overline{\lambda}_{1,j} = \frac{1}{j}\sum_{i=1}^{j}\lambda_{1,i}$$

of all simulated matrices preceding, and including, the latest update in the chain.

To evaluate the proposed algorithms, a sample of size $n_1 = 20$ is simulated from the multivariate normal model $N_m(\mathbf{0},\boldsymbol{I})$. The priors under consideration for $\boldsymbol{\Sigma}$ are

- Inverse Wishart prior $W_m^{-1}(\boldsymbol{\Phi}_{W^{-1}}, n_{W^{-1}})$ (Definition 2.1.2)
- Wishart prior $W_m(\boldsymbol{\Phi}_W, n_W)$ (Definition 2.1.1)
- Hypergeometric Wishart prior $HW_m(\boldsymbol{\Phi}_{HW}, a_1^*, b_1^*, c, n_{HW})$ (3.17)
- Kummer Wishart prior $KW_m(\boldsymbol{\Phi}_{KW}, \boldsymbol{I}, n_{KW})$ (4.11)

with $m = 3, \boldsymbol{\Phi}_{W^{-1}} = 4\boldsymbol{I}, n_{W^{-1}} = 9, \boldsymbol{\Phi}_W = 0.25\boldsymbol{I}, n_W = 3, \boldsymbol{\Phi}_{HW} = 0.5\boldsymbol{I}, a_1^* = 1, b_1^* = 2, c = 0.9, n_{HW} = 3, \boldsymbol{\Phi}_{KW} = 0.5\boldsymbol{I}, n_{KW} = 3$.
The initial values are $\boldsymbol{\mu}_0 = [0.5\ \ 0.5\ \ 0.5]$ and $\boldsymbol{\Sigma}_0 = 1.1\boldsymbol{I}$. A 10% burn-in period is used, hence the first 1000 observations are discarded. An improper objective prior (see R.55) for $\boldsymbol{\mu}, \pi(\mu) = 1$, is assumed.

Remark 5.1.1. *Note that in Algorithms 1 and 2, a subjective prior was assumed for $\boldsymbol{\mu}$. These algorithms can easily be adjusted for an improper objective prior for $\boldsymbol{\mu}$.*

Figure 5.1 illustrates the single convergence measures for Algorithms 1,2,6 and 8 while Figure 5.2 illustrates the composite convergence measures for Algorithms 1,2,6 and 8.

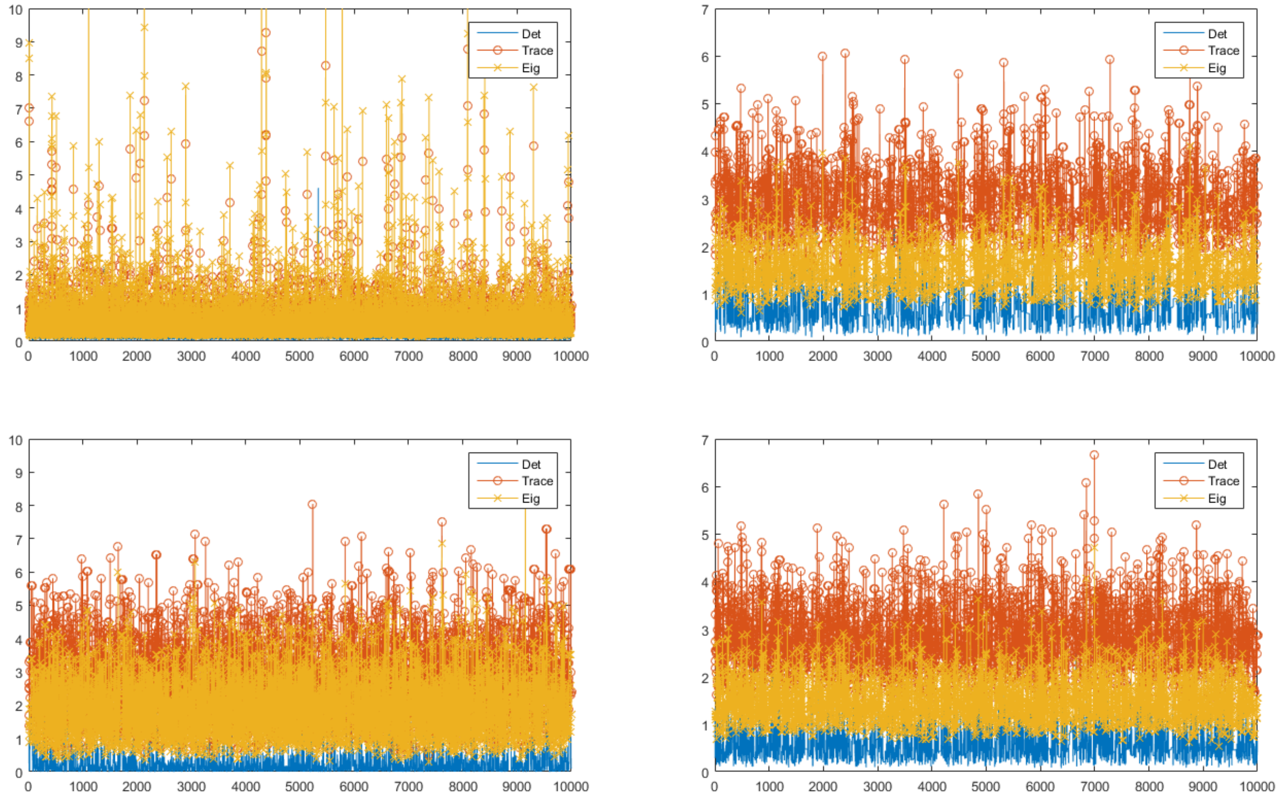


FIGURE 5.1: Single convergence measures for Algorithms 1 (top left), 2 (top right), 6 (bottom left) and 8 (bottom right)

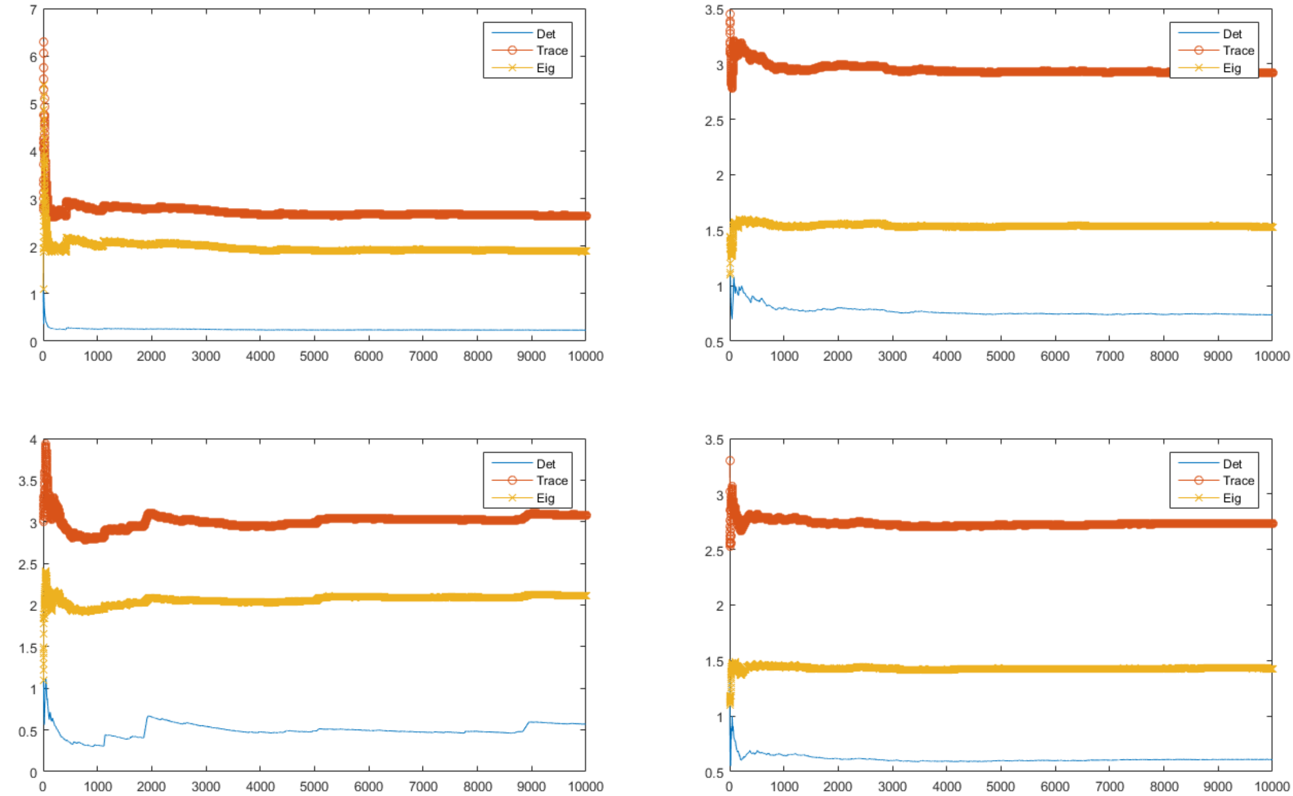


FIGURE 5.2: Composite convergence measures for Algorithms 1 (top left), 2 (top right), 6 (bottom left) and 8 (bottom right)

It is clear from Figures 5.1 and 5.2 that the proposed algorithms are stable and converge.

5.1.2 Effect of initial values

In the preceeding section some measures of convergence were introduced. Now, the effect of the initial values on the convergence of the chain is investigated. To this purpose, the simulation setup is repeated but with initial value $\boldsymbol{\Sigma}_0 = 5\boldsymbol{I}$.

Figure 5.3 illustrates the updated convergence measures for Algorithms 1,2,6 and 8.

The proposed algorithms are quite robust to initial values since from Figure 5.3, the effect of the initial value diminish very quickly within the 10% burn-in period, thus having no effect on the final estimates.

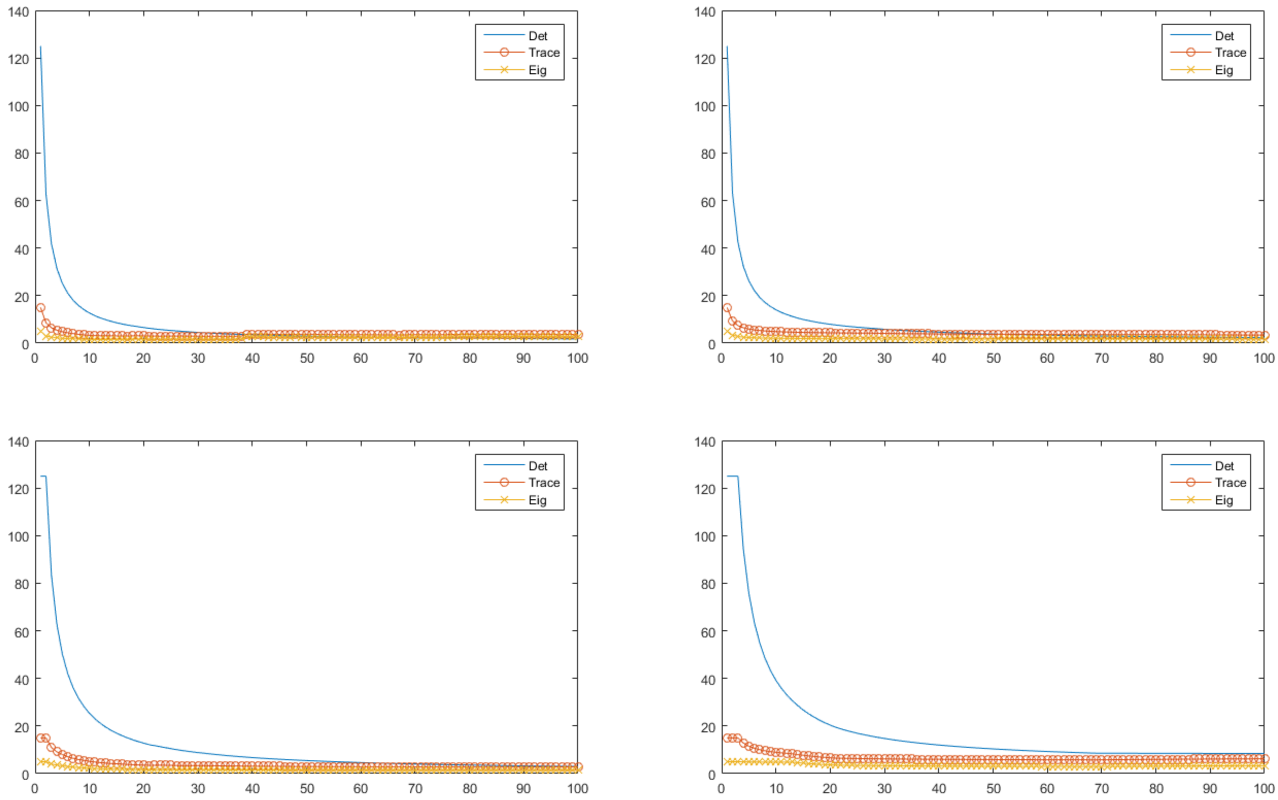


FIGURE 5.3: Composite convergence measures for the first 100 updates for Algorithms 1 (top left), 2 (top right), 6 (bottom left) and 8 (bottom right)

It is clear from Figure 5.3 that the proposed algorithms converge, for different initial values.

5.2 Univariate comparative analysis

5.2.1 Simulation study

In this section Bayesian analysis of the univariate normal is conducted (see code in Appendix E.3). A random sample of size $n_1 = 18$ is simulated from a standard normal distribution hence $X \sim N(0,1)$. The four priors for the variance under consideration are:

- Inverse gamma prior $\mathrm{IG}(\alpha_1, \beta_1)$ (R.40)
- Gamma prior $\mathrm{G}(\alpha_2, \beta_2)$ (R.41)
- Hypergeometric gamma prior $\mathrm{HG}(a,b,c,\phi)$ (Definition 3.4.1)
- Kummer gamma prior $\mathrm{KG}(\alpha_3, \beta_3, 1)$ (4.12)

5.2.1.1. Choice of hyperparameter values

The hyperparameters of the four priors are chosen according to the prior belief $E(\sigma^2) = \sigma_0^2 = 0.9$. The expected values of the four prior distributions are:

- Inverse gamma prior IG(α_1, β_1) (R.40): $\frac{\beta_1}{\alpha_1 - 1}$
- Gamma prior G(α_2, β_2) (R.41): $\frac{\alpha_2}{\beta_2}$
- Hypergeometric gamma prior HG(a, b, c, φ, n) ((3.19) and (3.47)): $\frac{n}{2}\varphi$
- Kummer gamma prior KG($\alpha_3, \beta_3, 1$) ((4.12) and (4.47)): $\frac{\alpha_3(1+\alpha_3+\beta_3)}{\beta_3(\alpha_3+\beta_3)}$

The four combinations of hyperparameters under consideration are given in Table 5.2 (see code in Appendix E.2). For each combination $E(\sigma^2) = \sigma_0^2 = 0.9$, except in combination 4, where $\sigma_0^2 = 6.8$ and $\sigma_0^2 = 4.75$ for the Kummer gamma and hypergeometric gamma priors, respectively. The new models are expected to perform poor with the hyperparameters in combination 4, since the prior belief is quite far from the target value. However, as shown in this section, the new models still perform well even under poor prior specification.

5.2.1.2. Results

To evaluate the performance of the priors, coverage probabilities and median credible interval length are used. For this purpose, 1000 independent samples of size $n_1 = 18$ are simulated and for each one the analytical posterior density functions (see Section 2.4.3.1 with $m = 1, p = 1$, (3.28) and (4.32)) under the four subjective priors and an objectve prior are calculated. This enables the calculation of the coverage probabilities and median credible interval width as given in Table 5.1.

Combination	Inverse gamma prior	Gamma prior	Hypergeometric gamma prior	Kummer gamma prior
1	0.87(0.9)	0.91(1)	0.99(0.8)	0.91(0.85)
2	0.88(0.9)	0.92(1)	0.97(0.8)	0.89(0.9)
3	0.90(0.9)	0.94(1)	0.98(0.8)	0.92(0.9)
4	0.88(0.9)	0.92(1)	0.89(1.3)	0.88(1.4)

TABLE 5.1: Coverage probabilities (median credible interval width) under the four priors

The coverage probability obtained under the gamma, Hypergeometric gamma and Kummer gamma priors are higher than for the well-known inverse-gamma prior, while the median width of the credible intervals (indicated in brackets) are competitive.

It is interesting to note that even under a completely false prior belief (see combination 4 in Table 5.2), the Hypergeometric gamma and Kummer gamma priors are still performing well.

The bias and mean squared error for each prior and each sample were calculated, for combinations 1 and 4, and are visually presented in Figure 5.4.

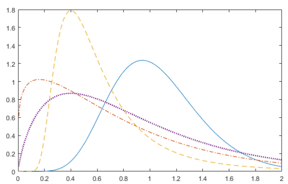
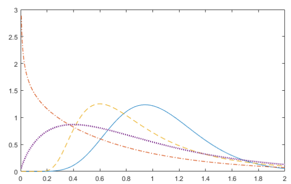
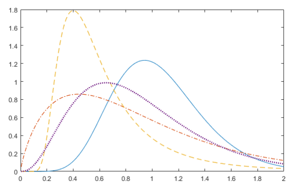
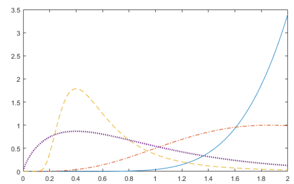
Combination	1	2	3	4
Inverse gamma prior	$\alpha_1 = 3.22, \beta_1 = 2$	$\alpha_1 = 4.33, \beta_1 = 3$	$\alpha_1 = 3.22, \beta_1 = 2$	$\alpha_1 = 3.22, \beta_1 = 2$
Gamma prior	$\alpha_2 = 1.8, \beta_2 = 2$	$\alpha_2 = 1.8, \beta_2 = 2$	$\alpha_2 = 3.6, \beta_2 = 4$	$\alpha_2 = 1.8, \beta_2 = 2$
Hypergeometric gamma prior	$a = 1, b = 2, c = 0.02, \varphi = 0.105, n = 19$	$a = 1, b = 2, c = 0.02, \varphi = 0.105, n = 19$	$a = 1, b = 2, c = 0.02, \varphi = 0.105, n = 19$	$a = 1, b = 2, c = 0.02, \varphi = 0.5, n = 19$
Kummer gamma prior	$\alpha_3 = 1.2, \beta_3 = 2.1$	$\alpha_3 = 0.8, \beta_3 = 1.8$	$\alpha_3 = 1.8, \beta_3 = 2.5$	$\alpha_3 = 5, \beta_3 = 2.5$
-.- Kummer gamma prior,- Hypergeometric gamma prior, - - Inverse gamma prior, ... Gamma prior				

TABLE 5.2: Influence of hyperparameters on the prior density functions

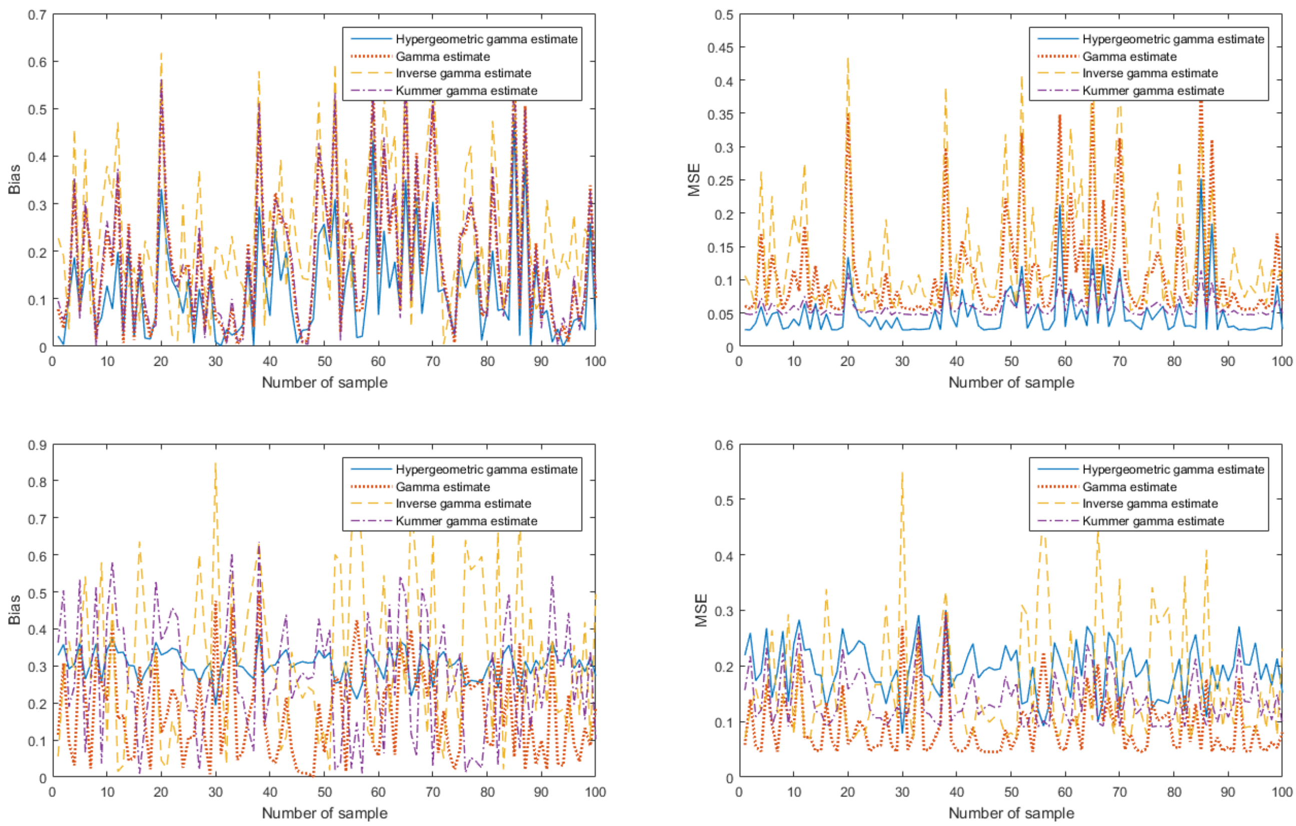


FIGURE 5.4: Bias (Left) and Mean squared error (MSE) (Right) for 100 samples of combination 1 (top row) and combination 4 (bottom row) (Hypergeometric gamma prior (-), Kummer gamma prior(-.-), gamma prior (...) and inverse-gamma prior (–))

The good performance of the hypergeometric gamma and Kummer gamma priors for combination 1 is evident from Figure 5.4 since the bias and MSE are small for these priors. More precise, and hence more accurate estimates are obtained for combination 1, compared to combination 4. This is supported by Table 5.1.

From Table 5.1 and Figure 5.4, the results of Griffin and Brown 2010 and Van Niekerk, Bekker, Arashi, and Roux 2015 are again apparent regarding the better performance of the gamma prior when compared to the inverse-gamma prior.

5.2.2 Forestry dataset

The forestry dataset used in previous chapters (see Sections 3.6.1 and 4.7.1) is revisited. Recall that the volume should be estimated based on a sample of 31 trees. The hypergeometric gamma, Kummer gamma, inverse gamma and gamma priors are considered.

The data is found to be non-normal and a log-transformation corrected this. Let X be the volume of the tree and $Y = log(X)$ be the log-transformed variable, then $Y \sim N(\mu, \sigma^2)$. The prior information is specified as $E[\sigma^2] = 0.3$ based on expert information. The hyperparameters are chosen as $\varphi = 0.01875, c = 0.01, n = 32, a = 1, b = 1, \alpha_1 = 11.5, \beta_1 = 3.5, \alpha_2 = 3, \beta_2 = 10, \alpha_3 = 4, \beta_3 = 0.8$, according to the methodology in Section 5.2.1.

The MLE of σ^2 is 0.277 and the Bayes estimates and credible intervals under the different priors are given in Table 5.3 (see code in Appendix E.4).

	Inverse gamma prior	Gamma prior	Hypergeometric gamma prior	Kummer gamma prior
Estimate	0.3125	0.3107	0.2933	0.2798
95% Credible interval	(0.235;0.52)	(0.195;0.465)	(0.21;0.385)	(0.18;0.40)

TABLE 5.3: Bayes estimates and credible intervals under the four subjective and one objective priors

The likelihood, prior and posterior density functions of σ^2 for this dataset is displayed in Figure 5.5.

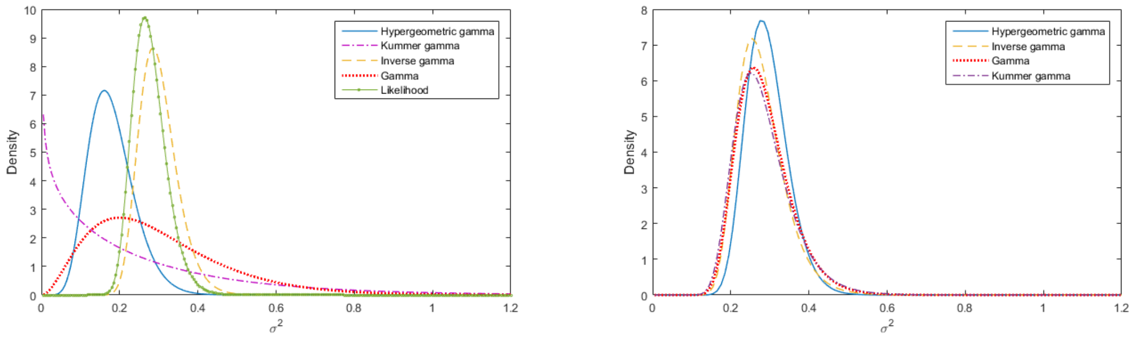


FIGURE 5.5: Likelihood and prior density functions (left) and analytical posterior density functions (right) of σ^2

The advantage of the hypergeometric and Kummer gamma priors are clearly seen in Figure 5.5 for the real dataset since these prior structures lead to a more concentrated posterior which in turn shows that the priors are not in conflict with the likelihood (see Lesaffre and Lawson 2012), and more than one prior may be appropriate.

The resulting credible intervals for the hypergeometric and Kummer gamma priors are narrower even though the prior is vaguer than the inverse gamma prior, leading to more precise estimates. The gamma prior still performs better than the inverse gamma prior.

5.3 Multivariate comparative analysis

5.3.1 Simulation study

A sample of size $n_1 = 20$ is simulated from a multivariate normal distribution with zero mean and identity covariance matrix i.e. $\boldsymbol{X} \sim N_3(\mathbf{0}, \boldsymbol{I})$ (see code Appendix E.5). The four priors under consideration are

- Inverse Wishart prior $W_3^{-1}(\boldsymbol{\Phi}_{W^{-1}}, n_{W^{-1}})$ (see Definition 2.1.2)
- Wishart prior $W_3(\boldsymbol{\Phi}_W, n_W)$ (see Definition 2.1.1)
- Hypergeometric Wishart prior $HW_3(\boldsymbol{\Phi}_{HW}, a_1, b_1, c, n_{HW})$ (see (3.17))
- Kummer Wishart prior $KW_3(\boldsymbol{\Phi}_{KW}, \boldsymbol{I}, n_{KW})$ (see (4.11))

5.3.1.1. Choice of hyperparameter values

Two different combinations of hyperparameter values are considered to assess the sensitivity of the posterior to the prior parameter values. **Combination 1** is $\boldsymbol{\Phi}_{W^{-1}} = 4\boldsymbol{I}_3, n_{W^{-1}} = 9, \boldsymbol{\Phi}_W = 0.25\boldsymbol{I}_3, n_W = 3, \boldsymbol{\Phi}_{HW} = 0.5\boldsymbol{I}_3, a_1 = 1, b_1 = 2, c = 0.9, n_{HW} = 3, \boldsymbol{\Phi}_{KW} = 0.5\boldsymbol{I}_3$ and $n_{KW} = 3$ whereas in **combination 2**; $n_W = 5, n_{HW} = 5$ and $n_{KW} = 5$.

The initial values are $\boldsymbol{\mu}_0 = [0.5 \;\; 0.5 \;\; 0.5]$ and $\boldsymbol{\Sigma}_0 = 1.1\boldsymbol{I}_3$. A 10% burn-in period is used.

5.3.1.2. Results

Posterior samples under each prior and each combination, of size 10 000, are simulated using Algorithms 1, 2, 6 and 8, respectively. The estimates calculated for $\boldsymbol{\Sigma}$ under the four different priors as well as the MLE with **combination 1** are

$$\widehat{\boldsymbol{\Sigma}}_{MLE} = \begin{bmatrix} 1.1845 & 0.5263 & 0.1733 \\ 0.5263 & 1.1293 & 0.2849 \\ 0.1733 & 0.2849 & 1.0234 \end{bmatrix}, \widehat{\boldsymbol{\Sigma}}_{B,W^{-1}} = \begin{bmatrix} 1.0123 & 0.4198 & 0.2996 \\ 0.4198 & 1.0762 & 0.2096 \\ 0.2996 & 0.2096 & 1.1746 \end{bmatrix}$$

$$\widehat{\boldsymbol{\Sigma}}_{B,W} = \begin{bmatrix} 1.2618 & 0.0913 & 0.3209 \\ 0.0913 & 0.9396 & 0.1519 \\ 0.3209 & 0.1519 & 1.5132 \end{bmatrix}, \widehat{\boldsymbol{\Sigma}}_{B,HW} = \begin{bmatrix} 0.9542 & 0.1392 & 0.1295 \\ 0.1392 & 1.0815 & 0.3494 \\ 0.1295 & 0.3494 & 1.1721 \end{bmatrix}$$

$$\widehat{\boldsymbol{\Sigma}}_{B,KW} = \begin{bmatrix} 0.9406 & 0.1588 & 0.1085 \\ 0.1588 & 0.9771 & 0.0655 \\ 0.1085 & 0.0655 & 0.9256 \end{bmatrix}$$

The estimates calculated for $\boldsymbol{\Sigma}$ under the four different priors as well as the MLE with **combination 2** are

$$\widehat{\mathbf{\Sigma}}_{MLE} = \begin{bmatrix} 0.9881 & -0.1275 & 0.2502 \\ -0.1275 & 1.6800 & -0.0587 \\ 0.2502 & -0.0587 & 0.5214 \end{bmatrix}, \widehat{\mathbf{\Sigma}}_{B,W^{-1}} = \begin{bmatrix} 0.9461 & -0.1330 & 0.2379 \\ -0.1330 & 1.5978 & -0.0503 \\ 0.2379 & -0.0503 & 0.5307 \end{bmatrix}$$

$$\widehat{\mathbf{\Sigma}}_{B,W} = \begin{bmatrix} 0.9940 & -0.0285 & 0.0765 \\ -0.0285 & 1.1844 & -0.0161 \\ 0.0765 & -0.0161 & 0.7896 \end{bmatrix}, \widehat{\mathbf{\Sigma}}_{B,HW} = \begin{bmatrix} 1.2754 & -0.0230 & 0.0248 \\ -0.0230 & 1.2607 & -0.0628 \\ 0.0248 & -0.0628 & 1.2549 \end{bmatrix}$$

$$\widehat{\mathbf{\Sigma}}_{B,KW} = \begin{bmatrix} 0.9868 & 0.0168 & 0.2119 \\ 0.0168 & 1.4037 & 0.0335 \\ 0.2119 & 0.0335 & 0.6645 \end{bmatrix}$$

The above estimates are obtained for one posterior sample under each prior and combination. The Frobenius norm of the errors (see Remark 2.5.4) is calculated for each estimate and the two different combinations and given in Table 5.4.

Frobenius norm	Combination 1	Combination 2
$\|\|\widehat{\mathbf{\Sigma}}_{MLE}-\mathbf{\Sigma}\|\|_F$	0.8156	0.7120
$\|\|\widehat{\mathbf{\Sigma}}_{W^{-1}}-\mathbf{\Sigma}\|\|_F$	0.7075	0.6327
$\|\|\widehat{\mathbf{\Sigma}}_{W}-\mathbf{\Sigma}\|\|_F$	0.7699	0.2360
$\|\|\widehat{\mathbf{\Sigma}}_{HW}-\mathbf{\Sigma}\|\|_F$	0.5403	0.3387
$\|\|\widehat{\mathbf{\Sigma}}_{KW}-\mathbf{\Sigma}\|\|_F$	0.2154	0.4411

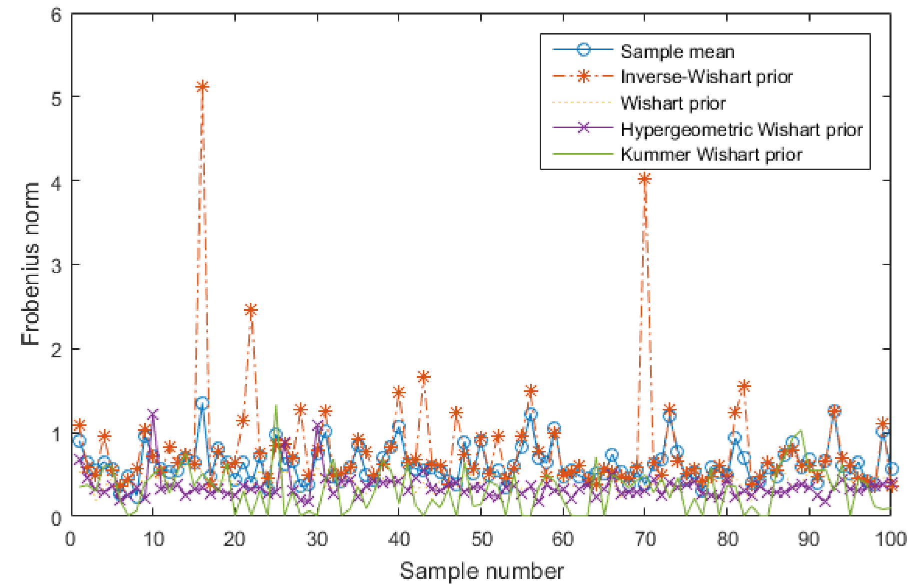
TABLE 5.4: Frobenius norm of the error of the estimates calculated from the simulated samples

It is clear from Table 5.4 that the Wishart, hypergeometric Wishart and Kummer Wishart priors perform well when compared with the sample estimate and the well-known inverse Wishart prior.

These results however are for just one posterior sample of size 10 000. 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 using both combinations of hyperparameter values.

The Frobenius norm of the error for each estimate and every repetition is calculated and given in Figure 5.6 for the 100 samples.

Additionally, the empirical cumulative distribution function (ecdf) of each set of Frobenius norms calculated for each estimator, is obtained and given in Figure 5.7 for the 100 samples. The ecdf which is 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.



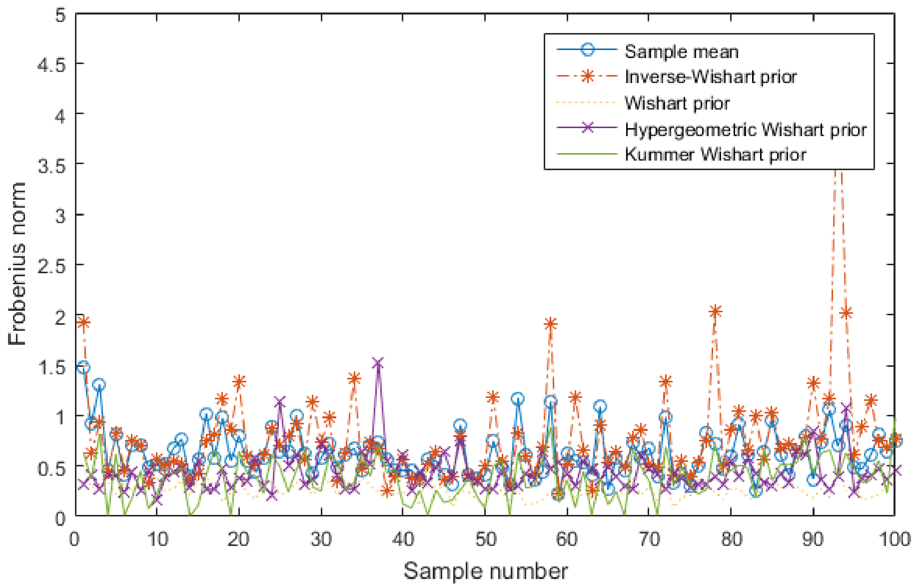


FIGURE 5.6: The Frobenius norm for the five estimates for combination 1 (top) and combination 2 (bottom)

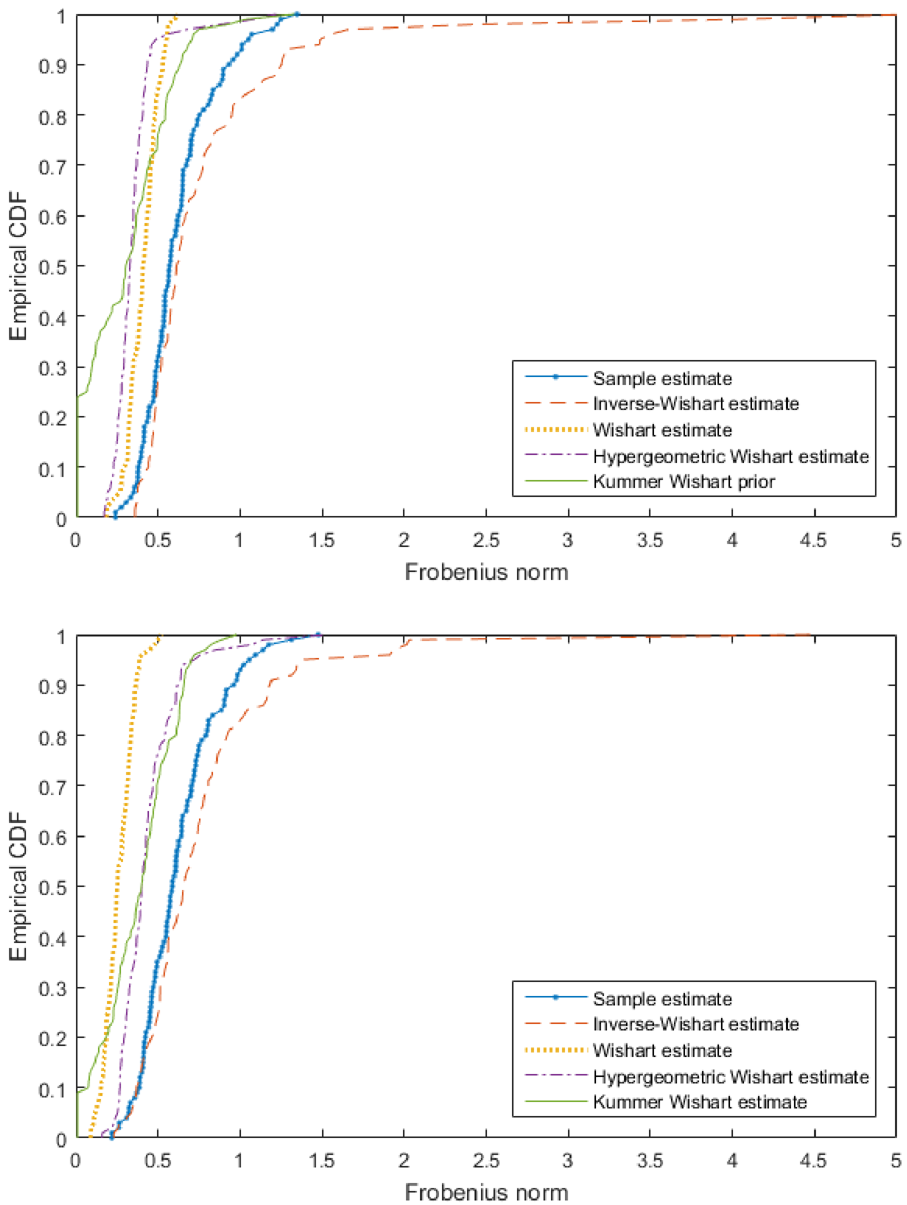


FIGURE 5.7: The empirical cumulative distribution function (ecdf) of the Frobenius norm of the estimation errors for combination 1 (top) and combination 2 (bottom)

The superior performance of the hypergeometric Wishart and Kummer Wishart priors are evident from Figures 5.6 and 5.7.

This simulation study was based on a small sample of size 20. For further investigation, the sample size, n_1, is increased from 20 to 100, and the ecdf of all the resulting estimates are given in Figure 5.8 for combination 1.

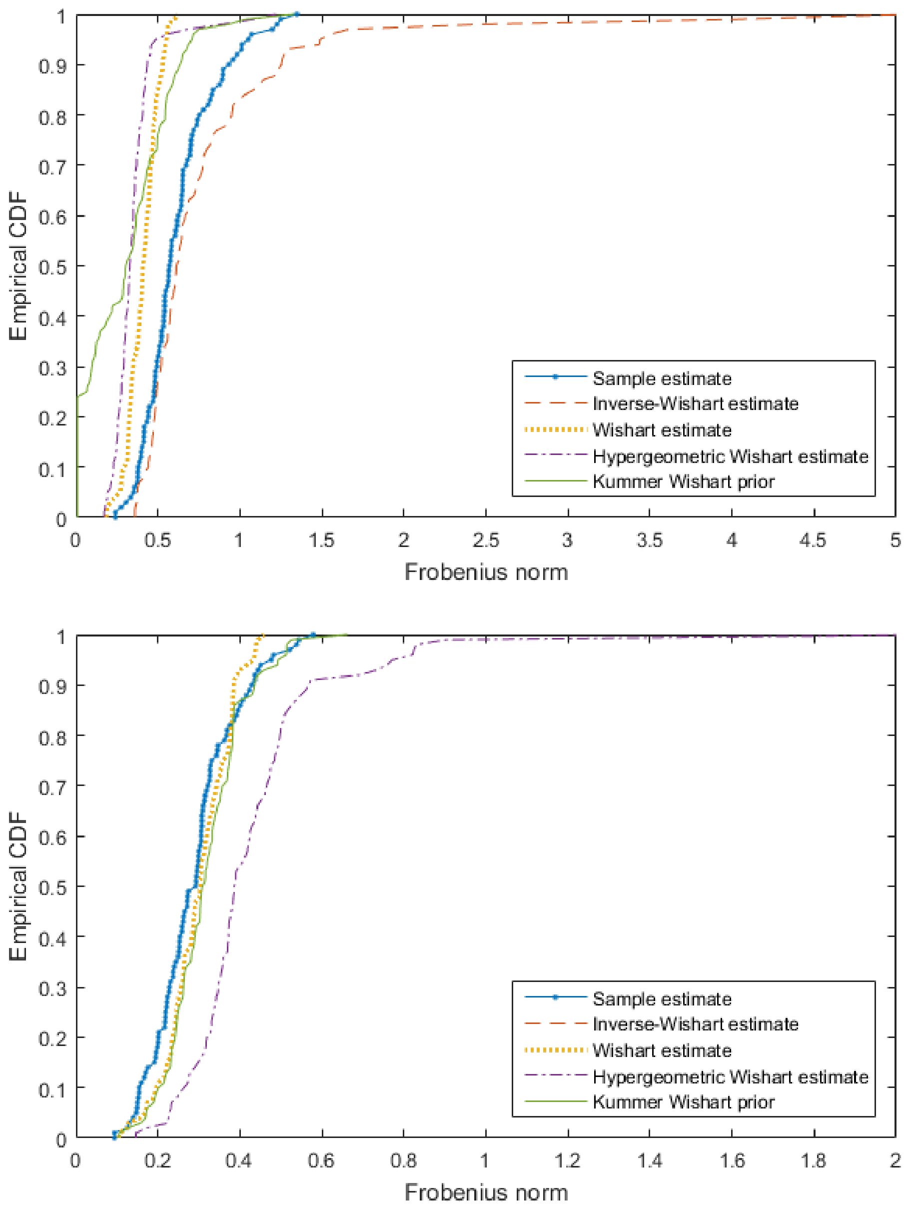


FIGURE 5.8: The empirical cumulative distribution function (ecdf) of the Frobenius norm of the estimation errors for $n = 20$ (top) and $n = 100$ (bottom) for combination 1

Note that due to poor performance, the ecdf under the inverse Wishart is omitted from the bottom graph in Figure 5.8.

It is evident from Figure 5.8 that the performance of the MLE improves as the sample size increases, which is to be expected, and the performance of the Kummer Wishart and Wishart priors are still competitive.

To validate the graphical interpretation, a two-sample Kolmogorov-Smirnov test is performed for $n_1 = 20$ using the hyperparameter values in combination 1, pairwise, on the five different ecdf's and the p-values from the tests are given in Table 5.5.

Pairwise comparison	p-value	Pairwise comparison	p-value
MLE and W^{-1}	< 0.001	W^{-1} and HW	< 0.001
MLE and W	< 0.001	W^{-1} and KW	0.482
MLE and HW	< 0.001	W and HW	< 0.384
MLE and KW	< 0.001	W and KW	< 0.510
W^{-1} and W	< 0.001	HW and KW	< 0.415

TABLE 5.5: p-values of the Kolmogorov-Smirnov two-sample test of the ecdf's based on samples ($n = 20$) of the Frobenius norms

From Table 5.5 it is clear that the ecdf of the errors under the hypergeometric Wishart, Kummer Wishart and Wishart priors are significantly different from the others. Therefore, the assertion can be made that the hypergeometric Wishart, Kummer Wishart and Wishart priors produce estimates that result in statistically significant less error and can be used in future.

McElreath (2016) stated: "People commonly ask what the correct prior is for a given analysis. The question sometimes implies that for a given set of data, there is a uniquely correct prior that must be used, or else the analysis will be invalid. This is a mistake. There is no more a uniquely correct prior than there is a uniquely correct likelihood."

This section illustrated that more than one prior may be appropriate for a specific situation. The priors and algorithms proposed in this thesis provide the practitioner with a variety of choices and eased implementation, with promising results.

5.3.2 Abalone dataset

The abalone dataset, used in previous Chapters (see Sections 3.6.2 and 4.7.2), is revisited. The aim is to estimate the age of abalone based on the number of rings derived from various other variables. A multivariate normal sample of size $n_1 = 20$ and dimension 6 is used.

Hence, $\boldsymbol{X}_{1\times 6} \sim N(\boldsymbol{\mu}_{1\times 6}, \boldsymbol{\Sigma}_{6\times 6})$. A Bayesian approach is followed with the four priors for $\boldsymbol{\Sigma}$,

- Inverse Wishart prior $W_6^{-1}(\boldsymbol{\Phi}_{W^{-1}}, n_{W^{-1}})$ (Definition 2.1.2)
- Wishart prior $W_6(\boldsymbol{\Phi}_W, n_W)$ (Definition 2.1.1)
- Hypergeometric Wishart prior $HW_6(\boldsymbol{\Phi}_{HW}, a_1^*, b_1^*, c, n_{HW})$ (3.17)
- Kummer Wishart prior $KW_6(\boldsymbol{\Phi}_{KW}, \boldsymbol{I}, n_{KW})$ (4.11)

The hyperparameters are chosen as $\boldsymbol{\Phi}_{W^{-1}} = 4\boldsymbol{S}_1, n_{W^{-1}} = 13, \boldsymbol{\Phi}_W = 0.25\boldsymbol{S}_1, n_W = 6, \boldsymbol{\Phi}_{HW} = 0.5\boldsymbol{S}_1, a_1^* = 1, b_1^* = 2, c = 0.9, n_{HW} = 6, \boldsymbol{\Phi}_{KW} = 0.5\boldsymbol{S}_1$ and $n_{KW} = 6$, with $\boldsymbol{S}_1$ equal to the sample covariance matrix of the available data excluding the 20 observations in the chosen subsample.

The initial values are $\boldsymbol{\mu}_0 = [0.5 \quad 0.5 \quad 0.5]$ and $\boldsymbol{\Sigma}_0 = 1.1\boldsymbol{I}_6$. The convergence of the chains were confirmed visually.

The Bayes estimates calculated for $\boldsymbol{\Sigma}$ under the four priors using Algorithms 1, 2, 6 and 8 (see code in Appendix E.6) as well as the MLE are

$$\widehat{\boldsymbol{\Sigma}}_{MLE} = \begin{bmatrix} 0.014 & 0.012 & 0.004 & 0.054 & 0.015 & 0.012 \\ 0.012 & 0.01 & 0.003 & 0.045 & 0.013 & 0.01 \\ 0.004 & 0.003 & 0.002 & 0.017 & 0.005 & 0.004 \\ 0.054 & 0.045 & 0.017 & 0.24 & 0.065 & 0.052 \\ 0.015 & 0.013 & 0.005 & 0.065 & 0.019 & 0.014 \\ 0.012 & 0.01 & 0.004 & 0.052 & 0.014 & 0.012 \end{bmatrix}$$

$$\widehat{\boldsymbol{\Sigma}}_{B,W^{-1}} = \begin{bmatrix} 0.006 & 0.005 & 0.002 & 0.026 & 0.001 & 0.006 \\ 0.005 & 0.004 & 0.002 & 0.022 & 0.006 & 0.005 \\ 0.002 & 0.002 & 0.001 & 0.008 & 0.002 & 0.002 \\ 0.026 & 0.022 & 0.008 & 0.120 & 0.033 & 0.027 \\ 0.007 & 0.006 & 0.002 & 0.033 & 0.009 & 0.007 \\ 0.006 & 0.005 & 0.002 & 0.027 & 0.007 & 0.006 \end{bmatrix}$$

$$\widehat{\boldsymbol{\Sigma}}_{B,W} = \begin{bmatrix} 0.009 & 0.008 & 0.003 & 0.038 & 0.01 & 0.008 \\ 0.008 & 0.006 & 0.002 & 0.032 & 0.009 & 0.007 \\ 0.003 & 0.002 & 0.001 & 0.011 & 0.003 & 0.003 \\ 0.038 & 0.032 & 0.011 & 0.186 & 0.047 & 0.038 \\ 0.01 & 0.009 & 0.003 & 0.047 & 0.013 & 0.01 \\ 0.008 & 0.007 & 0.003 & 0.038 & 0.01 & 0.009 \end{bmatrix}$$

$$\widehat{\boldsymbol{\Sigma}}_{B,HW} = \begin{bmatrix} 0.011 & 0.009 & 0.003 & 0.045 & 0.012 & 0.01 \\ 0.009 & 0.008 & 0.003 & 0.038 & 0.01 & 0.008 \\ 0.003 & 0.003 & 0.001 & 0.013 & 0.004 & 0.003 \\ 0.045 & 0.038 & 0.013 & 0.209 & 0.055 & 0.045 \\ 0.012 & 0.01 & 0.004 & 0.055 & 0.015 & 0.012 \\ 0.01 & 0.008 & 0.003 & 0.045 & 0.012 & 0.011 \end{bmatrix}$$

$$\widehat{\boldsymbol{\Sigma}}_{B,KW} = \begin{bmatrix} 0.016 & 0.013 & 0.005 & 0.067 & 0.018 & 0.015 \\ 0.013 & 0.011 & 0.004 & 0.058 & 0.016 & 0.013 \\ 0.005 & 0.004 & 0.002 & 0.019 & 0.005 & 0.004 \\ 0.067 & 0.058 & 0.019 & 0.315 & 0.084 & 0.068 \\ 0.018 & 0.016 & 0.005 & 0.084 & 0.023 & 0.018 \\ 0.015 & 0.013 & 0.004 & 0.068 & 0.018 & 0.016 \end{bmatrix}$$

The true covariance matrix of the dataset is unknown and error norms can therefore not be used as a measure of performance as was done in Section 5.3.1. The following approach is proposed as a measure of performance of the estimates based on a real dataset.

Based on the estimates of $\mathbf{\Sigma}$, $\widehat{\mathbf{\Sigma}}$, a multivariate normal dataset of size 20 is simulated for each estimate, i.e. $\boldsymbol{Y}_{i,1\times 6} \sim N_6(\overline{\boldsymbol{X}}, \widehat{\mathbf{\Sigma}}_Q), Q = (MLE, W^{-1}, W, HW, KW)$, and visually presented as an image where each observation forms a row and each variable a column of pixels of the image. Each entry of the simulated sample is used as the intensity of a pixel. This allows for visual inspection of the plausibility of the estimates. Figure 5.9 displays the images based on the real dataset and the simulated samples using the calculated estimates.

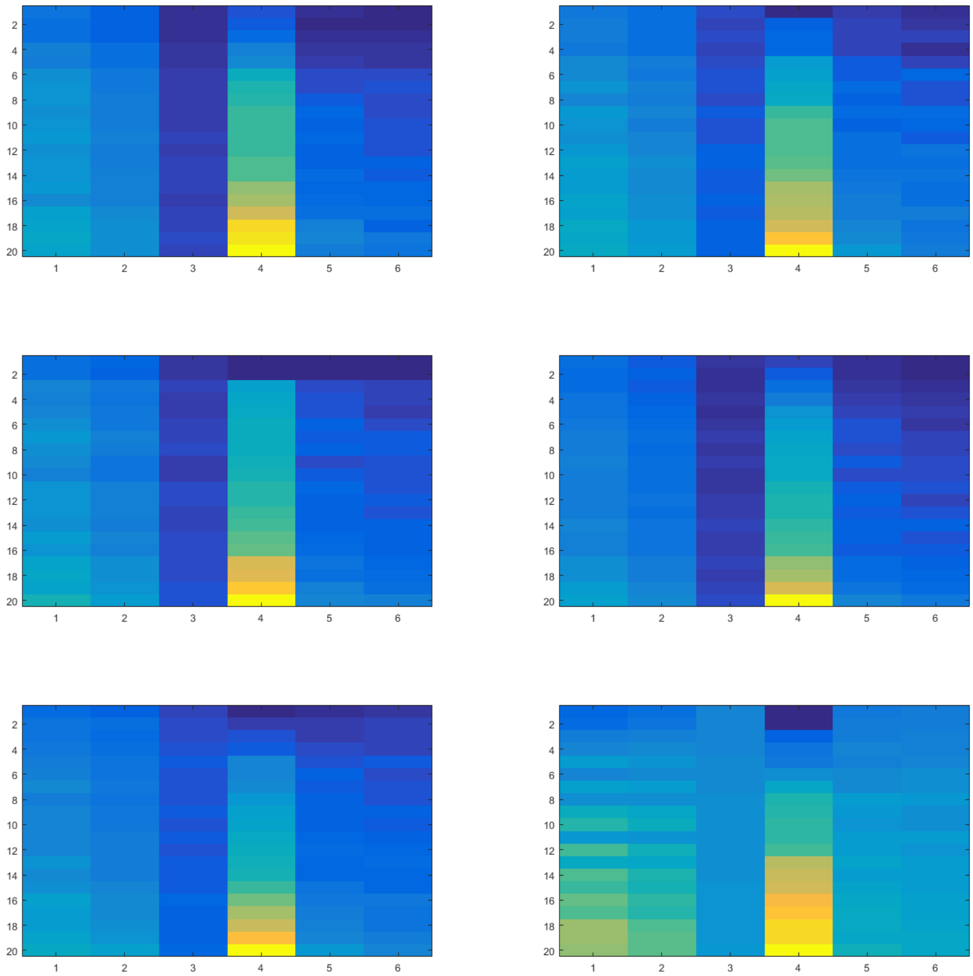


FIGURE 5.9: Images based on the real dataset (top left) and simulated samples - based on MLE (top right), HW (middle left), KW (middle right), W (bottom left), W^{-1} (bottom right)

It is clear from Figure 5.9, that the samples based on $\widehat{\mathbf{\Sigma}}_{B,KW}$ and $\widehat{\mathbf{\Sigma}}_{B,HW}$, the estimates of $\mathbf{\Sigma}$ under the Kummer Wishart and hypergeometric Wishart priors, respectively, are the closest to the real dataset. It is thus concluded that the Kummer Wishart and hypergeometric Wishart Bayes estimates are more plausible estimates of $\mathbf{\Sigma}$ for this dataset, than the MLE, since they reproduce a sample closer to the original datset.

For further investigation, the Mahalanobis distance (see R.18) 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 5.6.

Estimate used	MLE	W^{-1}	W	HW	KW
Mean Mahalanobis distance	17.3247	29.0114	7.0470	5.4523	5.2785

TABLE 5.6: Mean Mahalanobis distance between the real dataset and each simulated sample

Table 5.6 supports the findings of Figure 5.9 since the distance between the real dataset and the simulated sample based on $\widehat{\mathbf{\Sigma}}_{KW}$ is the smallest. Thus, the estimate under the Kummer Wishart prior, $\widehat{\mathbf{\Sigma}}_{KW}$, fits the data best. The estimates under the hypergeometric Wishart and Wishart priors are plausible from Figure 5.9 and Table 5.6. These results support the results from Section 5.3.1.

5.4 Conclusion

The focus of this chapter was the illustration and evaluation of models developed in this thesis, as priors for the univariate and multivariate normal models. The algorithms introduced in each chapter was evaluated with different convergence measures for different initial values and found to be stable. Simulation studies and real datasets were used to evaluate and compare the different priors for various combinations of hyperparameter values. The new priors were shown to perform well and can therefore be applied in future. The main results in this chapter are summarized as follows:

- Single and composite measures of convergence were defined and illustrated for the Algorithms 1,2,6 and 8.
- The algorithms developed in Chapters 2-4, were shown to be stable and quite robust with respect to different inital values.
- The new univariate models, the hypergeometric gamma and Kummer gamma distributions were shown to be competitive priors in the Bayesian analysis of the univariate normal model for different hyperparameter values.
- The matrix variate models, the hypergeometric Wishart and Kummer Wishart distributions were shown to perform well in the Bayesian analysis of the multivariate normal model for different combinations of hyperparameter values.

Chapter 6

Conclusion and future directions

Multivariate statistics are receiving ever increasing attention and are growing in popularity under statisticians and practitioners, alike. The availability of computing power makes it possible to analyse more extended data systems of multivariate type, hence the need for more complex models.

The Wishart distribution (Chapter 2) is amongst the most popular multivariate distributions and various extensions have been proposed in literature. In this study, the Wishart distribution is the foundation for further developments, and therefore the title is as such developments within Wishart ensemble.

Two general forms were developed within the Wishart ensemble. Firstly, the hypergeometric Wishart generator in Chapter 3, which generalizes the Wishart model by using a Borel measurable function and the generalised hypergeometric function of matrix argument. This model produces special cases known in literature as well as unknown new models. Secondly, weighted-type Wishart distributions are the focus of Chapter 4. These are constructed by introducing a weight function to the Wishart model. Three different weighted-type Wishart distributions were considered, each with some properties.

The advantages of the new models were illustrated through some properties and the application as priors for the normal model. In each Chapter, special cases of the new models were considered as priors for the variance and covariance matrix of the univariate and matrix variate normal models, respectively. The Bayesian results are promising, and support the advocation of the new models.

Numerical algorithms are developed for the application of the new models as matrix variate priors. These algorithms are further investigated in Chapter 5, where convergence, and the effect of initial values on the convergence are considered. The algorithms are shown to be stable and quite robust, and can therefore be used with ease. The effect of different hyperparameter values on the posterior results are illustrated, and the priors are shown to be preferable when compared to the well known inverse Wishart prior for the covariance matrix of the multivariate normal model using simulation studies and real datasets.

This study paves the way for future consideration of more complicated prior structures to improve on current Bayesian models. The computational advancements presented in this study, aids in the practical use of multivariate models.

There are many opportunities for future research based on this study, to name a few:

- Frequentist estimates of the proposed models can be derived and used to model phenomena leading to multivariate data.
- The noncentral Wishart generator distribution (see (3.14)) can be explored and applied in hypothesis testing regarding elliptical models (see Caro-Lopera, Gonzàlez-Farìas, and Balakrishnan (2014)).
- Following the methodology in Chapter 4, other weighted-type matrix variate distributions can be developed using other matrix variate distributions for the main density function in (4.2)-(4.4).
- In Chapter 4 different weight functions can be explored. A weight function of eigenvalues is of particular interest.
- The Bayesian application can be extended two-fold: the assumption of the underlying normal model can be replaced by the elliptical model (see Chapter 2) and other special cases of the HWGD (see Chapter 3) and Weighted-type Wishart distributions (see Chapter 4) can be considered as priors. This is illustrated in Figure 6.1.
- Other possible numerical algorithms for the Bayesian analysis can be investigated.

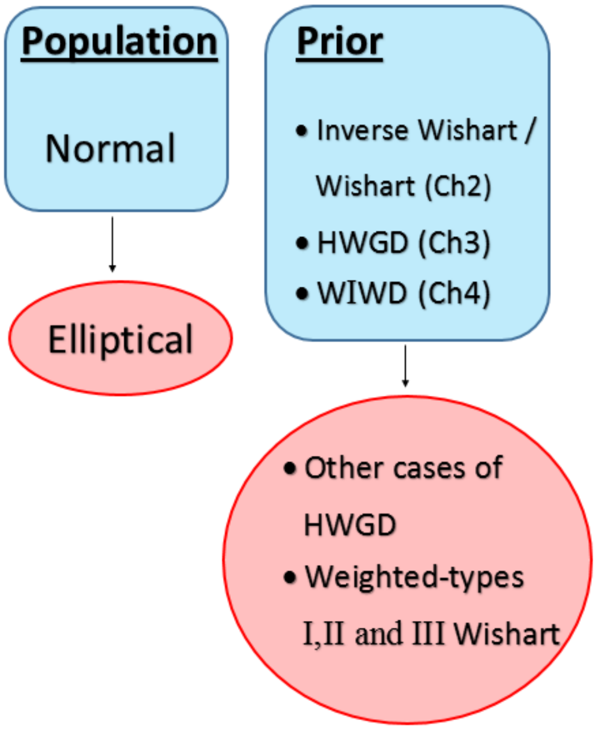


FIGURE 6.1: Graphical illustration of future directions regarding the Bayesian application

"Statisticians, like artists, have the bad habit of falling in love with their models..."

- George E.P. Box (1919-2013)

Appendix A

Preliminaries

A.1 Matrix theory and Jacobians

R. 1. *(Gupta and Nagar 2000, p.12)*
Let $\boldsymbol{X}_1,...,\boldsymbol{X}_k$ be matrix variates with n independent elements and joint density function $f(\boldsymbol{X}_1,...,\boldsymbol{X}_k)$. Consider the matrix transformations

$$\boldsymbol{Y}_i = t_i(\boldsymbol{X}_1,...,\boldsymbol{X}_k), i = 1,...,m$$

where the inverse set of transformations exist as

$$\boldsymbol{X}_j = t_j'(\boldsymbol{Y}_1,...,\boldsymbol{Y}_m), j = 1,...,k$$

hence $\boldsymbol{Y}_1,...,\boldsymbol{Y}_m$ are m matrix variates with n independent elements. The joint density function of $\boldsymbol{Y}_1,...,\boldsymbol{Y}_m$ is then given by

$$g(\boldsymbol{Y}_1,...,\boldsymbol{Y}_m) = f(t_1'(\boldsymbol{Y}_1,...,\boldsymbol{Y}_m),...,t_k'(\boldsymbol{Y}_1,...,\boldsymbol{Y}_m)).J((\boldsymbol{X}_1,...,\boldsymbol{X}_k) \to (\boldsymbol{Y}_1,...,\boldsymbol{Y}_m)),$$

where $J((\boldsymbol{X}_1,...,\boldsymbol{X}_k) \to (\boldsymbol{Y}_1,...,\boldsymbol{Y}_m))$ is the Jacobian of the transformation.

R. 2. *(Gupta and Nagar 2000, Eq.1.3.6, p.13)*
If $\boldsymbol{X}_{p\times p}$ and $\boldsymbol{Y}_{p\times p}$ are symmetric matrices and $\boldsymbol{Y} = c\boldsymbol{X}$ where $c \neq 0$ is a constant scalar then the Jacobian is given by

$$J(\boldsymbol{X} \to \boldsymbol{Y}) = c^{\frac{p(p+1)}{2}}.$$

R. 3. *(Muirhead 1982, Theorem 2.1.6, p.58)*
If $\boldsymbol{X}_{p\times p}$ and $\boldsymbol{Y}_{p\times p}$ are symmetric matrices and $\boldsymbol{Y} = \boldsymbol{A}\boldsymbol{X}\boldsymbol{A}'$ where $\boldsymbol{A}_{p\times p}$ is nonsingular then the Jacobian is given by

$$J(\boldsymbol{X} \to \boldsymbol{Y}) = |\boldsymbol{A}|^{-(p+1)}.$$

R. 4. *(Muirhead 1982, Theorem 2.1.8, p.59)*
If $\boldsymbol{X}_{p\times p}$ and $\boldsymbol{Y}_{p\times p}$ are symmetric matrices and $\boldsymbol{Y} = \boldsymbol{X}^{-1}$ then the Jacobian is given by

$$J(\boldsymbol{X} \to \boldsymbol{Y}) = |\boldsymbol{Y}|^{-(p+1)}.$$

R. 5. *(Muirhead 1982, pp.582-583)*
The eigenvalues of $\boldsymbol{X}_{p\times p}$ is defined as $\lambda_i, i = 1,...,p, \lambda_1 \geq ... \geq \lambda_p$ such that

$$|\boldsymbol{X} - \lambda_i \boldsymbol{I}| = 0,$$

so that $\boldsymbol{X} - \lambda_i \boldsymbol{I}$ is singular.

R. 6. *(Gupta and Nagar 2000, Eq.1.3.18, p.14)*
For matrices $\boldsymbol{X}, \boldsymbol{Y}, \boldsymbol{U}, \boldsymbol{V} \in S_p$, *if* $\boldsymbol{U} = \boldsymbol{X} + \boldsymbol{Y}$ *and* $\boldsymbol{V} = \boldsymbol{U}^{-\frac{1}{2}} \boldsymbol{Y} \boldsymbol{U}^{-\frac{1}{2}}$ *where* $\boldsymbol{U} = \boldsymbol{U}^{\frac{1}{2}} \boldsymbol{U}^{\frac{1}{2}}$, *then*

$$J(\boldsymbol{X}, \boldsymbol{Y} \to \boldsymbol{U}, \boldsymbol{V}) = |\boldsymbol{U}|^{-\frac{p+1}{2}}.$$

A.2 Special functions

A.2.1 Expansions

R. 7. *(Gradshteyn and Ryzhik 2007, Eq.0.318*$(1.^{11})$*, p.18)*
The Taylor series polynomial expansion is defined as

$$f(x+b) = \sum_{i=0}^{\infty} \frac{1}{i!} f^{(i)}(b) x^i,$$

where $f^{(i)}(b)$ *is the i-th derivative of* $f(.)$ *at the point b.*

A.2.2 Gamma functions

R. 8. *(Gradshteyn and Ryzhik 2007, Eq.8.310(1), p.892)*
The gamma function $\Gamma(\cdot)$ *is defined as*

$$\Gamma(a) = \int_0^{\infty} u^{a-1} exp(-u) du,$$

for $Re(a) > 0$. *The gamma function possesses the following property (see Gradshteyn and Ryzhik 2007, Eq.8.331(1), p.895):*

$$\Gamma(a) = (a-1)\Gamma(a-1).$$

R. 9. *(Gupta and Nagar 2000, Definition 1.4.2, p.18)*
The multivariate gamma function of dimension p is defined as

$$\Gamma_p(a) = \int_{S_p} |\boldsymbol{U}|^{a-\frac{p+1}{2}} etr(-\boldsymbol{U}) d\boldsymbol{U} = \pi^{\frac{p(p-1)}{4}} \prod_{j=1}^{p} \Gamma\left(a - \frac{j-1}{2}\right),$$

where $\boldsymbol{U} \in S_p$ *and* $Re(a) > \frac{1}{2}(p-1)$ *and* $\Gamma(\cdot)$ *defined in R.8.*

R. 10. *(Gupta and Nagar 2000, Eq.1.5.6, p.30)*
Let $\kappa = (k_1, \dots, k_p), k_1 \geq \dots \geq k_p, k_1 + \dots + k_p = k$. *The generalised hypergeometric coefficient* $(a)_\kappa$, *also known as the generalised Pochammer symbol of weight* κ, *is defined as*

$$(a)_\kappa = \prod_{i=1}^{p} \left(a - \frac{i-1}{2}\right)_{k_i} = \frac{\pi^{\frac{p(p-1)}{4}} \prod_{j=1}^{p} \Gamma(a + k_j - \frac{j-1}{2})}{\Gamma_p(a)},$$

where $Re(a) > \frac{p-1}{2} - k_p$, $\Gamma(.)$ *and* $\Gamma_p(.)$ *as defined in R.8 and R.9, respectively. For* $p = 1$ *the standard Pochammer symbol is*

$$(a)_k = a(a+1)...(a+k-1) = \frac{\Gamma(a+k)}{\Gamma(a)}, (a)_0 = 1, a \neq 0.$$

R. 11. *(Dìaz-Garcìa 2009, Eq.2.4, p.5)*
The generealised gamma function of weight κ *is defined as*

$$\Gamma_p(a,\kappa) = (a)_\kappa \Gamma_p(a),$$

where $(a)_\kappa$ *is the generalised hypergeometric coefficient (see R.10) and* $\Gamma_p(a)$ *is the multivariate gamma function (see R.9).*

A.2.3 Beta function

R. 12. *(Gradshteyn and Ryzhik 2007, Eq.8.380(1), p.908)*
The beta function is defined as

$$B(a;b) = \int_0^1 u^{a-1}(1-u)^{b-1} du,$$

with $a > 0$ *and* $b > 0$.

A.2.4 Dirac delta function

R. 13. *(Dirac 1958, p.58)*
The Dirac delta function is defined as

$$\delta(z-1) = \lim_{d \to 0} \frac{1}{d\sqrt{\pi}} \exp\left(-\frac{(z-1)^2}{d^2}\right),$$

and possesses the following property: $\delta(z-1) = 1$ *if* $z = 1$ *and* $\delta(z-1) = 0$ *if* $z \neq 1$.

A.2.5 Bessel functions

R. 14. *(Erdelyi, Magnus, Oberhettinger, and Tricomi 1953, Eq.2, p.43)*
Let $K_a(\cdot)$ *be the Bessel function of the third kind (Gradshteyn and Ryzhik, 2007) then for* $Re(\alpha) > 0, Re(\beta) > 0$

$$\int_0^\infty u^{a-1} exp\left(-\frac{\alpha}{u} - \beta u\right) du = 2K_a\left(2\sqrt{\alpha\beta}\right)\left(\frac{\alpha}{\beta}\right)^{\frac{a}{2}}.$$

R. 15. *(Herz 1955, p.506)*
Let $B_a(\boldsymbol{D})$ *be the Bessel function of the second kind with matrix argument. Then for* $\boldsymbol{U}, \boldsymbol{A}, \boldsymbol{Z} \in S_p$,

$$\int_{\boldsymbol{U} \in S_p} etr(-\boldsymbol{AU}) etr(-\boldsymbol{ZU}^{-1}) |\boldsymbol{U}|^{-a-\frac{p+1}{2}} d\boldsymbol{U} = B_a(\boldsymbol{AZ}) |\boldsymbol{A}|^a.$$

The integral is absolutely convergent for $Re(\boldsymbol{D}) \in S_p$ *if and only if* $Re(a) > \frac{p-1}{2}$.

A.2.6 Laplace transform

R. 16. *(Erdelyi, Magnus, Oberhettinger, and Tricomi 1953, p.127)*
The Laplace transform, $g(z)$ of an integrable function $f(u), 0 < u < \infty$ is given by

$$g(z) = \mathscr{L}[f(x)] = \int_0^\infty \exp(-zu) f(u) du.$$

A.2.7 Multivariate measures

R. 17. *(Golub and Loan 1996)*
The Frobenius norm of a vector(matrix), $\boldsymbol{X}$, is defined as

$$||\boldsymbol{X}||_F = \sqrt{tr(\boldsymbol{X}'\boldsymbol{X})}.$$

R. 18. *(Mahalanobis 1936)*
The Mahalanobis distance measures the distance between a point and a distribution with mean $\boldsymbol{\mu}_{1\times p}$ and covariance matrix $\boldsymbol{\Sigma}_{p\times p}$. It is defined, for any point $\boldsymbol{x} \in \mathbb{R}^p$, as

$$D_M(\boldsymbol{x}) = \sqrt{(\boldsymbol{x}-\boldsymbol{\mu})\boldsymbol{\Sigma}^{-1}(\boldsymbol{x}-\boldsymbol{\mu})'}.$$

A.2.8 Zonal polynomials

R. 19. *(James 1964, Definition 1.5.1, p.29)*
Let $\boldsymbol{X}_{p\times p}$ be a symmetric matrix and let V_k be the vector space of homogeneous polynomials $\phi(\boldsymbol{X})$ of degree k in the $\frac{p(p+1)}{2}$ distinct elements of $\boldsymbol{X}$. Let $\boldsymbol{L}_{p\times p}$ be a nonsingular matrix. For any congruence transformation $\boldsymbol{X} \to \boldsymbol{LXL}'$, a linear transformation of the space V_k of polynomials $\phi(\boldsymbol{X})$ can be defined, that is

$$\phi \to \boldsymbol{L}\phi : (\boldsymbol{L}\phi)(\boldsymbol{X}) = \phi(\boldsymbol{L}^{-1}\boldsymbol{X}\boldsymbol{L}^{-1}).$$

A subspace $V' \subset V_k$ is called invariant if $\boldsymbol{L}V' \subset V'$ for all $\boldsymbol{L}$ and is called an irreducible invariant subspace if it has no proper invariant subspace. The space V_k decomposes into a direct sum of irreducible invariant subspaces V_κ corresponding to each partition κ of k into not more than p parts

$$V_k = \bigoplus V_\kappa,$$

where $\kappa = (k_1, ..., k_p), k_1 \geq ... \geq k_p \geq 0. k_1 + ... + k_p = k$.
The polynomial $(tr\boldsymbol{X})^k \in V_k$ has a unique decomposition

$$(tr\boldsymbol{X})^k = \sum_\kappa C_\kappa(\boldsymbol{X}),$$

into polynomials, $C_\kappa(\boldsymbol{X}) \in V_\kappa$, belonging to the respective invariant subspaces. The zonal polynomial $C_\kappa(\boldsymbol{S})$ is defined as the component of $(tr\boldsymbol{S})^k$ in the subspace V_κ. It is a symmetric homogeneous polynomial of degree k in the latent roots of $\boldsymbol{X}$ and holds for all p. If the partition κ has more than p parts, the corresponding zonal polynomial will be identically zero.

R. 20. *(Gross and Richards 1987, Eq.6, p.784)*
Let C_κ be a zonal polynomial indexed by the ordered partitions κ of the nonnegative integers k, then

$$etr(\boldsymbol{X}) = \sum_{k=0}^{\infty} \frac{(tr(\boldsymbol{X}))^k}{k!} = \sum_{k=0}^{\infty} \frac{1}{k!} \sum_{\kappa} C_\kappa(\boldsymbol{X}),$$

for a symmetric matrix $\boldsymbol{X}$.

R. 21. *(Muirhead 1982, Theorem 7.2.5, p.243)*
Let $\boldsymbol{H}_{p\times p}$ be an orthogonal matrix, i.e. $\boldsymbol{H} \in O_p$ then

$$\int_{O_p} C_\kappa(\boldsymbol{XHYH}')d\boldsymbol{H} = \frac{C_\kappa(\boldsymbol{X})C_\kappa(\boldsymbol{Y})}{C_\kappa(\boldsymbol{I}_p)},$$

with $\boldsymbol{X} \in S_p$ and $\boldsymbol{Y} \in S_p$, where $d\boldsymbol{H}$ is the normalised Haar measure on O_p (Muirhead 1982, p.72).

R. 22. *(Davis 1979, Eq. 2.7, p.467)*
Let $\boldsymbol{X}_{p\times p}$ be a symmetric matrix and a, a complex constant, then

$$C_\kappa(a\boldsymbol{X}) = a^k C_\kappa(\boldsymbol{X}).$$

R. 23. *(Muirhead 1982, Theorem 7.2.7., p.248)*
Assume $\boldsymbol{Z}_{p\times p}$ is a symmetric matrix and $\boldsymbol{X}_{p\times p}$ is a complex symmetric matrix with $Re(\boldsymbol{X}) \in S_p$. Then

$$\int_{S_p} |\boldsymbol{U}|^{a-\frac{p+1}{2}} etr(-\boldsymbol{UX})C_\kappa(\boldsymbol{UZ})d\boldsymbol{U} = (a)_\kappa \Gamma_p(a)|X|^{-a} C_\kappa(\boldsymbol{ZX}^{-1}),$$

for $Re(a) > \frac{p-1}{2}$, $(a)_\kappa$ as defined in R.10 and $\Gamma_p(a)$ as defined in R.9.

R. 24. *(Teng, Fang, and Deng 1989)*
Assume $\boldsymbol{Z}_{p\times p}$ is a symmetric matrix, $\boldsymbol{X}_{p\times p}$ is a complex symmetric matrix with $Re(\boldsymbol{X}) \in S_p$ and h is a real function over $\mathbb{R}^+$. Then

$$\int_{S_p} |\boldsymbol{U}|^{a-\frac{p+1}{2}} C_\kappa(\boldsymbol{UZ})h(tr\boldsymbol{XU})d\boldsymbol{U} = \frac{(a)_\kappa \Gamma_p(a)\gamma_k(a)}{\Gamma(ap+k)} |\boldsymbol{X}|^{-a} C_\kappa(\boldsymbol{ZX}^{-1}),$$

where $Re(a) > \frac{p-1}{2}$, $(a)_\kappa$ as defined in R.10, $\Gamma_p(a)$ as defined in R.9 and

$$\gamma_k(a) = \int_{\mathbb{R}^+} y^{ap+k-1} h(y)dy.$$

A.2.9 Invariant polynomials

R. 25. *(Davis 1980)*
A class of invariant polynomials $C_\phi^{(\kappa,\tau)}(\boldsymbol{X},\boldsymbol{Y})$ in the elements of the symmetric matrices $\boldsymbol{X}_{p\times p}$ and $\boldsymbol{Y}_{p\times p}$ is defined as having the property of invariance under the simultaneous transformations

$$\boldsymbol{X} \to \boldsymbol{HXH}', \boldsymbol{Y} \to \boldsymbol{HYH}', \boldsymbol{H} \in O_p.$$

These satisfy the relationship

$$\int_{O_p} C_\kappa(\boldsymbol{AH}'\boldsymbol{XH})C_\tau(\boldsymbol{BH}'\boldsymbol{YH})d\boldsymbol{H} = \sum_{\phi\in\kappa,\tau} \frac{C_\phi^{\kappa,\tau}(\boldsymbol{A},\boldsymbol{B})C_\phi^{\kappa,\tau}(\boldsymbol{X},\boldsymbol{Y})}{C_\phi(\boldsymbol{I}_p)},$$

where C_κ, C_τ, C_ϕ *are zonal polynomials indexed by the ordered partitions* κ, τ, ϕ *of the nonnegative integers* $k, t, f = k+t$ *respectively into* $\leq m$ *parts.* $\phi \in \kappa, \tau$ *signifies that the irreducible representation of the group of* $p \times p$ *real nonsingular matrices, indexed by* 2ϕ *occurs in the decomposition of the Kronecker product* $2\kappa \otimes 2\tau$ *of the irreducible representations indexed by* 2κ *and* 2τ *and* $d\boldsymbol{H}$ *is the normalised Haar measure on* O_p *(Muirhead 1982, p.72).*

R. 26. *(Davis 1979, Eq.2.4, p.466)*
Let $\boldsymbol{H}_{p\times p}$ *be an orthogonal matrix, then*

$$\int_{O_p} C_\phi^{\kappa,\tau}(\boldsymbol{XHYH}', \boldsymbol{XHZH}')d\boldsymbol{H} = \frac{C_\phi^{\kappa,\tau}(\boldsymbol{Y},\boldsymbol{Z})C_\phi(\boldsymbol{X})}{C_\phi(\boldsymbol{I}_p)},$$

with $\boldsymbol{X}$, $\boldsymbol{Y}$ *and* $\boldsymbol{Z}$ *symmetric matrices and* $\phi \in (\kappa, \tau)$ *where* $d\boldsymbol{H}$ *is the normalised Haar measure on* O_p *(Muirhead 1982, p.72).*

R. 27. *(Davis 1979, Eq.2.8, p.467)*
Let C_κ, C_τ *and* $C_\phi^{\kappa,\tau}$ *be zonal polynomials indexed by the ordered partitions* κ, τ *and* ϕ *of the nonnegative integers* k, t *and* $f = k+t$ *respectively. Then,*

$$C_\kappa(\boldsymbol{X})C_\tau(\boldsymbol{Y}) = \sum_{\phi\in\,\kappa,\tau} \theta_\phi^{\kappa,\tau} C_\phi^{\kappa,\tau}(\boldsymbol{X},\boldsymbol{Y})$$

with $\theta_\phi^{\kappa,\tau} = \frac{C_\phi^{\kappa,\tau}(\boldsymbol{I}_p,\boldsymbol{I}_p)}{C_\phi(\boldsymbol{I}_p)}$.

R. 28. *(Dìaz-Garcìa 2009, Eq.5.13, p.22)*
For $\boldsymbol{A},\boldsymbol{B}$ *symmetric matrices and* $Re(a) > \frac{p-1}{2} + (k+t)_1$,

$$\int_{S_p} |\boldsymbol{X}|^{a-\frac{p+1}{2}} C_\phi^{\kappa,\tau}(\boldsymbol{AX},\boldsymbol{BX})etr(-\boldsymbol{XZ})d\boldsymbol{X} = \Gamma_p(a,\phi)|\boldsymbol{Z}|^{-a}C_\phi^{\kappa,\tau}(\boldsymbol{AZ}^{-1},\boldsymbol{BZ}^{-1}),$$

where $\Gamma_p(.,.)$ *is the generalised gamma function (see R.11) and* $(k+t)_1$ *s defined in R.11.*

A.2.10 Hypergeometric functions

R. 29. *(Mathai 1993, Definition 2.2, p.96)*
The hypergeometric series is defined as

$${}_pF_q(a_1,...,a_p;b_1,...,b_q;x) = \sum_{k=0}^{\infty} \frac{(a_1)_k...(a_p)_k}{(b_1)_k...(b_q)_k}\frac{x^k}{k!},$$

where $(a)_k$ *is the standard Pochammer symbol given in R.10. The following holds for the hypergeometric series:*

i The series converges for all x *if* $p \leq q$ *and for* $|x| < 1$ *if* $p = q+1$.

ii The series diverges for all $x, x \neq 0$ *for* $p > q+1$.

R. 30. *(Mathai 1993, p.97)*

$$_0F_0(x) = \exp(x).$$

R. 31. *(Mathai 1993, p.97)*

$$_1F_0(a;x) = (1-x)^{-a},$$

for $|x| < 1$.

R. 32. *(Gradshteyn and Ryzhik 2007, Eq.9.210[10], p.1023)*
The confuent hypergeometric function of scalar argument is defined as

$$_1F_1(a;b;x) = \sum_{k=0}^{\infty} \frac{(a)_k}{(b)_k} \frac{x^k}{k!}.$$

R. 33. *(Mathai 1993, p.97)*

$$_2F_1(a_1,a_2;b_1;x) = \frac{\Gamma(b_1)}{\Gamma(a_1)\Gamma(b_1-a_1)} \int_0^1 z^{a_1-1}(1-z)^{b_1-a_1-1}(1-xz)^{-a_2}dz,$$

where $|x| < 1, Re(b_1) > Re(a_1) > 0$. *This is known as the Gauss hypergeometric function.*

R. 34. *(Gross and Richards 1987, Definition 6.1, p.803)*
The hypergeometric function of matrix argument is defined as

$$_sF_t(a_1,...,a_s;b_1,...,b_t;\boldsymbol{X}) = \sum_{k=0}^{\infty} \sum_{\kappa} \frac{(a_1)_\kappa ... (a_s)_\kappa}{k!(b_1)_\kappa ... (b_t)_\kappa} C_\kappa(\boldsymbol{X}),$$

where $a_i, i = 1,...,s; b_j, j = 1,...,t$ *are arbitrary numbers,* $\boldsymbol{X}_{p\times p}$ *is a real symmetric matrix,* $\sum_\kappa$ *denotes the summation over all partitions* κ *of* k *and* $C_\kappa(\boldsymbol{X})$ *is the zonal polynomial of* $\boldsymbol{X}$ *(see R.19). The generalised hypergeometric coefficient* $(a)_\kappa$ *is given in R.10. The following properties hold for the hypergeometric function of matrix argument:*

i If $s < t+1$ *the series converges for all* $\boldsymbol{X}$.

ii If $s = t+1$ *the series converges for* $||\boldsymbol{X}|| < 1$ *(where* $||\boldsymbol{X}||$ *denotes the maximum of the absolute characteristic roots of* $\boldsymbol{X}$*).*

iii If $s > t+1$ *the series diverges for all* $\boldsymbol{X} \neq \boldsymbol{0}$ *unless the series terminates.*

R. 35. *(Gross and Richards 1987, Eq.6.2.1, p.803)*
For a real symmetric matrix $\boldsymbol{X}$,

$$_0F_0(\boldsymbol{X}) = etr(\boldsymbol{X}).$$

R. 36. *(Gross and Richards 1987, Eq.6.2.2, p.804)*
For a real symmetric matrix $\boldsymbol{X}$,

$$_1F_0(a;\boldsymbol{X}) = |\boldsymbol{I}-\boldsymbol{X}|^{-a}$$

R. 37. *(Muirhead 1982, Definition 7.3.2, p.259)*
The hypergeometric function of two matrix arguments is defined for symmetric matrices $\boldsymbol{X}_{p\times p}$ *and* $\boldsymbol{Y}_{p\times p}$*, as follows:*

$$ {}_sF_t(a_1,...,a_s;b_1,...,b_t;\boldsymbol{X},\boldsymbol{Y}) = \sum_{k=0}^{\infty}\sum_{\kappa}\frac{(a_1)_\kappa...(a_s)_\kappa C_\kappa(\boldsymbol{X})C_\kappa(\boldsymbol{Y})}{(b_1)_\kappa...(b_t)_\kappa k! C_\kappa(\boldsymbol{I}_m)}, $$

where $\sum_\kappa$ *denotes the summation over all partitions* κ *of* k *and* $C_\kappa(\boldsymbol{X})$ *is the zonal polynomial of* $\boldsymbol{X}$ *(see R.19). The generalised hypergeometric coefficient* $(a)_\kappa$ *is given in R.10.*

R. 38. *(Muirhead 1982, Theorem 7.3.3, p.260)*
If $\boldsymbol{X} \in S_p$ *and* $\boldsymbol{Y}_{p\times p}$ *a symmetric matrix, then*

$$ \int_{O_p} {}_sF_t(a_1,...,a_s;b_1,...,b_t;\boldsymbol{XHYH}')d\boldsymbol{H} = {}_sF_t(a_1,...,a_s;b_1,...,b_t;\boldsymbol{X};\boldsymbol{Y}), $$

where ${}_sF_t(.)$ *is the hypergeometric function of matrix argument (see R.34) and* ${}_sF_t(.;.)$ *is the hypergeometric argument of two matrices (see R.37).*

A.3 Distributions

A.3.1 Univariate distributions

R. 39. *(DeGroot 1970, Eq.1, p.37)*
A random variable X *has a normal distribution with mean* μ *and variance* $\sigma^2(-\infty < \mu < \infty; \sigma > 0)$ *if* X *has an absolutely continuous distribution with density function specified at any point* $x \in \mathbb{R}$ *by the equation*

$$ f(X) = (2\pi\sigma^2)^{-\frac{1}{2}} exp\left[\frac{-(X-\mu)^2}{2\sigma^2}\right] \equiv f_{N_{\mu,\sigma^2}}(X), $$

denoted as $X \sim N(\mu,\sigma^2)$.

R. 40. *(Raìffa and Schlaifer 1961, Eq.7-54, p.227)*
A random variable X *has an inverse gamma distribution with parameters* α *and* $\beta(\alpha > 0; \beta > 0)$ *if* X *has an absolutely continuous distribution with density function specified at any point* $x(x > 0)$ *by the equation*

$$ f(X) = \frac{\beta^\alpha}{\Gamma(\alpha)}X^{-\alpha-1}exp\left(-\frac{\beta}{X}\right), $$

denoted as $X \sim IG(\alpha,\beta)$*. Note that the expected value of* X *is* $\frac{\beta}{\alpha-1}$.

R. 41. *(DeGroot 1970, Eq.1, p.39)*
A random variable X *has a gamma distribution with parameters* α *and* $\beta(\alpha > 0; \beta > 0)$ *if* X *has an absolutely continuous distribution with density function specified at any point* $x(x > 0)$ *by the equation*

$$ f(X) = \frac{\beta^\alpha}{\Gamma(\alpha)}X^{\alpha-1}\exp(-\beta X), $$

denoted as $X \sim G(\alpha,\beta)$*. Note that the expected value of* X *is* $\frac{\alpha}{\beta}$.

R. 42. *(DeGroot 1970, Eq.1, p.41)*
A random variable X has a Student t-distribution with parameters $\mu, \sigma^2 (\mu \in \mathbb{R}; \sigma > 0)$ *and* $\nu > 0$ *if X has an absolutely continuous distribution with density function specified at any point* $x \in \mathbb{R}$ *by the equation*

$$f(X) = \frac{c\sigma^{-1}}{[\nu + (\frac{X-\mu}{\sigma})^2]^{\frac{n+1}{2}}},$$

denoted as $X \sim T(\mu, \sigma^2, \nu)$*, where* $c = \frac{\nu^{\frac{\nu}{2}}}{B(\frac{1}{2};\frac{n}{2})}$*,* $B(.,.)$ *is the beta function defined in R.12. Note that X has expected value* μ *and variance* $\frac{\nu}{\nu-2}\sigma^2$*.*

R. 43. *(Rao 1990)*
A random variable X has an elliptical distribution with parameters μ *and* $\sigma^2 (\mu \in \mathbb{R}; \sigma > 0)$ *and generator function* $g''(.)$ *if the characteristic function of X is given by*

$$\psi_X(t) = exp(it\mu)\varphi(\sigma^2 t^2),$$

for some function $\varphi(.)$*. If the density function of X exists, it is given by*

$$f(X) = (\sqrt{\sigma^2})^{-1} g''\left(-\frac{1}{2\sigma^2}(X-\mu)^2\right),$$

denoted as $X \sim E(\mu, \sigma^2, g'')$*. X can have a bounded or unbounded support, depending on* $g''(.)$*. Note that the expected value of X is* μ*.*

R. 44. *(Chu 1973, Theorem 1, p.647)*
If X is an elliptical random variable (see R.43) with parameters μ *and* σ^2*, with a density function* $f(X)$*, then there is a scalar function* $w(z)$ *defined on* $0 < z < \infty$ *such that*

$$f(X) = \int_0^\infty w(z) f_{N_{\mu, z^{-1}\sigma^2}}(X) dz, \tag{A.1}$$

where $f_{N_{\mu,z^{-1}\sigma^2}}(X)$ *is the normal density function (see R.39) with expected value* μ *and variance* $z^{-1}\sigma^2$*.*

A.3.2 Multivariate distributions

R. 45. *(Muirhead 1982, Theorem 1.2.9, p.9)*
The random vector $\boldsymbol{X}_{m\times 1} \in \mathbb{R}^m$ *is said to follow a multivariate normal distribution with mean vector* $\boldsymbol{\mu}_{m\times 1} \in \mathbb{R}^m$ *and covariance matrix* $\boldsymbol{\Sigma}_{m\times m} \in S_m$ *if its density function is given by*

$$\begin{aligned} f(\boldsymbol{X}) &= (2\pi)^{-\frac{m}{2}} |\boldsymbol{\Sigma}|^{-\frac{1}{2}} exp\left[-\frac{1}{2}(\boldsymbol{X}-\boldsymbol{\mu})'\boldsymbol{\Sigma}^{-1}(\boldsymbol{X}-\boldsymbol{\mu})\right] \equiv f_{N_{\boldsymbol{\mu},\boldsymbol{\Sigma}}}(\boldsymbol{x}) \\ &= (2\pi)^{-\frac{m}{2}} |\boldsymbol{\Sigma}|^{-\frac{1}{2}} etr\left[-\frac{1}{2}\boldsymbol{\Sigma}^{-1}(\boldsymbol{X}-\boldsymbol{\mu})(\boldsymbol{X}-\boldsymbol{\mu})'\right], \end{aligned}$$

denoted as $\boldsymbol{X} \sim N_m(\boldsymbol{\mu}, \boldsymbol{\Sigma})$*.*

R. 46. *(Press 1982, Eq. 6.2.3, p.136)*
The random vector $\boldsymbol{X}_{m\times 1} \in \mathbb{R}^m$ *is said to follow a multivariate t distribution with parameters*

$\boldsymbol{\mu}_{m\times 1} \in \mathbb{R}^m, \boldsymbol{\Sigma}_{m\times m} \in S_m$ *and* $\nu > 0$ *if its density function is given by*

$$f(\boldsymbol{X}) = \frac{c|\boldsymbol{\Sigma}|^{-\frac{1}{2}}}{[\nu + (\boldsymbol{X}-\boldsymbol{\mu})'\boldsymbol{\Sigma}^{-1}(\boldsymbol{X}-\boldsymbol{\mu})]^{\frac{\nu+m}{2}}},$$

where $c = \frac{\nu^{\frac{\nu}{2}}\Gamma(\frac{\nu+m}{2})}{\pi^{\frac{m}{2}}\Gamma(\frac{\nu}{2})}$ *and it is denoted as* $\boldsymbol{X} \sim T_m(\boldsymbol{\mu}, \boldsymbol{\Sigma}, \nu)$, *with* $\Gamma(\cdot)$ *defined in R.8. The expected value of* $\boldsymbol{X}$ *is* $\boldsymbol{\mu}$.

R. 47. *(Gupta, Varga, and Bodnar 2013, Remark 2.1, p.15)*
The random vector $\boldsymbol{X}_{m\times 1} \in \mathbb{R}^m$ *is said to follow a multivariate elliptical distribution with parameters* $\boldsymbol{\mu}_{m\times 1} \in \mathbb{R}^m, \boldsymbol{\Sigma}_{m\times m} \in S_m$ *and generator function* $g''(.)$ *if its characteristic function is given by*

$$\psi_{\boldsymbol{X}}(\boldsymbol{T}) = exp(i\boldsymbol{T}'\boldsymbol{\mu})\varphi(\boldsymbol{T}'\boldsymbol{\Sigma}\boldsymbol{T}),$$

for some function $\varphi(.)$. *If its density function exists, it is given by*

$$f(\boldsymbol{X}) = |\boldsymbol{\Sigma}|^{-\frac{1}{2}} g''\left[-\frac{1}{2}(\boldsymbol{X}-\boldsymbol{\mu})'\boldsymbol{\Sigma}^{-1}(\boldsymbol{X}-\boldsymbol{\mu})\right],$$

denoted as $\boldsymbol{X} \sim E_m(\boldsymbol{\mu}, \boldsymbol{\Sigma}, g'')$. *The expected value of* $\boldsymbol{X}$ *is* $\boldsymbol{\mu}$.

R. 48. *(Chu 1973, Theorem 1, p.647)*
If $\boldsymbol{X}_{m\times 1} \in \mathbb{R}^m$ *is an ellipticaly distributed random variable (see R.47) with parameters* $\boldsymbol{\mu}$ *and* $\boldsymbol{\Sigma}$, *with* $f(\boldsymbol{X})$ *as its density function, then there is a scalar function* $w(z)$ *defined in* $0 < z < \infty$ *such that*

$$f(\boldsymbol{X}) = \int_0^\infty w(z) f_{N_{\boldsymbol{\mu}, z^{-1}\boldsymbol{\Sigma}}}(\boldsymbol{X}) dz, \tag{A.2}$$

where $f_{N_{\boldsymbol{\mu}, z^{-1}\boldsymbol{\Sigma}}}(\boldsymbol{X})$ *is the multivariate normal density function (see R.49) with expected value* $\boldsymbol{\mu}$ *and covariance matrix* $z^{-1}\boldsymbol{\Sigma}$.

Proof. Assume without loss of generality that $\boldsymbol{\mu} = \boldsymbol{0}$. Let

$$f(\boldsymbol{X}) = h\left(\frac{1}{2}\boldsymbol{X}'\boldsymbol{\Sigma}^{-1}\boldsymbol{X}\right) = h(s), s = \frac{1}{2}\boldsymbol{X}'\boldsymbol{\Sigma}^{-1}\boldsymbol{X} \geq 0.$$

Define the scalar function

$$w(z) = (2\pi)^{\frac{m}{2}}|\boldsymbol{\Sigma}|^{\frac{1}{2}} z^{-\frac{m}{2}} \mathscr{L}^{-1}[h(s)], 0 < z < \infty$$

where $\mathscr{L}$ is the Laplace transform operator (See R.16). Therefore

$$\mathscr{L}^{-1}[h(s)] = w(z)(2\pi)^{-\frac{m}{2}}|\boldsymbol{\Sigma}|^{-\frac{1}{2}} z^{\frac{m}{2}}. \tag{A.3}$$

From (A.3) it follows that

$$\begin{aligned} f(\boldsymbol{X}) &= h(s) = \mathscr{L}(\mathscr{L}^{-1}[h(s)]) \\ &= \mathscr{L}[w(z)(2\pi)^{-\frac{m}{2}}|\boldsymbol{\Sigma}|^{-\frac{1}{2}}z^{\frac{m}{2}}] \\ &= \int_0^\infty w(z)(2\pi)^{-\frac{m}{2}}|z^{-1}\boldsymbol{\Sigma}|^{-\frac{1}{2}}\exp(-zs)dz \\ &= \int_0^\infty w(z)f_{N_{\boldsymbol{\mu},z^{-1}\boldsymbol{\Sigma}}}(\boldsymbol{X})dz. \end{aligned}$$

□

A.3.3 Matrix variate distributions

R. 49. *(Gupta and Nagar 2000, Theorem 2.2.1, p.55)*
The random matrix $\boldsymbol{X}_{m\times p} \in \mathbb{R}^{m\times p}$ is said to follow a matrix variate normal distribution with mean matrix $\boldsymbol{\mu} \in \mathbb{R}^{m\times p}$ and covariance matrix $\boldsymbol{\Sigma}_{m\times m} \otimes \boldsymbol{\Omega}_{p\times p}, \boldsymbol{\Sigma} \in S_m, \boldsymbol{\Omega} \in S_p$ if its density function is given by

$$f(\boldsymbol{X}) = (2\pi)^{-\frac{mp}{2}}|\boldsymbol{\Sigma}|^{-\frac{p}{2}}|\boldsymbol{\Omega}|^{-\frac{m}{2}} etr\left[-\frac{1}{2}\boldsymbol{\Sigma}^{-1}(\boldsymbol{X}-\boldsymbol{\mu})\boldsymbol{\Omega}^{-1}(\boldsymbol{X}-\boldsymbol{\mu})'\right] \equiv f_{N_{\boldsymbol{\mu},\boldsymbol{\Sigma}\otimes\boldsymbol{\Omega}}}(\boldsymbol{X}),$$

denoted as $\boldsymbol{X} \sim N_{m,p}(\boldsymbol{\mu}, \boldsymbol{\Sigma}\otimes\boldsymbol{\Omega})$. The expected value of $\boldsymbol{X}$ is $\boldsymbol{\mu}$.

R. 50. *(Gupta and Nagar 2000, Definition 4.2.1, p.134)*
The random matrix $\boldsymbol{X}_{m\times p} \in \mathbb{R}^{m\times p}$ is said to follow a matrix variate t distribution with parameters $\boldsymbol{\mu}_{m\times p} \in \mathbb{R}^{m\times p}, \boldsymbol{\Sigma}_{m\times m} \in S_m, \boldsymbol{\Omega}_{p\times p} \in S_p$ and $\nu > 0$ if its density function is given by

$$f(\boldsymbol{X}) = \frac{\Gamma_m\left[\frac{\nu+m+p-1}{2}\right]}{\pi^{\frac{mp}{2}}\Gamma_m\left[\frac{\nu+m-1}{2}\right]}|\boldsymbol{\Sigma}|^{-\frac{p}{2}}|\boldsymbol{\Omega}|^{-\frac{m}{2}}|\boldsymbol{I}_m + \boldsymbol{\Sigma}^{-1}(\boldsymbol{X}-\boldsymbol{\mu})\boldsymbol{\Omega}^{-1}(\boldsymbol{X}-\boldsymbol{\mu})'|^{-\frac{\nu+m+p-1}{2}},$$

denoted as $\boldsymbol{X} \sim T_{m,p}(\boldsymbol{\mu}, \boldsymbol{\Sigma}, \boldsymbol{\Omega}, \nu)$ with $\Gamma_m(.)$ defined in R.9. The expected value of $\boldsymbol{X}$ is $\boldsymbol{\mu}$.

R. 51. *(Gupta, Varga, and Bodnar 2013, Definition 2.1, p.15 and Theorem 2.6, p.20)*
The random matrix $\boldsymbol{X}_{m\times p}$ is said to follow a matrix variate elliptical distribution with parameters $\boldsymbol{\mu}_{m\times p} \in \mathbb{R}^{m\times p}, \boldsymbol{\Sigma}_{m\times m} \in S_m$ and $\boldsymbol{\Omega}_{p\times p} \in S_p$ and generator $g''(.)$, if its characteristic function is given by

$$\psi_{\boldsymbol{X}}(\boldsymbol{T}) = etr(i\boldsymbol{T}'\boldsymbol{X})\varphi(tr(\boldsymbol{T}'\boldsymbol{\Sigma}\boldsymbol{T}\boldsymbol{\Omega})),$$

for some function $\varphi(.)$, denoted as $\boldsymbol{X} \sim E_{m,p}(\boldsymbol{\mu}, \boldsymbol{\Sigma}, \boldsymbol{\Omega}, g'')$. The expected value of $\boldsymbol{X}$ is $\boldsymbol{\mu}$. If its density function exists, it is given by

$$f(\boldsymbol{X}) = |\boldsymbol{\Sigma}|^{-\frac{p}{2}}|\boldsymbol{\Omega}|^{-\frac{m}{2}} g''\left[tr\left(\boldsymbol{\Sigma}^{-1}(\boldsymbol{X}-\boldsymbol{\mu})\boldsymbol{\Omega}^{-1}(\boldsymbol{X}-\boldsymbol{\mu})'\right)\right].$$

Similarly as in R.48, the density function of $\boldsymbol{X}$ can also be written as

$$f(\boldsymbol{X}) = \int_0^\infty w(z)f_{N_{\boldsymbol{\mu},z^{-1}\boldsymbol{\Sigma}\otimes\boldsymbol{\Omega}}}(\boldsymbol{X})dz \tag{A.4}$$

for a scalar function $-1 \le w(z) \le 1, z > 0$ such that $\int_0^\infty w(z)dz = 1$ where $f_{N_{\boldsymbol{\mu},z^{-1}\boldsymbol{\Sigma}\otimes\boldsymbol{\Omega}}}(\cdot)$ is the matrix variate normal density function (R.49) with expected value $\boldsymbol{\mu}$ and covariance matrix

$z^{-1}\boldsymbol{\Sigma}\otimes\boldsymbol{\Omega}$. *See R.52, Arashi, Saleh, and Tabatabaey (2013) and Arashi, Iranmanesh, and Salarzadeh Jenatabadi (2013) for more details.*

R. 52. *(Provost and Cheong 2002)*
It is essential to note that (A.1), (A.2) and (A.4) are different from the class of scale mixtures of the normal distribution. If a scale mixture of the normal distribution is considered then the scalar function, which is referred to as the weighting function, is actually a density function where $0 \leq w(z) \leq 1$ *for all values of* z*. However, in the case of (A.1), (A.2) and (A.4), the weighting function does not have the same restriction but* $-1 \leq w(z) \leq 1$ *for all values of* z*. The restriction* $\int_0^\infty w(z)dz = 1$ *holds for all cases.*

R. 53. *(Muirhead 1982, Theorem 3.2.17, p.104)*
If $\boldsymbol{X}_{p\times p} \in S_p$ *is a random matrix with density function* $f(\boldsymbol{X})$ *then the joint density function of the diagonal matrix of the eigenvalues (see R.5) of* $\boldsymbol{X}$*, that is,* $\boldsymbol{\Lambda} = diag(\lambda_1, ..., \lambda_p)$*,* $\lambda_1 \geq ... \geq \lambda_p > 0$ *is given by*

$$g'(\boldsymbol{\Lambda}) = \frac{\pi^{\frac{p^2}{2}}}{\Gamma_p\left(\frac{p}{2}\right)} \prod_{i<j}^{p} (\lambda_i - \lambda_j) \int_{O_p} f(\boldsymbol{H\Lambda H}')d\boldsymbol{H}$$

for any $\boldsymbol{H} \in O_p$*, where* $d\boldsymbol{H}$ *is the normalised Haar measure on* O_p *(Muirhead 1982, p.72).*

A.4 Bayesian concepts

R. 54. *(Berger 2013, Definition 11, p.27)*
The likelihood function is defined as a function that gives the likelihood of a specific sample based on certain values of the population parameters. If the value of the function is large, the values of the population parameters are more likely to be true.

R. 55. *(Berger 2013, p.82)*
An objective prior is used when no or minimal prior information exists. This prior contains no information about the parameter of interest and does not favour one possible value of the parameter above another. An objective prior function may be improper (the area under the function is not equal to one).

R. 56. *(Berger 2013, p.74)*
A subjective prior quantifies a prior belief about the parameter of interest. It is usually expressed as a density function, assigning different probabilities to different possible values of the parameter of interest.

R. 57. *(Berger 2013, Definition 1, p.130)*
A conjugate prior distribution of a parameter, is a prior distribution that results in a posterior distribution with a density function of the same functional form as the prior density function.

R. 58. *(Berger 2013, p.126)*
The posterior distribution is the conditional distribution of the parameter of interest, based on and subsequent to obtaining a sample. The posterior density function is proportional to the prior density function multiplied by the likelihood function as folllows,

$$posterior \propto prior \times likelihood.$$

R. 59. *(Berger 2013, p.161)*
The Bayes estimator of a parameter is the function of prior and sample information, for which the Bayes risk is a minimum.

R. 60. *(Berger 2013, Result 3, p.161)*
The squared error loss (SEL) function is defined for an estimator $\widehat{\omega}$ of ω as

$$\pounds(\omega;\widehat{\omega}) = [\widehat{\omega} - \omega]^2$$

for the univariate case and for the estimator $\widehat{\boldsymbol{\omega}}$ of $\boldsymbol{\omega}$ as

$$\pounds(\boldsymbol{\omega};\widehat{\boldsymbol{\omega}}) = (\widehat{\boldsymbol{\omega}} - \boldsymbol{\omega})^{'}(\widehat{\boldsymbol{\omega}} - \boldsymbol{\omega})$$

for the multivariate case.
The Bayes estimator of $\boldsymbol{\omega}$ under the squared error loss function is the mean of the posterior distribution (PM estimator).

R. 61. *(Das and Dey 2010, Theorem 1)*
Under the loss function $L(\boldsymbol{\omega},\widehat{\boldsymbol{\omega}}) = \log\left[\frac{q(\boldsymbol{\omega}|\boldsymbol{X})}{q(\widehat{\boldsymbol{\omega}}|\boldsymbol{X})}\right]$ where $q(.;\boldsymbol{X})$ is the posterior density function (see R.58), the Bayes estimator of $\boldsymbol{\omega}$ under this loss function, is the mode of the posterior distribution (MAP estimator).

A.5 Computational methods

R. 62. *(Casella and George 1992)*
Gibbs sampling is a sampling method based on joint and conditional density functions. Consider the random variables X and Y with joint density function $f(X,Y)$. To obtain a sample, $X_1,\ldots,X_n$, from the marginal distribution of X, the marginal density function, $f(X)$, is needed. Note that

$$f(X) = \int_{\Omega_Y} f(X,Y)dY$$

In some cases, the integration leads to intractable expressions or cannot be solved. Hence, the form of $f(X)$ is unknown. To circumvent this problem, Gibbs sampling can be used if the conditional density functions, $f(X|Y)$ and $f(Y|X)$ are known.

The process is as follows:

1. *Set an intial value x_0.*
2. *Repeat the following k times:*
   (a) *Simulate $y_i \sim f(Y|x_{i-1})$*
   (b) *Simulate $x_i \sim f(X|y_i)$*

Extract the sample of size n, from $x_1,\ldots,x_k$. There are various methods of extracting the sample from the simulated sequence, most commonly select x_i if $i > \ell$ for a certain value $\ell \leq k$. The value ℓ is chosen based on the convergence of the sequence . The sample $x_1,\ldots,x_n$ can only be extracted from the sequence once convergence is reached and the effect of the initial values has

diminished.
If ℓ is large enough, the density function of the sample, $x_1,...,x_n$, converges to $f(X)$.

R. 63. *(Metropolis, Rosenbluth, Rosenbluth, Teller, and Teller 1953)*
The Metropolis-Hastings algorithm is an accept-reject type algorithm often used with Gibbs sampling (see R.62). Suppose that in R.62, $f(X|Y)$ is an unknown density but there exists a c such that $f(X|Y) \leq ch(X|Y)$ for all x, with $h(X|Y)$ a known density function.

The sequence is simulated as follows:

1. *Generate a random variate x^* from $h(X|y)$.*
2. *Calculate $\alpha = \frac{f(x^*|y)}{f(x|y)}$*
3. *If $min(\alpha, 1) \geq u, u \sim uniform(0,1)$ then $x_{i+1} = x^*$ else $x_{i+1} = x_i$.*

The sample $x_1,...,x_n$ can only be extracted from the sequence once convergence is reached and the effect of the initial values can be disgarded.

Appendix B

Code for Chapter 2

B.1 Univariate simulation study - Section 2.5.1

OpenBUGS code for normal-inverse gamma prior: Normal population

```
list(h = c( −0.95, 1.439, −0.247, −0.523, 0.839, 1.248, 0.4286,
 1.1729, −0.6748, 0.00), N = 10, mu0=0.5, n0=0.5, alpha0=4,
  beta0=3)
list(mu=0,vari=1)

model
{
        for (i in 1:N) {
                h[i] ~ dnorm(mu,vari)
        }
        mu ~ dnorm(mu0,v)
        v1 ~ dgamma(alpha0,beta0)
        vari<−1/v1
}
```

Student t population

```
list(h = c(8.93, 10.32, 12.72, 11.68, 11.75, 15.15, 15.61, 6.61,
 7.13, 7.95, 6.05, 9.37, 9.34, 10.59, 11.38, 8.02, 8.15, 11.88,
  9.32, 7.73), N = 20, mu0=12, n0=2, alpha0=4, beta0=3)

list(mu=0,vari=1)

model
{
        for (i in 1:N) {
                h[i] ~ dt(mu,vari,nu)
        }
        mu ~ dnorm(mu0,v)
        vari ~ dgamma(alpha0,beta0)
        nu~dunif(1,10)
        v<−vari*n0
        sigma2<−nu/(nu−2)*1/vari
}
```

OpenBUGS code for normal-gamma prior:

Normal population

```
list(h = c( −0.95, 1.439, −0.247, −0.523, 0.839, 1.248, 0.4286,
 1.1729, −0.6748, 0.00), N = 10, mu0=0.5, n0=0.5, alpha0=4,
  beta0=0.3333)
```

```
list(mu=0,vari=1)

model
{
for (i in 1:N) {
h[i] ~ dnorm(mu,vari)
}
mu ~ dnorm(mu0,v)
vari ~ dgamma(alpha0,beta0)
}
```

Student t population

```
list(h = c(8.93, 10.32, 12.72, 11.68, 11.75, 15.15, 15.61, 6.61,
 7.13, 7.95, 6.05, 9.37, 9.34, 10.59, 11.38, 8.02, 8.15, 11.88,
  9.32, 7.73), N = 20, mu0=12, n0=2, alpha0=4, beta0=0.3333)

list(mu=0,vari=1)

model
{
for (i in 1:N) {
h[i] ~ dt(mu,vari,nu)
}
mu ~ dnorm(mu0,v)
vari ~ dgamma(alpha0,beta0)
nu~dunif(1,10)
v<-vari*n0
sigma2<-nu/(nu-2)*vari
}
```

B.2 Kanamycin dataset - Section 2.5.1

OpenBUGS code for normal-inverse gamma prior:

```
list(h = c( 9.57, 7.59, 10.57, 13.61, 16.94, 13.13, 11.67,
 11.84,  12.31, 8.96, 9.47, 11.85, 11.38, 10.66, 7.3,
  14.71, 8.7, 15.19,   10.08, 7.36), ca=c(25.2, 26, 16.3,
   27.2, 23.2, 18.1, 22.2, 17.2, 18.8, 16.4, 24.8, 26.8,
    25.4, 14.9, 18.1, 16.3, 31.3, 31.2, 18,
    15.6), N = 20, n0=2, alpha0=4,beta=3)

model
{
        for (i in 1:N) {
                ca[i]~dt(mu[i],m,nu)
                mu[i]<-beta0+beta1*h[i]
        }
        beta0~dnorm(0,1)
        beta1~dnorm(0,1)
        vari~dgamma(alpha0,beta)
        nu<-7
        m<-1/vari
}
```

OpenBUGS code for normal-gamma prior:

```
list(h = c( 9.57, 7.59, 10.57, 13.61, 16.94, 13.13, 11.67,
```

```
11.84,  12.31, 8.96, 9.47, 11.85, 11.38, 10.66, 7.3,
14.71, 8.7, 15.19,   10.08, 7.36), ca=c(25.2, 26, 16.3,
27.2, 23.2, 18.1, 22.2, 17.2, 18.8, 16.4, 24.8, 26.8,
25.4, 14.9, 18.1, 16.3, 31.3, 31.2, 18,
     15.6), N = 20, n0=2, alpha0=4,beta=0.3333)

model
{
for (i in 1:N) {
ca[i]~dt(mu[i],vari,nu)
mu[i]<-beta0+beta1*h[i]
}
beta0~dnorm(0,1)
beta1~dnorm(0,1)
vari~dgamma(alpha0,beta)
nu<-7
}
```

B.3 Multivariate simulation study - Section 2.5.2

Matlab code:

Multivariate normal population

```
p=3;
n=5;
n0=1;
s=1;
theta0=0.5;
sigma=eye(p);
mu=zeros(1,p);

c=10000;
X=mvnrnd(mu,sigma,n);
Xbar=ones(1,n)*X/n;
V=(X-ones(n,1)*Xbar)'*(X-ones(n,1)*Xbar);
Phi=4*eye(p);
omega=eye(s);
theta=ones(1,p)*theta0;

mW=p;
mIW=2*p+1;

SigmaW=1.1*eye(p);
SigmaIW=1.1*eye(p);

b=1/(n+n0)*(n*Xbar+n0*theta);
Sx=V/(n-1);
Y=n*n0/(n+n0)*(Xbar-theta)*(Xbar-theta)'+V;

meanmuIW=zeros(1,p);
meanmuW=zeros(1,p);
meanSIW=zeros(p,p);
meanSW=zeros(p,p);

mapmuIW=meanmuIW;
```

```
mapmuW=meanmuIW;
mapSIW=meanSIW;
mapSW=meanSW;

maxmIW=-10000000;
maxmW=-100000000;
maxSIW=-100000000;
maxSW=-100000000;

for i=1:c;

%Inverse-Wishart;

muIW=mvnrnd(b,SigmaIW,1);

B=n*(muIW-theta)'*(muIW-theta)+Phi;
A=V+n*(Xbar-muIW)'*(Xbar-muIW);
SigmaIW=iwishrnd(B+A,mIW+s*n+s);

meanmuIW=meanmuIW+muIW*(1/c);
meanSIW=meanSIW+SigmaIW*(1/c);

if mvnpdf(muIW,b,(n+n0)*SigmaIW)>maxmIW mapmuIW=muIW;
 maxmIW=mvnpdf(muIW,b,(n+n0)*SigmaIW); end;
f=det(SigmaIW)^(-0.5*(mIW+s+n*s))*exp(-0.5*trace(
inv(SigmaIW)*(B+A));
if f>maxSIW mapSIW=SigmaIW ;maxSIW=f; end;

%Wishart;
muW=mvnrnd(b,SigmaW,1);
D=n0*(muW-theta)*(muW-theta)';

%Metropolis-Hastings;
w=0.5;
Sstar=w*iwishrnd(Phi,s+n*s)+(1-w)*wishrnd(Phi,mW);
fSigmaW=det(SigmaW)^(0.5*(mW-p-1-s-n*s))*exp(trace(
-0.5*SigmaW*inv(Phi)-0.5*inv(SigmaW)*(D+(n+n0)
/(n*n0)*Y)));
fSstar=det(Sstar)^(0.5*(mW-p-1-s-n*s))*exp(trace(
-0.5*Sstar*inv(Phi)-0.5*inv(Sstar)*(D+(n+n0)
/(n*n0)*Y)));
if min(fSstar/fSigmaW,1)>unifrnd(0,1), SigmaW=Sstar,
 end;

if i>0.1*c meanmuW=meanmuW+muW/c;
if i>0.1*c meanSW=meanSW+SigmaW/c;

if mvnpdf(muW,b,(n+n0)*SigmaW)>maxmW mapmuW=muW;
 maxmW=mvnpdf(muW,b,(n+n0)*SigmaW); end;
f=det(SigmaW)^(-0.5*(mW+s+n*s))*exp(-0.5*(trace
(SigmaW*inv(Phi))
+trace(SigmaW*(D+(n+n0)/(n*n0)*Y))));
if f>maxSW mapSW=SigmaW; maxSW=f; end;

end;

r=[1:10];
```

```
Norms=zeros(length(r),10);
for j=1:length(r);
Norms(j,1)=norm((Xbar-mu).^r(j))^(1/r(j));
Norms(j,2)=norm((meanmuIW-mu).^r(j))^(1/r(j));
Norms(j,3)=norm((meanmuW-mu).^r(j))^(1/r(j));
Norms(j,4)=norm((mapmuIW-mu).^r(j))^(1/r(j));
Norms(j,5)=norm((mapmuW-mu).^r(j))^(1/r(j));
Norms(j,6)=norm((Sx-sigma).^r(j))^(1/r(j));
Norms(j,7)=norm((meanSIW-sigma).^r(j))^(1/r(j));
Norms(j,8)=norm((meanSW-sigma).^r(j))^(1/r(j));
Norms(j,9)=norm((mapSIW-sigma).^r(j))^(1/r(j));
Norms(j,10)=norm((mapSW-sigma).^r(j))^(1/r(j));
end;

figure1=figure('Color',[1 1 1]);
plot(r,Norms(1:10,1),'-o',r,Norms(1:10,2),'-.*',
r,Norms(1:10,3),':');
legend('Sample mean','Inverse-Wishart prior'
,'Wishart prior')
title('Expected value - SEL')
xlabel('r')
ylabel('Frobenius norm');

figure2=figure('Color',[1 1 1]);
plot(r,Norms(1:10,1),'-o',r,Norms(1:10,4),'-.*',
r,Norms(1:10,5),':')
legend('Sample mean','Inverse-Wishart prior'
,'Wishart prior')
xlabel('r')
ylabel('Frobenius norm')
title('Expected value - MAP');

figure3=figure('Color',[1 1 1]);
plot(r,Norms(1:10,6),'-o',r,Norms(1:10,7),'-.*',
r,Norms(1:10,8),':')
legend('Sample mean','Inverse-Wishart prior'
,'Wishart prior')
xlabel('r')
ylabel('Frobenius norm')
title('Covariance matrix - SEL');

figure4=figure('Color',[1 1 1]);
plot(r,Norms(1:10,6),'-o',r,Norms(1:10,9),'-.*',
r,Norms(1:10,10),':')
legend('Sample mean','Inverse-Wishart prior'
,'Wishart prior');
xlabel('r')
ylabel('Frobenius norm')
title('Covariance matrix - MAP');
```

Multivariate t population

```
p=3;
n0=5;
s=1;
n=30;
theta0=0.5;
sigma=eye(p)%
```

```
mu=zeros(1,p);
c=10000;
nu0=5;
z=1;
TRB=zeros(z,10);
DF=zeros(c,11);


for j=1:z;

X=mvtrnd(sigma,nu0,n)+ones(n,1)*mu;
Xbar=ones(1,n)*X/n;

V=(X-ones(n,1)*Xbar)'*(X-ones(n,1)*Xbar);
Phi=4*eye(p);
omega=eye(s);
theta=ones(1,p)*theta0;
mW=p;
mIW=9.5;

SigmaW=1.1*eye(p);
SigmaIW=1.1*eye(p);
muIW=0.5*ones(1,p);
muW=0.5*ones(1,p);

b=1/(n+n0)*(n*Xbar+n0*theta);
Sx=V/(n-1);

meanmuIW=zeros(1,p);
meanmuW=zeros(1,p);
meanSIW=zeros(p,p);
meanSW=zeros(p,p);

mapmuIW=meanmuIW;
mapmuW=meanmuIW;
mapSIW=meanSIW;
mapSW=meanSW;

maxmIW=-10000000000000;
maxmW=-100000000000000;
maxSIW=-10000000000000;
maxSW=-100000000000000;


for i=1:c;
DF(i,2)=mean(muIW);
DF(i,3)=median(muIW);
DF(i,1)=i;
DF(i,4)=det(SigmaIW);
DF(i,5)=trace(SigmaIW);
detsen=eigs(SigmaIW);
DF(i,6)=detsen(p);
DF(i,7)=mean(muW);
DF(i,8)=median(muW);
DF(i,9)=det(SigmaW);
DF(i,10)=trace(SigmaW);
detsen=eigs(SigmaW);
```

```
DF(i,11)=detsen(p);

%Inverse-Wishart;
 mustar1=mvnrnd(b,1/(n+n0)*SigmaIW,1);

D1=n*(muIW-theta)'*(muIW-theta);
A1=V+n*(Xbar-muIW)'*(Xbar-muIW);
D2=n*(mustar1-theta)'*(mustar1-theta);
A2=V+n*(Xbar-mustar1)'*(Xbar-mustar1);

fmuIW=(0.5*nu0+0.5*trace(inv(SigmaIW)*(A1+D1)))^(
-0.5*p*(p-mIW-n)-0.5*nu0);
fmustar=(0.5*nu0+0.5*trace(inv(SigmaIW)*(A2+D2)))
^(-0.5*p*(p-mIW-n)-0.5*nu0);

 if  min(fmustar/fmuIW,1)>unifrnd(0,1)  muIW=mustar1;
  end;

B=n*(muIW-theta)'*(muIW-theta)+Phi;
A=V+n*(Xbar-muIW)'*(Xbar-muIW);

w=0.5;

 Sigmastar=(1-w)*iwishrnd(Phi,mIW)+w*wishrnd(Phi,mIW);

fSigmaIW=det(SigmaIW)^(0.5*(mIW+n+1))*(0.5*nu0+
0.5*trace(inv(SigmaIW)*(A+B)))^(-0.5*p*(p-mIW-n)
-0.5*nu0);
fSigmastar=det(Sigmastar)^(0.5*(mIW+n+1))*(0.5*nu0
+0.5*trace(inv(Sigmastar)*(A+B)))^(-0.5*p*(p-mIW-n)
-0.5*nu0);
 if  min(fSigmastar/fSigmaIW,1)>unifrnd(0,1)
SigmaIW=Sigmastar;
end;


meanmuIW=meanmuIW+muIW*(1/c);
meanSIW=meanSIW+SigmaIW*(1/c);

 if  (0.5*nu0+0.5*trace(inv(SigmaIW)*(A+B-Phi)))
^(-0.5*p*(p-mIW-n)
-0.5*nu0)>maxmIW  mapmuIW=muIW;
  maxmIW=(0.5*nu0+0.5*trace(inv(SigmaIW)
  *(A+B-Phi)))^(-0.5*p*(p-mIW-n)-0.5*nu0);
    end;
 f=det(SigmaIW)^(0.5*(mIW+n+1))*(0.5*nu0+
 0.5*trace(inv(SigmaIW)*(A+B)))^(-0.5*p
 *(p-mIW-n)-0.5*nu0);
 if  f>maxSIW  mapSIW=SigmaIW  ;maxSIW=f;  end;

%Wishart;
 mustar=mvnrnd(b,1/(n+n0)*SigmaW,1);

D1=n*(muW-theta)'*(muW-theta);
A1=V+n*(Xbar-muW)'*(Xbar-muW);
D2=n*(mustar-theta)'*(mustar-theta);
A2=V+n*(Xbar-mustar)'*(Xbar-mustar);
```

```
fmuW=(0.5*nu0+0.5*trace(inv(SigmaW)*(A1+D1)))
^(-0.5*p*(p-mW-n)-0.5*nu0);
fmustar=(0.5*nu0+0.5*trace(inv(SigmaW)*(A2+D2)))
^(-0.5*p*(p-mW-n)-0.5*nu0);
if min(fmustar/fmuW,1)>unifrnd(0,1) muW=mustar; end;


B=n*(muW-theta)'*(muW-theta)+Phi;
A=V+n*(Xbar-muW)'*(Xbar-muW);

w=0.5;
Sigmastar=w*iwishrnd(Phi,mW*2+1)
+(1-w)*wishrnd(Phi,mW);

fSigmaW=det(SigmaW)^(0.5*(mW+n+1))*(0.5*nu0
+0.5*trace(inv(SigmaW)*(A+B))+0.5*trace(SigmaW*Phi))
^(-0.5*p*(p-mW-n)-0.5*nu0);
fSigmastar=det(Sigmastar)^(0.5*(mW+n+1))*(0.5*nu0
+0.5*trace(inv(Sigmastar)*(A+B))+0.5*
trace(Sigmastar*Phi))^(-0.5*p*(p-mW-n)-0.5*nu0);
if min(fSigmastar/fSigmaW,1)>unifrnd(0,1)
SigmaW=Sigmastar; end;

meanmuW=meanmuW+muW/c;
meanSW=meanSW+SigmaW/c;


if (0.5*nu0+0.5*trace(inv(SigmaIW)*(A+B-Phi)))
^(-0.5*p*(p-mW-n)-0.5*nu0)>maxmW mapmuW=muW;
 maxmW=(0.5*nu0+0.5*trace(inv(SigmaIW)*(A+B-Phi)))
 ^(-0.5*p*(p-mW-n)-0.5*nu0); end;
f=det(SigmaW)^(0.5*(mW+n+1))*(0.5*nu0+0.5
*trace(inv(SigmaW)*(A+B))+0.5*trace(SigmaW*Phi))
^(-0.5*p*(p-mW-n)-0.5*nu0);
if f>maxSW mapSW=SigmaW; maxSW=f; end;
end;

TRB(j,1)=sum((Xbar-mu).^2);
TRB(j,2)=sum((meanmuIW-mu).^2);
TRB(j,3)=sum((meanmuW-mu).^2);
TRB(j,4)=sum((mapmuIW-mu).^2);
TRB(j,5)=sum((mapmuW-mu).^2);
TRB(j,6)=sum(sum((Sx-sigma).^2));
TRB(j,7)=sum(sum((meanSIW-sigma).^2));
TRB(j,8)=sum(sum((meanSW-sigma).^2));
TRB(j,9)=sum(sum((mapSIW-sigma).^2));
TRB(j,10)=sum(sum((mapSW-sigma).^2));


end;

r=[1:10];
Norms=zeros(length(r),10);
for j=1:length(r);
Norms(j,1)=norm((Xbar-mu).^r(j))^(1/r(j));
Norms(j,2)=norm((meanmuIW-mu).^r(j))^(1/r(j));
```

```
Norms(j,3)=norm((meanmuW-mu).^r(j))^(1/r(j));
Norms(j,4)=norm((mapmuIW-mu).^r(j))^(1/r(j));
Norms(j,5)=norm((mapmuW-mu).^r(j))^(1/r(j));
Norms(j,6)=norm((Sx-sigma).^r(j))^(1/r(j));
Norms(j,7)=norm((meanSIW-sigma).^r(j))^(1/r(j));
Norms(j,8)=norm((meanSW-sigma).^r(j))^(1/r(j));
Norms(j,9)=norm((mapSIW-sigma).^r(j))^(1/r(j));
Norms(j,10)=norm((mapSW-sigma).^r(j))^(1/r(j));
end;

figure;%muIW
plot(DF(5000:10000,1),DF(5000:10000,2),'--',
DF(5000:10000,1),
DF(5000:10000,3));
legend('Mean','Median');
figure;%SigmaIW
plot(DF(1:1000,1),DF(1:1000,4),'--',
DF(1:1000,1),DF(1:1000,5),
'-o',DF(1:1000,1),DF(1:1000,6),'-x');
legend('Det','Trace','Eig');

figure;%muW
plot(DF(5000:10000,1),DF(5000:10000,7),'--',
DF(5000:10000,1),
DF(5000:10000,8));
legend('Mean','Median');
figure;%SigmaW
plot(DF(5000:10000,1),DF(5000:10000,9),'--',
DF(5000:10000,1),
DF(5000:10000,10),'-o',DF(5000:10000,1),
DF(5000:10000,11),'-x');
legend('Det','Trace','Eig');

figure1=figure('Color',[1 1 1]);
plot(r,Norms(1:10,1),'-o',r,Norms(1:10,2),'-.*',
r,Norms(1:10,3),':');
legend('Sample mean','Inverse-Wishart prior',
'Wishart prior')
title('Expected value - SEL')
xlabel('r')
ylabel('Frobenius norm');

figure2=figure('Color',[1 1 1]);
plot(r,Norms(1:10,1),'-o',r,Norms(1:10,4),'-.*',
r,Norms(1:10,5),':')
legend('Sample mean','Inverse-Wishart prior',
'Wishart prior')
xlabel('r')
ylabel('Frobenius norm')
title('Expected value - MAP');

figure3=figure('Color',[1 1 1]);
plot(r,Norms(1:10,6),'-o',r,Norms(1:10,7),'-.*',
r,Norms(1:10,8),':')
legend('Sample mean','Inverse-Wishart prior',
'Wishart prior')
xlabel('r')
```

```
ylabel('Frobenius norm')
title('Covariance matrix - SEL');

figure4=figure('Color',[1 1 1]);
plot(r,Norms(1:10,6),'-o',r,Norms(1:10,9),'-.*',
r,Norms(1:10,10),':')
legend('Sample mean','Inverse-Wishart prior',
'Wishart prior');
xlabel('r')
ylabel('Frobenius norm')
title('Covariance matrix - MAP');
```

B.4 Iris dataset - Section 2.5.2

Matlab code:

```
X=[5.1     3.5     1.4     0.2,
 4.9     3.0     1.4     0.2,
 4.7     3.2     1.3     0.2,
 4.6     3.1     1.5     0.2,
 5.0     3.6     1.4     0.2,
 5.4     3.9     1.7     0.4,
 4.6     3.4     1.4     0.3,
 5.0     3.4     1.5     0.2,
 4.4     2.9     1.4     0.2,
 4.9     3.1     1.5     0.1];

[n,p]=size(X);
n0=5;
s=1;
c=100000;
nu0=7;

for j=1:z;
Xbar=ones(1,n)*X/n;
V=(X-ones(n,1)*Xbar)'*(X-ones(n,1)*Xbar);
Phi=V/n;
omega=eye(s);
theta=Xbar;
mW=5;
mIW=9;
SigmaW=V/n;
SigmaIW=V/n;
muIW=Xbar;
muW=Xbar;

b=1/(n+n0)*(n*Xbar+n0*theta);
Sx=V/(n-1);

meanmuIW=zeros(1,p);
meanmuW=zeros(1,p);
meanSIW=zeros(p,p);
meanSW=zeros(p,p);

mapmuIW=meanmuIW;
mapmuW=meanmuIW;
```

```
mapSIW=meanSIW;
mapSW=meanSW;

maxmIW=-10000000000000;
maxmW=-100000000000000;
maxSIW=-10000000000000;
maxSW=-100000000000000;

for i=1:c;

%Inverse-Wishart;
mustar1=mvnrnd(b,1/(n+n0)*SigmaIW,1);

D1=n*(muIW-theta)'*(muIW-theta);
A1=V+n*(Xbar-muIW)'*(Xbar-muIW);
D2=n*(mustar1-theta)'*(mustar1-theta);
A2=V+n*(Xbar-mustar1)'*(Xbar-mustar1);

fmuIW=(0.5*nu0+0.5*trace(inv(SigmaIW)*(A1+D1)))
^(-0.5*p*(p-mIW-n)-0.5*nu0);
fmustar=(0.5*nu0+0.5*trace(inv(SigmaIW)
*(A2+D2)))^(-0.5*p*(p-mIW-n)-0.5*nu0);

if min(fmustar/fmuIW,1)>unifrnd(0,1)
 muIW=mustar1; end;

B=n*(muIW-theta)'*(muIW-theta)+Phi;
A=V+n*(Xbar-muIW)'*(Xbar-muIW);

w=0.5;
Sigmastar=(1-w)*iwishrnd(Phi,mIW)
+w*wishrnd(Phi,mIW);

fSigmaIW=det(SigmaIW)^(0.5*(mIW+n+1))*(0.5*nu0
+0.5*trace(inv(SigmaIW)*(A+B)))^
(-0.5*p*(p-mIW-n)-0.5*nu0);
fSigmastar=det(Sigmastar)^(0.5*(mIW+n+1))
*(0.5*nu0+0.5*trace(inv(Sigmastar)*(A+B)))
^(-0.5*p*(p-mIW-n)-0.5*nu0);
if min(fSigmastar/fSigmaIW,1)>unifrnd(0,1)
SigmaIW=Sigmastar; end;


meanmuIW=meanmuIW+muIW*(1/c);
meanSIW=meanSIW+SigmaIW*(1/c);

if (0.5*nu0+0.5*trace(inv(SigmaIW)
*(A+B-Phi)))^(-0.5*p*(p-mIW-n)-0.5*nu0)>maxmIW
 mapmuIW=muIW;
 maxmIW=(0.5*nu0+0.5*trace(inv(SigmaIW)
 *(A+B-Phi)))^(-0.5*p*(p-mIW-n)-0.5*nu0);
   end;
f=det(SigmaIW)^(0.5*(mIW+n+1))*(0.5*nu0
+0.5*trace(inv(SigmaIW)*(A+B)))^
(-0.5*p*(p-mIW-n)-0.5*nu0);
if f>maxSIW mapSIW=SigmaIW ;maxSIW=f; end;
```

```
%Wishart;
mustar=mvnrnd(b,1/(n+n0)*SigmaW,1);

D1=n*(muW-theta)'*(muW-theta);
A1=V+n*(Xbar-muW)'*(Xbar-muW);
D2=n*(mustar-theta)'*(mustar-theta);
A2=V+n*(Xbar-mustar)'*(Xbar-mustar);

fmuW=(0.5*nu0+0.5*trace(inv(SigmaW)*(A1+D1)))
^(-0.5*p*(p-mW-n)-0.5*nu0);
fmustar=(0.5*nu0+0.5*trace(inv(SigmaW)
*(A2+D2)))^(-0.5*p*(p-mW-n)-0.5*nu0);

if min(fmustar/fmuW,1)>unifrnd(0,1) muW=mustar; end;


B=n*(muW-theta)'*(muW-theta)+Phi;
A=V+n*(Xbar-muW)'*(Xbar-muW);

w=0.5;

Sigmastar=w*iwishrnd(Phi,mW*2+1)
+(1-w)*wishrnd(Phi,mW);
fSigmaW=det(SigmaW)^(0.5*(mW+n+1))*(0.5*nu0+
0.5*trace(inv(SigmaW)*(A+B))
+0.5*trace(SigmaW*Phi))^(-0.5*p*(p-mW-n)
-0.5*nu0);
fSigmastar=det(Sigmastar)^(0.5*(mW+n+1))
*(0.5*nu0+0.5*trace(inv(Sigmastar)*(A+B))
+0.5*trace(Sigmastar*Phi))^
(-0.5*p*(p-mW-n)-0.5*nu0);

if min(fSigmastar/fSigmaW,1)>0.1*unifrnd(0,1)
SigmaW=Sigmastar; end;

meanmuW=meanmuW+muW/c;
meanSW=meanSW+SigmaW/c;


if (0.5*nu0+0.5*trace(inv(SigmaIW)*(A+B-Phi)))
^(-0.5*p*(p-mW-n)-0.5*nu0)>maxmW mapmuW=muW;
 maxmW=(0.5*nu0+0.5*trace(inv(SigmaIW)
  *(A+B-Phi)))^(-0.5*p*(p-mW-n)-0.5*nu0); end;
f=det(SigmaW)^(0.5*(mW+n+1))*(0.5*nu0
+0.5*trace(inv(SigmaW)*(A+B))+0.5
*trace(SigmaW*Phi))^(-0.5*p*(p-mW-n)-0.5*nu0);
if f>maxSW mapSW=SigmaW; maxSW=f; end;
end;

end;
```

Appendix C

Code for Chapter 3

C.1 Code for Figures 3.1 and 3.2

Matlab code:

```
%Grid of values
inc=0.1;
l1=5:inc:10;
l2=0:inc:5;

%Parameters
n=10;
m=2;
a=1;
b=2;
c1=1;
c2=-2;

g=zeros(length(l1),length(l2));
sum=0;

for i=1:length(l1);
for j=1:length(l2);

%1F1(a,b,x)
g(i,j)=(l1(i)-l2(j))*c1^(-n*m/2)*((l1(i)*l2(j))^(n/2-(m+1)/2))
*hypergeom(a,b,1/c1*(l1(i)+l2(j)))*exp(c2*(l1(i)+l2(j)));end;

%1+x
%g(i,j)=(l1(i)-l2(j))*c1^(-n*m/2)*((l1(i)*l2(j))^(n/2-(m+1)/2))
*(1+1/c1*(l1(i)+l2(j)))*exp(c2*(l1(i)+l2(j)));end;

%exp(x)
%g(i,j)=(l1(i)-l2(j))*c1^(-n*m/2)*((l1(i)*l2(j))^(n/2-(m+1)/2))
*exp(1/c1*(l1(i)+l2(j)))*exp(c2*(l1(i)+l2(j)));end;
sum=sum+g(i,j);
end;
end;

g=g/(inc*inc*sum);

figure;
surfc(l2,l1,g)
```

C.2 Code for Figure 3.3

Matlab code:

```
%Parameters
m=2;
n=5;
a=1;
b=2;
c=0.5;
inc=0.1;
sigma=eye(m);

%Grid
x1=0:inc:10;
x2=0:inc:10;
sum=0;
g=zeros(length(x1),length(x2));

for i=1:length(x1);
for j=1:length(x2);
X=[x1(i) 0;0 x2(j)];
g(i,j)=det(X)^(n/2-(m+1)/2)*hypergeom(a,b,c*trace(inv(sigma)*X))
*exp(trace(-inv(sigma)*X));
sum=sum+g(i,j);
end;
end;

g=g/(sum*inc*inc);

figure;
surfc(x1,x2,g)
```

C.3 Evaluation of algorithm and methodology - Section 3.6.1

Matlab code:

```
%Target values
mu=0;
s=1;

%Parameters
Sprior=0.9;
a=1;%
b=2;%
phi=0.105;
n=Sprior*2/phi;

%Sample
m=18;
x=normrnd(mu,s,[1 m]); %Simulated data;
xbar=sum(x)/m;
ss=(sum((x-xbar).^2))/(m-1);

%MCMC
nz=50;
rep=10000;
```

```
burn=0.1*rep;
mus=ones(rep+burn,1)*xbar;
sigmas=ones(rep+burn,1)*ss;
countzm=0;
countzs=0;
ints=zeros(nz,4);

for l=1:nz;
mus=ones(rep+burn,1)*xbar;
sigmas=ones(rep+burn,1)*ss;

for ii=2:rep+burn;

u1=unifrnd(xbar-ss,xbar+ss);
if min(exp(-0.5/(sigmas(ii-1))*sum((x-u1).^2))
/exp(-0.5/(sigmas(ii-1))*sum((x-mus(ii-1)).^2)),1)
>unifrnd(0,1);
mus(ii)=u1; else mus(ii)=mus(ii-1); end;

u2=unifrnd(0,4*ss);
if min((u2).^(n/2-m/2-0.5)*hypergeom(a,b,c/phi*u2)
*exp(-0.5/u2*(sum((x-mus(ii)).^2))-u2/phi)/((sigmas(ii-1))
.^(n/2-m/2-0.5)*hypergeom(a,b,c/phi*sigmas(ii-1))
*exp(-0.5/sigmas(ii-1)*(sum((x-mus(ii)).^2))-
sigmas(ii-1)/phi)),1)>unifrnd(0,1);
sigmas(ii)=u2; else sigmas(ii)=sigmas(ii-1); end;
end;

int1=x1;% -1:0.1:1;
int2=ss1;
cumd1=x1*0;
cumd2=ss1*0;
mus=mus(burn+1:length(mus));
sigmas=sigmas(burn+1:length(sigmas));

for jj=2:length(int1);
count=0;

for ii=1:length(mus);
if mus(ii)<int1(jj) if mus(ii)>int1(jj-1) count=count+1;
end; end;

cumd1(jj)=count/rep;
end;
end;

for jj=1:length(int2);
count1=0;

for ii=2:length(sigmas);
if sigmas(ii)<int2(jj) if sigmas(ii)>int2(jj-1)
count1=count1+1;
end; end;

cumd2(jj)=count1/rep;
end;
end;
```

```
%Analytical
inc=0.01;

nz=1;
inc2=0.2;

x1=-5:inc:5;
ss1=0.005:inc:2;
ss2=0.005:inc2:2;

postm1=x1;
posts1=ss1;

postm=zeros(length(x1),length(nz));
dhx=zeros(length(x1),length(x1))+triu(ones(length(x1),
length(x1)));
posts=zeros(length(ss1),length(nz));
dhs=zeros(length(ss1),length(ss1))+triu(ones(length(ss1),
length(ss1)));


for z=1:nz

c=max([max(ss2)/phi max(ss1)/phi]);
c=phi/c-0.0001;

%Posterior mu
for j=1:length(x1)
sum1=zeros(length(ss2),1);
for ii=1:length(ss2);
sum1(ii)=phi.^(m/2-n/2)/gamma(n/2-m/2)*ss2(ii).^(n/2-m/2)
*hypergeom(a,b,c/phi*ss2(ii))*exp(-0.5/ss2(ii)*
sum((x-x1(j)).^2) - ss2(ii)/phi);
end;
aa=sum1'*inc*dhs1;
aa=aa(length(ss2));
postm1(j)=(2*22/7/m).^(-0.5)*phi.^(-0.5)*gamma(n/2-m/2)
/gamma(n/2-m/2+0.5)*(aa);
end;

%Posterior sigma
for k=1:length(ss1)

posts1(k)=ss1(k).^(n/2-m/2 -0.5)*hypergeom(a,b,c/phi*ss1(k))
*exp(-0.5/ss1(k)*(sum(x.^2)-m*xbar^2)-ss1(k)/phi);

end;

%Normalising
aream=(postm1*inc)*dhx;
areas=(posts1*inc)*dhs;

postm(:,z)=postm1/aream(length(x1));
posts(:,z)=(posts1').*/areas(length(ss1));

figure;
```

```
subplot(2,1,1)
plot(x1,postm);
xlabel('Population mean');
ylabel('Density');
title 'Exact posterior density function';
subplot (2,1,2)
bar(x1,cumd1);
xlabel('Population mean');
ylabel('Relative frequency');
title 'Histogram of MCMC sample';

figure;
plotyy(int1 ,cumd1,int1 ,postm,'bar','plot');
figure;
plotyy(int2 ,cumd2,int2 ,posts ,'bar','plot');

figure;
subplot(2,1,1)
plot(int2 ,posts);
xlabel('Population variance');
ylabel('Density');
title 'Exact posterior density function';
subplot (2,1,2)
bar(int2 ,cumd2);
xlabel('Population variance');
ylabel('Relative frequency');
title 'Histogram of MCMC sample';
```

C.4 Choice of hyperparameter values - Section 3.6.1

Matlab code:

```
n=1:1:50;
y=n;
a=1;
b=1;
c=0.1;

for i=1:length(n);
y(i)=hypergeom([n(i)/2+1 a],b,c)/hypergeom([n(i)/2 a],b,c);
end;

plot(n,y)
```

C.5 Univariate simulation study - Section 3.6.1

Matlab code:

```
%Target values
mu=0;
s=1;

%Sample
m=18;
```

```
Sprior=0.9;
a=1;
b=2;
phi=0.105;
n=19; %Sprior*2/phi

inc=0.01;
nz=1; %100
inc2=0.2;

alpha=0.05;

est=zeros(nz,5);
int=zeros(nz,6);
x1=-5:inc:5;
ss1=0.005:inc:2;
ss2=0.005:inc2:2;

postm1=x1;
posts1=ss1;
posts2=ss1;

counthg=0;
postm=zeros(length(x1),length(nz));
posts=zeros(length(ss1),length(nz));
dhx=zeros(length(x1),length(x1))+triu(ones(length(x1),
length(x1)));
dhs=zeros(length(ss1),length(ss1))+triu(ones(length(ss1),
length(ss1)));
dhs1=zeros(length(ss2),length(ss2))+triu(ones(length(ss2),
length(ss2)));

for z=1:nz
x=normrnd(mu,s,[1 m]); %Simulated data;
xbar=sum(x)/m;
ss=(sum((x-xbar).^2))/(m-1);

%Posterior mu
for j=1:length(x1)
sum1=zeros(length(ss2),1);
for ii=1:length(ss2);
sum1(ii)=phi.^(m/2-n/2)/gamma(n/2-m/2)*ss2(ii).^(n/2-m/2)
*hypergeom(a,b,c/phi*ss2(ii))*exp(-0.5/ss2(ii)*
sum((x-x1(j)).^2) - ss2(ii)/phi);
end;
aa=sum1'*inc*dhs1;
aa=aa(length(ss2));
postm1(j)=(2*22/7/m).^(-0.5)*phi.^(-0.5)*gamma(n/2-m/2)
/gamma(n/2-m/2+0.5)*(aa);
end;

%Posterior sigma
for k=1:length(ss1)

posts1(k)=ss1(k).^(n/2-m/2-0.5)*hypergeom(a,b,c/phi*ss1(k))
*exp(-0.5/ss1(k)*(sum(x.^2)-m*xbar^2)-ss1(k)/phi); %HG
posts2(k)=ss1(k).^(n/2-m/2+0.5)*hypergeom(a,b,c/phi*ss1(k))
```

```
*exp(-0.5/ss1(k)*(sum(x.^2)-m*xbar^2)-ss1(k)/phi); %HG est

end;

%Normalising constants
aream=(postm1*inc)*dhx;
areas=(posts1*inc)*dhs;
areas2=(posts2*inc)*dhs;

postm(:,z)=postm1/aream(length(x1));
posts(:,z)=(posts1').areas(length(ss1));

sum2=areas(length(ss1));
sum3=areas2(length(ss1));

aream=(postm(:,z)')*inc*dhx;
areas=(posts(:,z)')*inc*dhs;

telmu=0;
telml=0;
telsl=0;
telsu=0;

%credible intervals
%mu
for i=1:length(x1);
if aream(i)<alpha/2; telml=telml+1; end;
if aream(i)<(1-alpha/2); telmu=telmu+1; end;
end;

%HG
for i=1:length(ss1);
if areas(i)<alpha/2; telsl=telsl+1; end;
if areas(i)<(1-alpha/2); telsu=telsu+1; end;
end;
if s<ss1(telsu); if s>ss1(telsl); counthg=counthg+1;
end;
end;

%Bayes estimates
mu111=xbar;
est(z,1)=ss; %MLE
est(z,2)=sum3/sum2; %HG

end;
end;

%Posterior df of sigma
figure;
plot(ss1,posts);
xlabel('Population variance');
ylabel('Density');

%Posterior df of mu
figure13=figure('Color',[1 1 1]);
plot(x1,postm);
xlabel('Population mean');
```

```
ylabel('Density');

%Coverage probabilities
coveragehgamma=counthg/nz*100;

%Median credible interval widths
medianwidthhgamma=median(width(:,1)); %HG
```

C.6 Forestry dataset - Section 3.6.1

Matlab code:

```
%Real data - trees volume;

x=[10.3
10.3
10.2
16.4
18.8
19.7
15.6
18.2
22.6
19.9
24.2
21
21.4
21.3
19.1
22.2
33.8
27.4
25.7
24.9
34.5
31.7
36.3
38.3
42.6
55.4
55.7
58.3
51.5
51
77];
x=log(x);
m=length(x);

xbar=sum(x)/m;
ss=(sum((x-xbar).^2))/(m-1);

Sprior=0.3;

%Priors
inc=0.005;
```

```
y=0.005:inc:2;
Sprior=0.3;

%Likelihood
beta=11.5;
alpha=beta/ss+1;

%HG
a=1;
b=1;
phi=0.01875;
c=0.01;
alpha1=32/2;
beta1=1/phi;

f=y.^(alpha1-1).*exp(-beta1.*y).*hypergeom(a,b,c/phi.*y);
g=y.^(-alpha-1).*exp(-beta./y);

c1=sum(f.*inc);
c2=sum(g.*inc);

f=f/c1;
g=g/c2;

figure;
plot(y,f,y,j,y,g1,'--',y,h,':',y,g);
legend('Hypergeometric gamma','Kummer gamma','Inverse
gamma','Gamma','Likelihood');

%Bayesian analysis

a=1;%HG
b=1;%HG
c=0.01;
phi=0.01875; %HG
n=Sprior*2/phi; %HG

inc=0.005;
nz=1;
inc2=0.005;

alpha=0.05;

est=zeros(nz,5);
int=zeros(nz,6);
x1=-5:inc:5;
ss1=0.005:inc:2;
ss2=0.005:inc2:2;

postm1=x1;
posts1=ss1;
posts2=posts1;

width=zeros(length(ss1),4);
postm=zeros(length(x1),length(nz));
posts=zeros(length(ss1),length(nz));
```

```
dhx=zeros(length(x1),length(x1))+triu(ones(length(x1),
length(x1)));
dhs=zeros(length(ss1),length(ss1))+triu(ones(length(ss1),
length(ss1)));
dhs1=zeros(length(ss2),length(ss2))+triu(ones(length(ss2),
length(ss2)));

for z=1:nz

%Posterior mu
for j=1:length(x1)
sum1=zeros(length(ss2),1);
for ii=1:length(ss2);
sum1(ii)=phi.^(m/2-n/2)/gamma(n/2-m/2)*ss2(ii).^(n/2-m/2)
*hypergeom(a,b,c/phi*ss2(ii))*exp(-0.5/ss2(ii)*
sum((x-x1(j)).^2) - ss2(ii)/phi);
end;
aa=sum1'*inc*dhs1;
aa=aa(length(ss2));
postm1(j)=(2*22/7/m).^(-0.5)*phi.^(-0.5)*gamma(n/2-m/2)
/gamma(n/2-m/2+0.5)*(aa);
end;

%Posterior sigma
for k=1:length(ss1)

posts1(k)=ss1(k).^(n/2-m/2 -0.5)*hypergeom(a,b,c/phi*ss1(k))
*exp(-0.5/ss1(k)*(sum(x.^2)-m*xbar^2)-ss1(k)/phi); %HG
posts2(k)=ss1(k).^(n/2-m/2+0.5)*hypergeom(a,b,c/phi*ss1(k))
*exp(-0.5/ss1(k)*(sum(x.^2)-m*xbar^2)-ss1(k)/phi); %HG est

end;

%Normalising constants
aream=(postm1*inc)*dhx;
areas=(posts1*inc)*dhs;
areas2=(posts2*inc)*dhs;

postm(:,z)=postm1/aream(length(x1));
posts(:,z)=(posts1').*/areas(length(ss1));%HG

aream=(postm(:,z)')*inc*dhx;
areas=(posts(:,z)')*inc*dhs;

telmu=0;
telml=0;
telsl=0;
telsu=0;

%credible intervals
%mu
for i=1:length(x1);
if aream(i)<alpha/2; telml=telml+1; end;
if aream(i)<(1-alpha/2); telmu=telmu+1; end;
end;

%HG
```

```
for i=1:length(ss1);
if areas(i)<alpha/2; telsl=telsl+1; end;
if areas(i)<(1-alpha/2); telsu=telsu+1; end;
end;

%credible interval widths
width(z,1)=ss1(telsu)-ss1(telsl); %HG


%Bayes estimates
mu111=xbar;
est(z,1)=ss; %MLE
est(z,2)=sum3/sum2; %HG

end;
end;

%Posterior df, prior and likelihood of sigma2
figure;
plot(ss1,posts,ss1,f,ss1,g);
xlabel('Population variance');
ylabel('Density');
legend('Posterior','Hypergeometric prior', 'Likelihood');
```

C.7 Multivariate simulation study - Section 3.6.2

Matlab code:

```
n0=1;
s=1;
p=3;
sigma=eye(p);
mu=zeros(1,p);
c=100;
z=100;
aa=1;

c1=0.9;
a=1;
b=2;

n=20;
DF=zeros(c,4);

for iz=1:z;

X=mvnrnd(mu,sigma,n);
Xbar=ones(1,n)*X/n;
V=(X-ones(n,1)*Xbar)'*(X-ones(n,1)*Xbar);
omega=eye(s);
mW=p;
mIW=2*p+1;
mHW=p;
mKW=p;

%Initial values;
```

```
HWSigmaW=4*eye(p);

b=Xbar;
Sx=V/(n-1);
thetaHW=0.5*eye(p);

Y=n*n0/(n+n0)*(Xbar-theta)*(Xbar-theta)'+V;

meanHW=zeros(p,p);

%Single convergence measures for initial values
DF(1,1)=det(HWSigma);
DF(1,2)=trace(HWSigma);
detsen=eigs(HWSigma);
DF(1,3)=detsen(p);
DF(1,4)=1;

%Gibbs sampling
for i=2:c;

%HW
muHW=mvnrnd(b,1/n*HWSigma,1);
A=V+n*(Xbar-muHW)'*(Xbar-muHW);
HWStar=wishrnd(thetaHW,mHW);

gsigmaHW=det(HWSigma)^(-0.5*(mHW-n*s-p-1))*exp(trace(
-0.5*inv(thetaHW)*HWSigma))*hypergeom(a,b,c1*trace(inv(
thetaHW)*HWSigma));%*exp(-0.5*trace((A)*inv(HWSigma)));

gsigmastar=det(HWStar)^(-0.5*(mHW-n*s-p-1))*exp(trace(
-0.5*inv(thetaHW)*HWStar))*hypergeom(a,b,c1*trace(inv(
thetaHW)*HWStar));%*exp(-0.5*trace((A)*inv(HWStar)));

if min(gsigmastar/gsigmaHW,1)>unifrnd(0,1),
HWSigma=HWStar; end;

%Single convergence mesures
DF(i,1)=det(HWSigma);
DF(i,2)=trace(HWSigma);
detsen=eigs(HWSigma);
DF(i,3)=detsen(p);
DF(i,4)=i;

%Burn-in period
if i>0.1*c, meanHW=meanHW+HWSigma/c; end;

end;

FN1=norm(Sx-sigma)
FN2=norm(meanHW-sigma)
end;

%Graphs of single convergence measures
%HW;
figure;
plot(DF(:,4),DF(:,1),DF(:,4),DF(:,2),'-o',DF(:,4),DF(:,3),'-x');
legend('Det','Trace','Eig');
```

C.8 Abalone dataset - Section 3.6.2

Matlab code:

```
X1=horzcat(length1 ,diameter ,height ,wweight ,sweight ,vweight);
X=X1(21:40,:);
Roystest(X,0.05); %Test for normality

[n,p]=size(X);
n0=1;
s=1;
theta0=0.1;
c=1000;
z=1;
aa=1;

c1=0.9;
a=1;
b=2;

for iz=1:z;

Xbar=ones(1,n)*X/n;
V=(X-ones(n,1)*Xbar)'*(X-ones(n,1)*Xbar);
omega=eye(s);
mHW=p;


%Initial values;
HWSigma=V/n;

b=Xbar;
Sx=V/(n-1);
thetaHW=Sx;

Y=n*n0/(n+n0)*(Xbar-theta)*(Xbar-theta)'+V;

meanHW=zeros(1,p);

for i=2:c;
%HW
muHW=mvnrnd(b,1/n*HWSigma,1);
A=V+n*(Xbar-muHW)'*(Xbar-muHW);

HWStar=wishrnd(thetaHW ,mHW);

gsigmaHW=det(HWSigma)^(-0.5*(mHW-n*s-p-1))*exp(trace(
-0.5*inv(thetaHW)*HWSigma))*hypergeom(a,b,c1*trace(
inv(thetaHW)*HWSigma))*exp(-0.5*trace((A)*inv(HWSigma)));
gsigmastar=det(HWStar)^(-0.5*(mHW-n*s-p-1))*exp(trace(
-0.5*inv(thetaHW)*HWStar))*hypergeom(a,b,c1*trace(inv(
thetaHW)*HWStar))*exp(-0.5*trace((A)*inv(HWStar)));
if min(gsigmastar/gsigmaHW,1)>0.1*unifrnd(0,1),

HWSigma=HWStar; end;

%Bayes estimates
```

```
if i>0.1*c, meanHW=meanHW+HWSigma/c; end;
end;
end;
```

Appendix D

Code for Chapter 4

D.1 Code for Example 1

Matlab code:

```
periods=200;
clients=100;
alpha=10;
beta=10;
Times=gamrnd(alpha,beta,clients,periods);
count=0;
Rem=zeros(periods*clients,1);

for i=1:periods;
for j=1:clients;
if Times(j,i)+i>200 count=count+1;
Rem(count)=Times(j,i); end;
end;
end;

X=Rem(1:count,:);
grid=0:0.5:300;
histogram(X,'Normalization','pdf','FaceAlpha'
,0.2,'FaceColor','Blue');
f=gampdf(grid,alpha,beta);%gamma(alpha)/beta^alpha
*grid.^(alpha-1).*exp(-beta*grid);
g=gampdf(grid,alpha+1,beta);%gamma(alpha+1)/beta^(alpha+1)
*grid.^(alpha).*exp(-beta*grid);
hold on
plot(grid,f,'r--','Linewidth',1.5)
hold on
plot(grid,g,'k','Linewidth',2.5)
legend('Histogram of times','Gamma distribution'
,'Length-biased gamma distribution');
fitdist(X,'gamma')
```

D.2 Code for Tables 4.1, 4.2 and 4.3

Matlab code:

```
%Grid of values
inc=0.1;
l1=5:inc:10;
l2=0:inc:5;
```

```
%Parameters
n=9;
m=2;
c1=1;
c2=-2;

g=zeros(length(l1),length(l2));
sum=0;

for i=1:length(l1);
for j=1:length(l2);

%sin(x)
g(i,j)=(l1(i)-l2(j))*c1^(-n*m/2)*((l1(i)*l2(j))^(n/2-(m+1)/2))
*sin(c2*(l1(i)+l2(j)))*exp(-c1/2*(l1(i)+l2(j)));end;

%exp(x)
%g(i,j)=(l1(i)-l2(j))*c1^(-n*m/2)*((l1(i)*l2(j))^(n/2-(m+1)/2))
*exp(c2*(l1(i)+l2(j)))*exp(-c1/2*(l1(i)+l2(j)));end;

%exp(1/x)
%g(i,j)=(l1(i)-l2(j))*c1^(-n*m/2)*((l1(i)*l2(j))^(n/2-(m+1)/2))
*exp(1/(c2*(l1(i)+l2(j))))*exp(-c1/2*(l1(i)+l2(j)));end;

%1+x
%g(i,j)=(l1(i)-l2(j))*c1^(-n*m/2)*((l1(i)*l2(j))^(n/2-(m+1)/2))
*(1+c2*(l1(i)+l2(j)))*exp(-c1/2*(l1(i)+l2(j)));end;

sum=sum+g(i,j);
end;
end;

g=g/(inc*inc*sum);

figure;
surfc(l2,l1,g)
```

D.3 Univariate simulation study - Section 4.7.1

Matlab code:

```
%Target values
mu=0;
s=1;

%Sample
m=18;

Sprior=0.9;
varphi=0.185;
n=4;
gamma1=1;
inc=0.01;
nz=1; %100
inc2=0.2;
```

```
alpha=0.05;

est=zeros(nz,5);
int=zeros(nz,6);
x1=-5:inc:5;
ss1=0.005:inc:2;
ss2=0.005:inc2:2;

postm1=x1;
posts1=ss1;
posts2=ss1;

countwg=0;
postm=zeros(length(x1),length(nz));
posts=zeros(length(ss1),length(nz));
dhx=zeros(length(x1),length(x1))+triu(ones(length(x1),
length(x1)));
dhs=zeros(length(ss1),length(ss1))+triu(ones(length(ss1),
length(ss1)));
dhs1=zeros(length(ss2),length(ss2))+triu(ones(length(ss2),
length(ss2)));

for z=1:nz
x=normrnd(mu,s,[1 m]); %Simulated data;
xbar=sum(x)/m;
ss=(sum((x-xbar).^2))/(m-1);

%Posterior mu
for j=1:length(x1)
sum1=zeros(length(ss2),1);
for ii=1:length(ss2);
sum1(ii)=ss2(ii).^(alp1-m/2-1)*exp(-0.5/ss2(ii)
*((sum(x.^2)-m*xbar^2))-lambda*ss2(ii))*(1+ss2(ii))
^gamma1;;
end;
aa=sum1'*inc*dhs1;
aa=aa(length(ss2));
postm1(j)=(2*22/7/m).^(-0.5)*phi.^(-0.5)*gamma(n/2-m/2)
/gamma(n/2-m/2+0.5)*(aa);
end;

%Posterior sigma
for k=1:length(ss1)

posts1(k)=ss1(k).^(alp1-m/2-1)*exp(-0.5/ss1(k)
*((sum(x.^2)-m*xbar^2))-lambda*ss1(k))*(1+ss1(k))
^gamma1; %KG
posts2(k)=ss1(k).^(alp1-m/2)*exp(-0.5/ss1(k)
*((sum(x.^2)-m*xbar^2))-lambda*ss1(k))*(1+ss1(k))
^gamma1; %KG est

end;

%Normalising constants
aream=(postm1*inc)*dhx;
areas=(posts1*inc)*dhs;
```

```
areas2=(posts2*inc)*dhs;

postm(:,z)=postm1/aream(length(x1));
posts(:,z)=(posts1').//areas(length(ss1));

sum2=areas(length(ss1));
sum3=areas2(length(ss1));

aream=(postm(:,z)')*inc*dhx;
areas=(posts(:,z)')*inc*dhs;

telmu=0;
telml=0;
telsl=0;
telsu=0;

%credible intervals
%mu
for i=1:length(x1);
if aream(i)<alpha/2; telml=telml+1; end;
if aream(i)<(1-alpha/2); telmu=telmu+1; end;
end;

%WG
for i=1:length(ss1);
if areas(i)<alpha/2; telsl=telsl+1; end;
if areas(i)<(1-alpha/2); telsu=telsu+1; end;
end;
if s<ss1(telsu); if s>ss1(telsl); countwg=countwg+1;
end;
end;

%Bayes estimates
mu111=xbar;
est(z,1)=ss; %MLE
est(z,2)=sum3/sum2; %KG

end;
end;

%Posterior df of sigma
figure;
plot(ss1,posts);
xlabel('Population variance');
ylabel('Density');

%Posterior df of mu
figure13=figure('Color',[1 1 1]);
plot(x1,postm);
xlabel('Population mean');
ylabel('Density');

%Coverage probabilities
coveragewgamma=countwg/nz*100;

%Median credible interval widths
medianwidthwgamma=median(width(:,1)); %KG
```

D.4 Forestry dataset - Section 4.7.1

Matlab code:

```
%Real data - trees volume;

x=[10.3
10.3
10.2
16.4
18.8
19.7
15.6
18.2
22.6
19.9
24.2
21
21.4
21.3
19.1
22.2
33.8
27.4
25.7
24.9
34.5
31.7
36.3
38.3
42.6
55.4
55.7
58.3
51.5
51
77];
x=log(x);
m=length(x);

xbar=sum(x)/m;
ss=(sum((x-xbar).^2))/(m-1);

Sprior=0.9;
varphi=0.185;
n=4;
gamma1=1;
inc=0.01;
nz=1; %100
inc2=0.2;

alpha=0.05;

est=zeros(nz,5);
int=zeros(nz,6);
x1=-5:inc:5;
ss1=0.005:inc:2;
ss2=0.005:inc2:2;
```

```
postm1=x1;
posts1=ss1;
posts2=ss1;

countwg=0;
postm=zeros(length(x1),length(nz));
posts=zeros(length(ss1),length(nz));
dhx=zeros(length(x1),length(x1))+triu(ones(length(x1),
length(x1)));
dhs=zeros(length(ss1),length(ss1))+triu(ones(length(ss1),
length(ss1)));
dhs1=zeros(length(ss2),length(ss2))+triu(ones(length(ss2),
length(ss2)));

%Priors
inc=0.005;
y=0.005:inc:2;
Sprior=0.3;

%Likelihood
beta=11.5;
alpha=beta/ss+1;

%KG
alpha3=0.625;
beta3=4;

f=y.^(alpha3-1).*exp(-beta3.*y).*(1.+y);
g=y.^(-alpha-1).*exp(-beta./y);

c1=sum(f.*inc);
c2=sum(g.*inc);

f=f/c1;
g=g/c2;

%Bayesian analysis
for z=1:nz
x=normrnd(mu,s,[1 m]); %Simulated data;
xbar=sum(x)/m;
ss=(sum((x-xbar).^2))/(m-1);

%Posterior mu
for j=1:length(x1)
sum1=zeros(length(ss2),1);
for ii=1:length(ss2);
sum1(ii)=ss2(ii).^(alp1-m/2-1)*exp(-0.5/ss2(ii)
*((sum(x.^2)-m*xbar^2))-lambda*ss2(ii))*(1+ss2(ii))
^gamma1;;
end;
aa=sum1'*inc*dhs1;
aa=aa(length(ss2));
postm1(j)=(2*22/7/m).^(-0.5)*phi.^(-0.5)*gamma(n/2-m/2)
/gamma(n/2-m/2+0.5)*(aa);
end;
```

```
%Posterior sigma
for k=1:length(ss1)

posts1(k)=ss1(k).^(alp1-m/2-1)*exp(-0.5/ss1(k)
*((sum(x.^2)-m*xbar^2))-lambda*ss1(k))*(1+ss1(k))
^gamma1; %KG
posts2(k)=ss1(k).^(alp1-m/2)*exp(-0.5/ss1(k)
*((sum(x.^2)-m*xbar^2))-lambda*ss1(k))*(1+ss1(k))
^gamma1; %KG est

end;

%Normalising constants
aream=(postm1*inc)*dhx;
areas=(posts1*inc)*dhs;
areas2=(posts2*inc)*dhs;

postm(:,z)=postm1/aream(length(x1));
posts(:,z)=(posts1').Iareas(length(ss1));

sum2=areas(length(ss1));
sum3=areas2(length(ss1));

aream=(postm(:,z)')*inc*dhx;
areas=(posts(:,z)')*inc*dhs;

telmu=0;
telml=0;
telsl=0;
telsu=0;

%credible intervals
%mu
for i=1:length(x1);
if aream(i)<alpha/2; telml=telml+1; end;
if aream(i)<(1-alpha/2); telmu=telmu+1; end;
end;

%KG
for i=1:length(ss1);
if areas(i)<alpha/2; telsl=telsl+1; end;
if areas(i)<(1-alpha/2); telsu=telsu+1; end;
end;
if s<ss1(telsu); if s>ss1(telsl); countwg=countwg+1;
end;
end;

%Bayes estimates
mu111=xbar;
est(z,1)=ss; %MLE
est(z,2)=sum3/sum2; %KG

end;
end;

%Posterior df of sigma
figure;
```

```
plot(ss1,posts);
xlabel('Population variance');
ylabel('Density');

%Posterior df of mu
figure13=figure('Color',[1 1 1]);
plot(x1,postm);
xlabel('Population mean');
ylabel('Density');

%Coverage probabilities
coveragewgamma=countwg/nz*100;

%Median credible interval widths
medianwidthwgamma=median(width(:,1)); %KG

%Posterior df, prior and likelihood of sigma2
figure;
plot(ss1,posts,ss1,f,ss1,g);
xlabel('Population variance');
ylabel('Density');
legend('Posterior','Kummer gamma prior', 'Likelihood');
```

D.5 Multivariate simulation study - Section 4.7.2

Matlab code:

```
n0=1;
s=1;
p=3;
sigma=eye(p);
mu=zeros(1,p);
c=100;
z=100;
aa=1;

n=10;
DF=zeros(c,4);

for iz=1:z;

X=mvnrnd(mu,sigma,n);
Xbar=ones(1,n)*X/n;
V=(X-ones(n,1)*Xbar)'*(X-ones(n,1)*Xbar);
omega=eye(s);
theta=ones(1,p)*theta0;
mKW=p;

%Initial values;
KWSigma=1.1*eye(p);

b=Xbar;
Sx=V/(n-1);
PhiKW=eye(p);
thetaKW=eye(p);
```

```
Y=n*n0/(n+n0)*(Xbar-theta)*(Xbar-theta)'+V;

meanKW=zeros(p,p);

%Single convergence measures for initial values
DF(1,1)=det(KWSigma);
DF(1,2)=trace(KWSigma);
detsen=eigs(KWSigma);

%Gibbs sampling
for i=2:c;
%KW
muKW=mvnrnd(b,1/n*KWSigma,1);
A=V+n*(Xbar-muKW)'*(Xbar-muKW);
KWStar=wishrnd(thetaKW,mKW);

gsigmaKW=det(KWSigma)^(0.5*(mKW-n*s-p-1))*exp(trace(
-0.5*inv(PhiKW)*KWStar))*(1+trace(thetaKW*KWSigma))*exp(
-0.5*trace((A)*inv(KWSigma)));

gsigmastar=det(KWStar)^(0.5*(mKW-n*s-p-1))*exp(trace(
-0.5*inv(PhiKW)*KWStar))*(1+trace(thetaKW*KWSigma))*exp(
-0.5*trace((A)*inv(KWStar)));

if min(gsigmastar/gsigmaKW,1)>unifrnd(0,1),
KWSigma=KWStar; end;

%Single convergence mesures
DF(i,1)=det(KWSigma);
DF(i,2)=trace(KHWSigma);
detsen=eigs(KWSigma);
DF(i,3)=detsen(p);

%Burn-in period
if i>0.1*c, meanKW=meanKW+KWSigma/c; end;

end;

FN1=norm(Sx-sigma);
FN2=norm(meanKW-sigma);
end;

%Graphs of single convergence measures
%KW;
figure;
plot(DF(:,4),DF(:,1),DF(:,4),DF(:,2),'-o',DF(:,4),DF(:,3),'-x');
legend('Det','Trace','Eig');
```

D.6 Abalone dataset - Section 4.7.2

Matlab code:

```
X1=horzcat(length1,diameter,height,wweight,sweight,vweight);
X=X1(21:40,:);
Roystest(X,0.05); %Test for normality
```

```
[n,p]=size(X);
n0=1;
s=1;
theta0=0.1;
c=1000;
z=1;
aa=1;

for iz=1:z;

Xbar=ones(1,n)*X/n;
V=(X-ones(n,1)*Xbar)'*(X-ones(n,1)*Xbar);
omega=eye(s);
mKW=p;

%Initial values;
SigmaW=V/n;
SigmaIW=V/n;
HWSigma=V/n;
KWSigma=V/n;

b=Xbar;
Sx=V/(n-1);
PhiKW=Sx;
thetaKW=Sx;

Y=n*n0/(n+n0)*(Xbar-theta)*(Xbar-theta)'+V;

meanKW=zeros(p,p);

for i=2:c;

%KW
muKW=mvnrnd(b,1/n*KWSigma,1);
A=V+n*(Xbar-muKW)'*(Xbar-muKW);
KWStar=wishrnd(thetaKW,mKW);
gsigmaKW=det(KWSigma)^(0.5*(mKW-n*s-p-1))*exp(trace(
-0.5*inv(PhiKW)*KWStar))*(1+trace(thetaKW*KWSigma))
*exp(-0.5*trace((A)*inv(KWSigma)));
gsigmastar=det(KWStar)^(0.5*(mKW-n*s-p-1))*exp(trace(
-0.5*inv(PhiKW)*KWStar))*(1+trace(thetaKW*KWSigma))
*exp(-0.5*trace((A)*inv(KWStar)));
if min(gsigmastar/gsigmaKW,1)>0.1*unifrnd(0,1),
KWSigma=KWStar; end;

%Bayes estimates
if i>0.1*c, meanKW=meanKW+KWSigma/c; end;

end;

end;
```

Appendix E

Code for Chapter 5

E.1 Convergence measures and intial values - Section 5.1

Matlab code:

```
n0=1;
s=1;
p=3;
theta0=0.1;
sigma=eye(p);
mu=zeros(1,p);
c=100;
z=100;
aa=1;

c1=0.9;
a=1;
b=2;

n=20;
TRB1=zeros(z,12);
DF=zeros(c,13);

for iz=1:z;
X=mvnrnd(mu,sigma,n);
Xbar=ones(1,n)*X/n;
V=(X-ones(n,1)*Xbar)'*(X-ones(n,1)*Xbar);
omega=eye(s);
theta=ones(1,p)*theta0;
mW=p;
mIW=2*p+1;
mHW=p;
mKW=p;

%Initial values;
SigmaW=1.1*eye(p);%5*eye(p);
SigmaIW=1.1*eye(p);%5*eye(p);
HWSigma=1.1*eye(p);%5*eye(p);
KWSigma=1.1*eye(p);%5*eye(p);

b=Xbar;
Sx=V/(n-1);
PhiIW=4*eye(p);
PhiW=0.25*eye(p);
PhiKW=eye(p);
```

```
thetaHW=0.5*eye(p);
thetaKW=0.5*eye(p);

Y=n*n0/(n+n0)*(Xbar-theta)*(Xbar-theta)'+V;

meanmuIW=zeros(1,p);
meanmuW=zeros(1,p);
meanSIW=zeros(p,p);
meanSW=zeros(p,p);
meanHW=meanSW;
meanKW=meanSW;

%Single convergence measures for initial values
DF(1,1)=det(HWSigma);
DF(1,2)=trace(HWSigma);
detsen=eigs(HWSigma);
DF(1,3)=detsen(p);
DF(1,4)=1;
DF(1,5)=det(KWSigma);
DF(1,6)=trace(KWSigma);
detsen=eigs(KWSigma);
DF(1,7)=detsen(p);
DF(1,8)=det(SigmaW);
DF(1,9)=trace(SigmaW);
detsen=eigs(SigmaW);
DF(1,10)=detsen(p);
DF(1,11)=det(SigmaIW);
DF(1,12)=trace(SigmaIW);
detsen=eigs(SigmaIW);
DF(1,13)=detsen(p);

%Gibbs sampling
for i=2:c;

%Inverse-Wishart;
muIW=mvnrnd(b,SigmaIW,1);
B=n*(muIW-theta)'*(muIW-theta)+PhiIW;
A=V+n*(Xbar-muIW)'*(Xbar-muIW);
SigmaIW=iwishrnd(A,mIW+s*n+s);

meanmuIW=meanmuIW+muIW*(1/c);
meanSIW=meanSIW+SigmaIW*(1/c);

muW=mvnrnd(b,SigmaW,1);
muHW=mvnrnd(b,1/n*HWSigma,1);
D=n0*(muW-theta)*(muW-theta)';
A=V+n*(Xbar-muHW)'*(Xbar-muHW);

%Metropolis-Hastings;
%W
w=0.5;
Sstar=w*iwishrnd(PhiIW,mIW)+(1-w)*wishrnd(PhiW,mW);
fSigmaW=det(SigmaW)^(0.5*(mW-p-1-s-n*s))*exp(trace(
-0.5*SigmaW*inv(PhiW)-0.5*inv(SigmaW)*A));
fSstar=det(Sstar)^(0.5*(mW-p-1-s-n*s))*exp(trace(
-0.5*Sstar*inv(PhiW)-0.5*inv(Sstar)*A));
if min(fSstar/fSigmaW,1)>unifrnd(0,1), SigmaW=Sstar; end;
```

```
%HW
muHW=mvnrnd(b,1/n*HWSigma,1);
A=V+n*(Xbar-muHW)'*(Xbar-muHW);
HWStar=wishrnd(thetaHW,mHW);

gsigmaHW=det(HWSigma)^(-0.5*(mHW-n*s-p-1))*exp(trace(
-0.5*inv(thetaHW)*HWSigma))*hypergeom(a,b,c1*trace(inv(
thetaHW)*HWSigma))*exp(-0.5*trace((A)*inv(HWSigma)));

gsigmastar=det(HWStar)^(-0.5*(mHW-n*s-p-1))*exp(trace(
-0.5*inv(thetaHW)*HWStar))*hypergeom(a,b,c1*trace(inv(
thetaHW)*HWStar))*exp(-0.5*trace((A)*inv(HWStar)));

if min(gsigmastar/gsigmaHW,1)>unifrnd(0,1),
HWSigma=HWStar; end;


%KW
muKW=mvnrnd(b,1/n*KWSigma,1);
A=V+n*(Xbar-muKW)'*(Xbar-muKW);
KWStar=wishrnd(thetaKW,mKW);

gsigmaKW=det(KWSigma)^(0.5*(mKW-n*s-p-1))*exp(trace(
-0.5*inv(PhiKW)*KWStar))*(1+trace(thetaKW*KWSigma))*exp(
-0.5*trace((A)*inv(KWSigma)));

gsigmastar=det(KWStar)^(0.5*(mKW-n*s-p-1))*exp(trace(
-0.5*inv(PhiKW)*KWStar))*(1+trace(thetaKW*KWSigma))*exp(
-0.5*trace((A)*inv(KWStar)));

if min(gsigmastar/gsigmaKW,1)>unifrnd(0,1),
KWSigma=KWStar; end;

%Single convergence mesures
DF(i,1)=det(HWSigma);
DF(i,2)=trace(HWSigma);
detsen=eigs(HWSigma);
DF(i,3)=detsen(p);
DF(i,4)=i;
DF(i,5)=det(KWSigma);
DF(i,6)=trace(KWSigma);
detsen=eigs(KWSigma);
DF(i,7)=detsen(p);
DF(i,8)=det(SigmaW);
DF(i,9)=trace(SigmaW);
detsen=eigs(SigmaW);
DF(i,10)=detsen(p);
DF(i,11)=det(SigmaIW);
DF(i,12)=trace(SigmaIW);
detsen=eigs(SigmaIW);
DF(i,13)=detsen(p);

%Composite convergence measures
block1=eye(c);
for i=1:c;
for j=1:c;
```

```
 if  j>=i ,  block1(i ,j)=1/j;end;
 end;
 end;

DF1=DF'*block1;
DF1=DF1';

%Graphs of single convergence measures
%HW;
 figure;
 plot(DF(:,4),DF(:,1),DF(:,4),DF(:,2),'-o',
DF(:,4),DF(:,3),'-x');
 legend('Det','Trace','Eig');

%KW;
 figure;
 plot(DF(:,4),DF(:,5),DF(:,4),DF(:,6),'-o',
DF(:,4),DF(:,7),'-x');
 legend('Det','Trace','Eig');

%W
 figure;
 plot(DF(:,4),DF(:,8),DF(:,4),DF(:,9),'-o',
DF(:,4),DF(:,10),'-x');
 legend('Det','Trace','Eig');

%IW
 figure;
 plot(DF(:,4),DF(:,11),DF(:,4),DF(:,12),'-o',
DF(:,4),DF(:,13),'-x');
 legend('Det','Trace','Eig');

%Graphs of composite convergence measures
%HW1;
 figure;
 plot(DF(:,4),DF1(:,1),DF(:,4),DF1(:,2),'-o',
DF(:,4),DF1(:,3),'-x');
 legend('Det','Trace','Eig');

%KW1;
 figure;
 plot(DF(:,4),DF1(:,5),DF(:,4),DF1(:,6),'-o',
DF(:,4),DF1(:,7),'-x');
 legend('Det','Trace','Eig');

%W1;
 figure;
 plot(DF(:,4),DF1(:,8),DF(:,4),DF1(:,9),'-o',
DF(:,4),DF1(:,10),'-x');
 legend('Det','Trace','Eig');

%IW1
 figure;
 plot(DF(:,4),DF1(:,11),DF(:,4),DF1(:,12),'-o',
DF(:,4),DF1(:,13),'-x');
 legend('Det','Trace','Eig');
```

E.2 Univariate prior density functions - Section 5.2.1

Matlab code:

```
inc=0.005;
y=0.005:inc:2;
Sprior=0.9;

%IG
betaIG=11.5;
alphaIG=betaIG/Sprior+1;

%HG
a=1;
b=1;
phi=0.01875;
c=phi/max(y);
alpha1=19/2;
beta1=1/phi;

%KG
alpha3=0.8;
beta3=4;

%G
alpha2=3;
beta2=10;

f=y.^(alpha1-1).*exp(-beta1.*y).*hypergeom(a,b,c/phi.*y);
g=y.^(-alpha-1).*exp(-beta./y);
g1=y.^(-alphaIG-1).*exp(-betaIG./y);
h=y.^(alpha2-1).*exp(-beta2.*y);
j=y.^(alpha3-1).*exp(-beta3.*y).*(1.+y);

c1=sum(f.*inc);
c2=sum(g.*inc);
c3=sum(h.*inc);
c4=sum(g1.*inc);
c5=sum(j.*inc);

f=f/c1;
g=g/c2;
h=h/c3;
g1=g1/c4;
j=j/c5;
c1=sum(f.*inc);
c2=sum(g.*inc);
c3=sum(h.*inc);

figure;
plot(y,f,y,j,y,g1,'--',y,h,':',y,g);
legend('Hypergeometric gamma','Kummer gamma','Inverse gamma',
'Gamma');
```

E.3 Univariate simulation study - Section 5.2.1

Matlab code:

```
%Target values
mu=0;
s=1;

%Sample
m=18;

Sprior=0.9;

phi1=2; %IG
alp=phi1/Sprior+1; %IG

alp2=1.8; %G
beta2=2; %G

a=1;%HG
b=2;%HG
phi=0.105; %HG
n=Sprior*2/phi; %HG

%KG parameters
alp1=1.2;
lambda=2.1;
gamma1=1;

inc=0.01;
%n1=200;
%n2=300;

nz=1; %100
inc2=0.2;

alpha=0.05;

est=zeros(nz,5);
int=zeros(nz,6);
x1=-5:inc:5;
ss1=0.005:inc:2;
ss2=0.005:inc2:2;

postm1=x1;
posts1=ss1;
posts1a=ss1;
posts1b=ss1;
posts2=posts1;
posts3=posts1;
posts4=posts1;
posts5=posts1;
posts6=posts1;
posts7=posts1;
posts8=posts1;

countg=0;
countig=0;
```

```
countwg=0;
counthg=0;
width=zeros(length(ss1),4);
postm=zeros(length(x1),length(nz));
posts=zeros(length(ss1),length(nz));
postsg=posts;
postsig=posts;
postswg=posts;
dhx=zeros(length(x1),length(x1))+triu(ones(length(x1),
length(x1)));
dhs=zeros(length(ss1),length(ss1))+triu(ones(length(ss1),
length(ss1)));
dhs1=zeros(length(ss2),length(ss2))+triu(ones(length(ss2),
length(ss2)));

for z=1:nz

x=normrnd(mu,s,[1 m]); %Simulated data;
xbar=sum(x)/m;
ss=(sum((x-xbar).^2))/(m-1);

c=max([max(ss2)/phi max(ss1)/phi]);
c=phi/c-0.0001;

%Posterior mu
for j=1:length(x1)
  sum1=zeros(length(ss2),1);
  for ii=1:length(ss2);
  sum1(ii)=phi.^(m/2-n/2)/gamma(n/2-m/2)*ss2(ii).^(n/2-m/2)
  *hypergeom(a,b,c/phi*ss2(ii))*exp(-0.5/ss2(ii)*
  sum((x-x1(j)).^2) - ss2(ii)/phi);
  end;
aa=sum1'*inc*dhs1;
aa=aa(length(ss2));
   postm1(j)=(2*22/7/m).^(-0.5)*phi.^(-0.5)*gamma(n/2-m/2)
   /gamma(n/2-m/2+0.5)*(aa);
end;

%Posterior sigma
for k=1:length(ss1)

posts1(k)=ss1(k).^(n/2-m/2-0.5)*hypergeom(a,b,c/phi*ss1(k))
*exp(-0.5/ss1(k)*(sum(x.^2)-m*xbar^2)-ss1(k)/phi); %HG
posts2(k)=ss1(k).^(n/2-m/2+0.5)*hypergeom(a,b,c/phi*ss1(k))
*exp(-0.5/ss1(k)*(sum(x.^2)-m*xbar^2)-ss1(k)/phi); %HG est

posts3(k)=ss1(k).^(alp2-m/2-1)*exp(-0.5/ss1(k)*
((sum(x.^2)-m*xbar^2))-beta2*ss1(k)); %G
posts4(k)=ss1(k).^(alp2-m/2)*exp(-0.5/ss1(k)*
((sum(x.^2)-m*xbar^2))-beta2*ss1(k)); G est

posts5(k)=ss1(k).^(-alp-m/2-1)*exp(-0.5/ss1(k)
*((sum(x.^2)-m*xbar^2)+phi1)); %IG
posts6(k)=ss1(k).^(-alp-m/2)*exp(-0.5/ss1(k)
*((sum(x.^2)-m*xbar^2)+phi1)); %IG est

posts7(k)=ss1(k).^(alp1-m/2-1)*exp(-0.5/ss1(k)
```

```
*((sum(x.^2)-m*xbar^2))-lambda*ss1(k))*(1+ss1(k))
^gamma1; %KG
posts8(k)=ss1(k).^(alp1-m/2)*exp(-0.5/ss1(k)
*((sum(x.^2)-m*xbar^2))-lambda*ss1(k))*(1+ss1(k))
^gamma1; %KG est

end;

%Normalising constants
aream=(postm1*inc)*dhx;
areas=(posts1*inc)*dhs;
areas2=(posts2*inc)*dhs;
areas3=(posts3*inc)*dhs;
areas4=(posts4*inc)*dhs;
areas5=(posts5*inc)*dhs;
areas6=(posts6*inc)*dhs;
areas7=(posts7*inc)*dhs;
areas8=(posts8*inc)*dhs;

postm(:,z)=postm1/aream(length(x1));
posts(:,z)=(posts1').//areas(length(ss1));%HG
postsg(:,z)=(posts3')/areas3(length(ss1)); %G
postsig(:,z)=(posts5')/areas5(length(ss1)); %IG
postswg(:,z)=(posts7')/areas7(length(ss1)); %KG

sum2=areas(length(ss1));
sum3=areas2(length(ss1));
sum4=areas3(length(ss1));
sum5=areas4(length(ss1));
sum6=areas5(length(ss1));
sum7=areas6(length(ss1));
sum8=areas7(length(ss1));
sum9=areas8(length(ss1));

aream=(postm(:,z)')*inc*dhx;
areas=(posts(:,z)')*inc*dhs;
areas1=(postsg(:,z)')*inc*dhs;
areas2=(postsig(:,z)')*inc*dhs;
areas3=(postswg(:,z)')*inc*dhs;

telmu=0;
telml=0;
telsl=0;
telsu=0;

%credible intervals
%mu
for i=1:length(x1);
if aream(i)<alpha/2; telml=telml+1; end;
if aream(i)<(1-alpha/2); telmu=telmu+1; end;
end;

%HG
for i=1:length(ss1);
if areas(i)<alpha/2; telsl=telsl+1; end;
if areas(i)<(1-alpha/2); telsu=telsu+1; end;
end;
```

```
if s<ss1(telsu); if s>ss1(telsl); counthg=counthg+1;
end;
end;

telsl1=0;
telsu1=0;
telsl2=0;
telsu2=0;
telsl3=0;
telsu3=0;

%G
for i=1:length(ss1);
if areas1(i)<alpha/2; telsl1=telsl1+1; end;
if areas1(i)<(1-alpha/2); telsu1=telsu1+1; end;
end;
if s<ss1(telsu1); if s>ss1(telsl1); countg=countg+1;
end;
end;

%IG
for i=1:length(ss1);
if areas2(i)<alpha/2; telsl2=telsl2+1; end;
if areas2(i)<(1-alpha/2); telsu2=telsu2+1; end;
end;
if s<ss1(telsu2); if s>ss1(telsl2); countig=countig+1;
end;
end;

%KG
for i=1:length(ss1);
if areas3(i)<alpha/2; telsl3=telsl3+1; end;
if areas3(i)<(1-alpha/2); telsu3=telsu3+1; end;
end;
if s<ss1(telsu3); if s>ss1(telsl3); countwg=countwg+1;
end;
end;

%credible interval widths
width(z,1)=ss1(telsu)-ss1(telsl); %HG
width(z,2)=ss1(telsu1)-ss1(telsl1); %G
width(z,3)=ss1(telsu2)-ss1(telsl2); %IG
width(z,4)=ss1(telsu3)-ss1(telsl3); %KG


%Bayes estimates
mu111=xbar;
est(z,1)=ss; %MLE
est(z,2)=sum3/sum2; %HG
est(z,3)=sum5/sum4; %G
est(z,4)=sum7/sum6; %IG
est(z,5)=sum9/sum8; %KG

end;
end;
%Posterior df of sigma
figure;
```

```
plot(ss1,posts,ss1,postsig,ss1,postsg,ss1,postswg);
xlabel('Population variance');
ylabel('Density');
legend('Hypergeometric gamma','Inverse gamma', 'Gamma',
'Kummer gamma');
title 'Exact posterior density functions';

%Posterior df of mu
figure13=figure('Color',[1 1 1]);
plot(x1,postm);
xlabel('Population mean');
ylabel('Density');

%Coverage probabilities
coveragemu=sum(int(:,5))/nz;
coveragegamma=countg/nz*100;
coveragekg=countwg/nz*100;
coverageigamma=countig/nz*100;
coveragehgamma=counthg/nz*100;

%Median credible interval widths
medianwidthhgamma=median(width(:,1)); %HG
medianwidthgamma=median(width(:,2)); %G
medianwidthigamma=median(width(:,3)); %IG
medianwidthwgamma=median(width(:,4)); %KG

%Bias and MSE
%HG
HGammaE=abs(est(:,2)-s);
MSEH=std(est(:,2))^2+HGammaE.^2;

%G
GammaE=abs(est(:,3)-s);
MSEG=std(est(:,3))^2+GammaE.^2;

%IG
IGammaE=abs(est(:,4)-s);
MSEI=std(est(:,4))^2+IGammaE.^2;

%KG
WGammaE=abs(est(:,5)-s);
MSEW=std(est(:,5))+WGammaE.^2;

figure;
plot(1:1:nz,HGammaE,1:1:nz,GammaE,':',1:1:nz,IGammaE,
'--',1:1:nz,WGammaE,'*-');
xlabel('Number of sample');
ylabel('Bias');
legend 'Hypergeometric gamma estimate' 'Gamma estimate'
 'Inverse gamma estimate' 'Kummer gamma estimate';
title 'Bias calculated from 100 samples';

figure;
plot(1:1:nz,MSEH,1:1:nz,MSEG,':',1:1:nz,MSEI,'--',
1:1:nz,MSEW,'*-');
xlabel('Number of sample');
ylabel('MSE');
```

```
legend 'Hypergeometric gamma estimate' 'Gamma estimate'
 'Inverse gamma estimate' 'Kummer gamma estimate';
title 'MSE calculated from 100 samples';
```

E.4 Forestry dataset - Section 5.2.2

Matlab code:

```
%Real data – trees volume;

x=[10.3
10.3
10.2
16.4
18.8
19.7
15.6
18.2
22.6
19.9
24.2
21
21.4
21.3
19.1
22.2
33.8
27.4
25.7
24.9
34.5
31.7
36.3
38.3
42.6
55.4
55.7
58.3
51.5
51
77];
x=log(x);
m=length(x);

xbar=sum(x)/m;
ss=(sum((x-xbar).^2))/(m-1);

Sprior=0.3;

%Priors
inc=0.005;
y=0.005:inc:2;
Sprior=0.3;

%Likelihood
beta=11.5;
```

```
alpha=beta/ss+1;

%IG
betaIG=11.5;
alphaIG=betaIG/Sprior+1;

%HG
a=1;
b=1;
phi=0.01875;
c=phi/max(y);
alpha1=19/2;
beta1=1/phi;

%KG
alpha3=0.8;
beta3=4;

%G
alpha2=3;
beta2=10;

f=y.^(alpha1-1).*exp(-beta1.*y).*hypergeom(a,b,c/phi.*y);
g=y.^(-alpha-1).*exp(-beta./y);
g1=y.^(-alphaIG-1).*exp(-betaIG./y);
h=y.^(alpha2-1).*exp(-beta2.*y);
j=y.^(alpha3-1).*exp(-beta3.*y).*(1.+y);

c1=sum(f.*inc);
c2=sum(g.*inc);
c3=sum(h.*inc);
c4=sum(g1.*inc);
c5=sum(j.*inc);

f=f/c1;
g=g/c2;
h=h/c3;
g1=g1/c4;
j=j/c5;

figure;
plot(y,f,y,j,y,g1,'--',y,h,':',y,g);
legend('Hypergeometric gamma','Kummer gamma','Inverse
 gamma','Gamma','Likelihood');

%Bayesian analysis
phi1=3.5; %IG
alp=phi1/Sprior+1; %IG

alp2=3; %G
beta2=10; %G

a=1;%HG
b=1;%HG
phi=0.01875; %HG
n=Sprior*2/phi; %HG
```

```
%KG parameters
alp1=0.8;
lambda=4;
gamma1=1;

inc=0.01;
%n1=200;
%n2=300;

nz=1;
inc2=0.2;

alpha=0.05;

est=zeros(nz,5);
int=zeros(nz,6);
x1=-5:inc:5;
ss1=0.005:inc:2;
ss2=0.005:inc2:2;

postm1=x1;
posts1=ss1;
posts2=posts1;
posts3=posts1;
posts4=posts1;
posts5=posts1;
posts6=posts1;
posts7=posts1;
posts8=posts1;

width=zeros(length(ss1),4);
postm=zeros(length(x1),length(nz));
posts=zeros(length(ss1),length(nz));
postsg=posts;
postsig=posts;
postswg=posts;

dhx=zeros(length(x1),length(x1))+triu(ones(length(x1),
length(x1)));
dhs=zeros(length(ss1),length(ss1))+triu(ones(length(ss1),
length(ss1)));
dhs1=zeros(length(ss2),length(ss2))+triu(ones(length(ss2),
length(ss2)));

for z=1:nz
xbar=sum(x)/m;
ss=(sum((x-xbar).^2))/(m-1);

c=max([max(ss2)/phi max(ss1)/phi]);
c=phi/c-0.0001;

%Posterior mu
for j=1:length(x1)
sum1=zeros(length(ss2),1);
for ii=1:length(ss2);
sum1(ii)=phi.^(m/2-n/2)/gamma(n/2-m/2)*ss2(ii).^(n/2-m/2)
*hypergeom(a,b,c/phi*ss2(ii))*exp(-0.5/ss2(ii)*
```

```
sum((x-x1(j)).^2) - ss2(ii)/phi);
end;
aa=sum1'*inc*dhs1;
aa=aa(length(ss2));
postm1(j)=(2*22/7/m).^(-0.5)*phi.^(-0.5)*gamma(n/2-m/2)
/gamma(n/2-m/2+0.5)*(aa);
end;

%Posterior sigma
for k=1:length(ss1)

posts1(k)=ss1(k).^(n/2-m/2-0.5)*hypergeom(a,b,c/phi*ss1(k))
*exp(-0.5/ss1(k)*(sum(x.^2)-m*xbar^2)-ss1(k)/phi); %HG
posts2(k)=ss1(k).^(n/2-m/2+0.5)*hypergeom(a,b,c/phi*ss1(k))
*exp(-0.5/ss1(k)*(sum(x.^2)-m*xbar^2)-ss1(k)/phi); %HG est

posts3(k)=ss1(k).^(alp2-m/2-1)*exp(-0.5/ss1(k)*
((sum(x.^2)-m*xbar^2))-beta2*ss1(k)); %G
posts4(k)=ss1(k).^(alp2-m/2)*exp(-0.5/ss1(k)*
((sum(x.^2)-m*xbar^2))-beta2*ss1(k)); G est

posts5(k)=ss1(k).^(-alp-m/2-1)*exp(-0.5/ss1(k)
*((sum(x.^2)-m*xbar^2)+phi1)); %IG
posts6(k)=ss1(k).^(-alp-m/2)*exp(-0.5/ss1(k)
*((sum(x.^2)-m*xbar^2)+phi1)); %IG est

posts7(k)=ss1(k).^(alp1-m/2-1)*exp(-0.5/ss1(k)
*((sum(x.^2)-m*xbar^2))-lambda*ss1(k))*(1+ss1(k))
^gamma1; %KG
posts8(k)=ss1(k).^(alp1-m/2)*exp(-0.5/ss1(k)
*((sum(x.^2)-m*xbar^2))-lambda*ss1(k))*(1+ss1(k))
^gamma1; %KG est

end;

%Normalising constants
aream=(postm1*inc)*dhx;
areas=(posts1*inc)*dhs;
areas2=(posts2*inc)*dhs;
areas3=(posts3*inc)*dhs;
areas4=(posts4*inc)*dhs;
areas5=(posts5*inc)*dhs;
areas6=(posts6*inc)*dhs;
areas7=(posts7*inc)*dhs;
areas8=(posts8*inc)*dhs;

postm(:,z)=postm1/aream(length(x1));
posts(:,z)=(posts1').*/areas(length(ss1));%HG
postsg(:,z)=(posts3')/areas3(length(ss1)); %G
postsig(:,z)=(posts5')/areas5(length(ss1)); %IG
postswg(:,z)=(posts7')/areas7(length(ss1)); %KG

aream=(postm(:,z)')*inc*dhx;
areas=(posts(:,z)')*inc*dhs;
areas1=(postsg(:,z)')*inc*dhs;
areas2=(postsig(:,z)')*inc*dhs;
areas3=(postswg(:,z)')*inc*dhs;
```

```
telmu=0;
telml=0;
telsl=0;
telsu=0;

%credible intervals
%mu
for i=1:length(x1);
if aream(i)<alpha/2; telml=telml+1; end;
if aream(i)<(1-alpha/2); telmu=telmu+1; end;
end;

%HG
for i=1:length(ss1);
if areas(i)<alpha/2; telsl=telsl+1; end;
if areas(i)<(1-alpha/2); telsu=telsu+1; end;
end;

%G
for i=1:length(ss1);
if areas1(i)<alpha/2; telsl1=telsl1+1; end;
if areas1(i)<(1-alpha/2); telsu1=telsu1+1; end;
end;

%IG
for i=1:length(ss1);
if areas2(i)<alpha/2; telsl2=telsl2+1; end;
if areas2(i)<(1-alpha/2); telsu2=telsu2+1; end;
end;

%KG
for i=1:length(ss1);
if areas3(i)<alpha/2; telsl3=telsl3+1; end;
if areas3(i)<(1-alpha/2); telsu3=telsu3+1; end;
end;

%credible interval widths
width(z,1)=ss1(telsu)-ss1(telsl); %HG
width(z,2)=ss1(telsu1)-ss1(telsl1); %G
width(z,3)=ss1(telsu2)-ss1(telsl2); %IG
width(z,4)=ss1(telsu3)-ss1(telsl3); %KG


%Bayes estimates
mu111=xbar;
est(z,1)=ss; %MLE
est(z,2)=sum3/sum2; %HG
est(z,3)=sum5/sum4; %G
est(z,4)=sum7/sum6; %IG
est(z,5)=sum9/sum8; %KG

end;
end;

%Posterior df of sigma
figure;
```

```
plot(ss1,posts,ss1,postsig,ss1,postsg,ss1,postswg);
xlabel('Population variance');
ylabel('Density');
legend('Hypergeometric gamma','Inverse gamma', 'Gamma',
'Kummer gamma');
title 'Exact posterior density functions';

%Posterior df of mu
figure13=figure('Color',[1 1 1]);
plot(x1,postm);
xlabel('Population mean');
ylabel('Density');

%Median credible interval widths
medianwidthhgamma=median(width(:,1)); %HG
medianwidthgamma=median(width(:,2)); %G
medianwidthigamma=median(width(:,3)); %IG
medianwidthwgamma=median(width(:,4)); %KG
```

E.5 Multivariate simulation study - Section 5.3.1

Matlab code:

```
n0=1;
s=1;
p=3;
sigma=eye(p);
mu=zeros(1,p);
c=100;
z=100;
aa=1;

c1=0.9;
a=1;
b=2;

n=20;
TRB1=zeros(z,12);
DF=zeros(c,13);

for iz=1:z;

X=mvnrnd(mu,sigma,n);
Xbar=ones(1,n)*X/n;
V=(X-ones(n,1)*Xbar)'*(X-ones(n,1)*Xbar);
omega=eye(s);
theta=ones(1,p)*theta0;
mW=p;
mIW=2*p+1;
mHW=p;
mKW=p;

%Initial values;
SigmaW=1.1*eye(p);
SigmaIW=1.1*eye(p);
HWSigma=V/n;
KWSigma=V/n;

b=Xbar;
Sx=V/(n-1);
PhiIW=4*eye(p);
PhiW=0.25*eye(p);
PhiKW=eye(p);
thetaHW=0.5*eye(p);
thetaKW=0.5*eye(p);

Y=n*n0/(n+n0)*(Xbar-theta)*(Xbar-theta)'+V;

meanmuIW=zeros(1,p);
meanmuW=zeros(1,p);
meanSIW=zeros(p,p);
meanSW=zeros(p,p);
meanHW=meanSW;
meanKW=meanSW;

%Single convergence measures for initial values
DF(1,1)=det(HWSigma);
```

```
DF(1,2)=trace(HWSigma);
detsen=eigs(HWSigma);
DF(1,3)=detsen(p);
DF(1,4)=1;
DF(1,5)=det(KWSigma);
DF(1,6)=trace(KWSigma);
detsen=eigs(KWSigma);
DF(1,7)=detsen(p);
DF(1,8)=det(SigmaW);
DF(1,9)=trace(SigmaW);
detsen=eigs(SigmaW);
DF(1,10)=detsen(p);
DF(1,11)=det(SigmaIW);
DF(1,12)=trace(SigmaIW);
detsen=eigs(SigmaIW);
DF(1,13)=detsen(p);

%Gibbs sampling
for i=2:c;

%Inverse-Wishart;
muIW=mvnrnd(b,SigmaIW,1);
B=n*(muIW-theta)'*(muIW-theta)+PhiIW;
A=V+n*(Xbar-muIW)'*(Xbar-muIW);
SigmaIW=iwishrnd(A,mIW+s*n+s);%B+A

meanmuIW=meanmuIW+muIW*(1/c);
meanSIW=meanSIW+SigmaIW*(1/c);

%W
muW=mvnrnd(b,SigmaW,1);
muHW=mvnrnd(b,1/n*HWSigma,1);
D=n0*(muW-theta)*(muW-theta)';
A=V+n*(Xbar-muHW)'*(Xbar-muHW);
w=0.5;
Sstar=w*iwishrnd(PhiIW,mIW)+(1-w)*wishrnd(PhiW,mW);
fSigmaW=det(SigmaW)^(0.5*(mW-p-1-s-n*s))*exp(trace(
-0.5*SigmaW*inv(PhiW)-0.5*inv(SigmaW)*A));
fSstar=det(Sstar)^(0.5*(mW-p-1-s-n*s))*exp(trace(
-0.5*Sstar*inv(PhiW)-0.5*inv(Sstar)*A));
if min(fSstar/fSigmaW,1)>unifrnd(0,1), SigmaW=Sstar; end;

%HW
muHW=mvnrnd(b,1/n*HWSigma,1);
A=V+n*(Xbar-muHW)'*(Xbar-muHW);
HWStar=wishrnd(thetaHW,mHW);

gsigmaHW=det(HWSigma)^(-0.5*(mHW-n*s-p-1))*exp(trace(
-0.5*inv(thetaHW)*HWSigma))*hypergeom(a,b,c1*trace(inv(
thetaHW)*HWSigma));%*exp(-0.5*trace((A)*inv(HWSigma)));

gsigmastar=det(HWStar)^(-0.5*(mHW-n*s-p-1))*exp(trace(
-0.5*inv(thetaHW)*HWStar))*hypergeom(a,b,c1*trace(inv(
thetaHW)*HWStar));%*exp(-0.5*trace((A)*inv(HWStar)));

if min(gsigmastar/gsigmaHW,1)>unifrnd(0,1),
 HWSigma=HWStar; end;
```

```
%KW
muKW=mvnrnd(b,1/n*KWSigma,1);
A=V+n*(Xbar-muKW)'*(Xbar-muKW);
KWStar=wishrnd(thetaKW,mKW);

gsigmaKW=det(KWSigma)^(0.5*(mKW-n*s-p-1))*exp(trace(
-0.5*inv(PhiKW)*KWStar))*(1+trace(thetaKW*KWSigma))*exp(
-0.5*trace((A)*inv(KWSigma)));

gsigmastar=det(KWStar)^(0.5*(mKW-n*s-p-1))*exp(trace(
-0.5*inv(PhiKW)*KWStar))*(1+trace(thetaKW*KWSigma))*exp(
-0.5*trace((A)*inv(KWStar)));

if min(gsigmastar/gsigmaKW,1)>unifrnd(0,1),
 KWSigma=KWStar; end;

%Single convergence mesures
DF(i,1)=det(HWSigma);
DF(i,2)=trace(HWSigma);
detsen=eigs(HWSigma);
DF(i,3)=detsen(p);
DF(i,4)=i;
DF(i,5)=det(KWSigma);
DF(i,6)=trace(KWSigma);
detsen=eigs(KWSigma);
DF(i,7)=detsen(p);
DF(i,8)=det(SigmaW);
DF(i,9)=trace(SigmaW);
detsen=eigs(SigmaW);
DF(i,10)=detsen(p);
DF(i,11)=det(SigmaIW);
DF(i,12)=trace(SigmaIW);
detsen=eigs(SigmaIW);
DF(i,13)=detsen(p);

%Burn-in period
if i>0.1*c,meanmuW=meanmuW+muW/c; end;
if i>0.1*c, meanSW=meanSW+SigmaW/c; end;
if i>0.1*c, meanHW=meanHW+HWSigma/c; end;
if i>0.1*c, meanKW=meanKW+KWSigma/c; end;

end;

%Frobenius norms
TRB1(iz,1)=((norm(Xbar-mu))^aa)^(1/aa);
TRB1(iz,2)=((norm(meanmuIW-mu))^aa)^(1/aa);
TRB1(iz,3)=((norm(meanmuW-mu))^aa)^(1/aa);
TRB1(iz,4)=((norm(mapmuIW-mu))^aa)^(1/aa);
TRB1(iz,5)=((norm(mapmuW-mu))^aa)^(1/aa);
TRB1(iz,6)=((norm(Sx-sigma))^aa)^(1/aa);
TRB1(iz,7)=((norm(meanSIW-sigma))^aa)^(1/aa);
TRB1(iz,8)=((norm(meanSW-sigma))^aa)^(1/aa);
TRB1(iz,9)=((norm(mapSIW-sigma))^aa)^(1/aa);
TRB1(iz,10)=((norm(mapSW-sigma))^aa)^(1/aa);
TRB1(iz,11)=((norm(meanHW-sigma))^aa)^(1/aa);
```

```
TRB1(iz,12)=((norm(meanKW-sigma))^aa)^(1/aa);
end;

%Graphs of single convergence measures
%HW;
figure;
plot(DF(:,4),DF(:,1),DF(:,4),DF(:,2),'-o',DF(:,4),
DF(:,3),'-x');
legend('Det','Trace','Eig');

%KW;
figure;
plot(DF(:,4),DF(:,5),DF(:,4),DF(:,6),'-o',DF(:,4),
DF(:,7),'-x');
legend('Det','Trace','Eig');

%W
figure;
plot(DF(:,4),DF(:,8),DF(:,4),DF(:,9),'-o',DF(:,4),
DF(:,10),'-x');
legend('Det','Trace','Eig');

%IW
figure;
plot(DF(:,4),DF(:,11),DF(:,4),DF(:,12),'-o',DF(:,4),
DF(:,13),'-x');
legend('Det','Trace','Eig');

%ECDF
[y1,x1]=ecdf(TRB1(1:z,6)); %MLE
[y2,x2]=ecdf(TRB1(1:z,7)); % IW
[y3,x3]=ecdf(TRB1(1:z,8)); %W
[y4,x4]=ecdf(TRB1(1:z,11)); %HW
[y5,x5]=ecdf(TRB1(1:z,12)); %KW

figure('Color',[1 1 1]);
plot(x1,y1,'-*',x2,y2,'--',x3,y3,':',x4,y4,'-.',x5,y5);
xlabel('Frobenius norm');
ylabel('Empirical CDF');
title('ECDF of the Frobenius norm for the four PM
estimators of sigma');
legend('Sample estimate','Inverse-Wishart estimate',
'Wishart estimate', 'Hypergeometric Wishart estimate',
'Kummer Wishart prior','Location','SouthEast');

%Kolmogorov-Smirnov test
[h7,p7,k7]=kstest2(TRB1(1:z,6),TRB1(1:z,7));
[h8,p8,k8]=kstest2(TRB1(1:z,6),TRB1(1:z,8));
[h9,p9,k9]=kstest2(TRB1(1:z,8),TRB1(1:z,7));
[h11,p11,k11]=kstest2(TRB1(1:z,8),TRB1(1:z,11));
```

E.6 Abalone dataset - Section 5.3.2

Matlab code:

```
X1=horzcat(length1,diameter,height,wweight,sweight,vweight);
X=X1(21:40,:);
Roystest(X,0.05); %Test for normality

[n,p]=size(X);
n0=1;
s=1;
theta0=0.1;
c=1000;
z=1;
aa=1;

c1=0.9;
a=1;
b=2;

for iz=1:z;

Xbar=ones(1,n)*X/n;
V=(X-ones(n,1)*Xbar)'*(X-ones(n,1)*Xbar);
omega=eye(s);
mW=p;
mIW=2*p+1;
mHW=p;
mKW=p;

%Initial values;
SigmaW=V/n;
SigmaIW=V/n;
HWSigma=V/n;
KWSigma=V/n;

b=Xbar;
Sx=V/(n-1);
PhiIW=4*eye(p);
PhiW=0.25*eye(p);
PhiKW=eye(p);
thetaHW=0.5*eye(p);
thetaKW=0.5*eye(p);

Y=n*n0/(n+n0)*(Xbar-theta)*(Xbar-theta)'+V;

meanmuIW=zeros(1,p);
meanmuW=zeros(1,p);
meanSIW=zeros(p,p);
meanSW=zeros(p,p);
meanHW=meanSW;
meanKW=meanSW;

for i=2:c;

%Inverse-Wishart;
muIW=mvnrnd(b,SigmaIW,1);
```

```
B=n*(muIW-theta)'*(muIW-theta)+PhiIW;
A=V+n*(Xbar-muIW)'*(Xbar-muIW);
SigmaIW=iwishrnd(B+A,3*mIW+s*n+s);%B+A

meanmuIW=meanmuIW+muIW*(1/c);
meanSIW=meanSIW+SigmaIW*(1/c);

%Wishart;
muW=mvnrnd(b,SigmaW,1);
muHW=mvnrnd(b,1/n*HWSigma,1);
D=n0*(muW-theta)*(muW-theta)';
A=V+n*(Xbar-muHW)'*(Xbar-muHW);
B=n*(muW-theta)'*(muW-theta)+PhiW;

%Metropolis-Hastings;
%W
w=0.5;
Sstar=w*iwishrnd(PhiIW,mIW)+(1-w)*wishrnd(PhiW,mW);
fSigmaW=det(SigmaW)^(0.5*(mW-p-1-s-n*s))*exp(trace(
-0.5*SigmaW*inv(PhiW)-0.5*inv(SigmaW)*(B+A)));
fSstar=det(Sstar)^(0.5*(mW-p-1-s-n*s))*exp(trace(
-0.5*Sstar*inv(PhiW)-0.5*inv(Sstar)*(B+A)));
if min(fSstar/fSigmaW,1)>unifrnd(0,1), SigmaW=Sstar; end;

%HW
muHW=mvnrnd(b,1/n*HWSigma,1);
A=V+n*(Xbar-muHW)'*(Xbar-muHW);
HWStar=wishrnd(thetaHW,mHW);
gsigmaHW=det(HWSigma)^(-0.5*(mHW-n*s-p-1))*exp(trace(
-0.5*inv(thetaHW)*HWSigma))*hypergeom(a,b,c1*trace(
inv(thetaHW)*HWSigma))*exp(-0.5*trace((A)*inv(HWSigma)));
gsigmastar=det(HWStar)^(-0.5*(mHW-n*s-p-1))*exp(trace(
-0.5*inv(thetaHW)*HWStar))*hypergeom(a,b,c1*trace(inv(
thetaHW)*HWStar))*exp(-0.5*trace((A)*inv(HWStar)));
if min(gsigmastar/gsigmaHW,1)>0.1*unifrnd(0,1),
 HWSigma=HWStar; end;

%KW
muKW=mvnrnd(b,1/n*KWSigma,1);
A=V+n*(Xbar-muKW)'*(Xbar-muKW);
KWStar=wishrnd(thetaKW,mKW);
gsigmaKW=det(KWSigma)^(0.5*(mKW-n*s-p-1))*exp(trace(
-0.5*inv(PhiKW)*KWStar))*(1+trace(thetaKW*KWSigma))
*exp(-0.5*trace((A)*inv(KWSigma)));
gsigmastar=det(KWStar)^(0.5*(mKW-n*s-p-1))*exp(trace(
-0.5*inv(PhiKW)*KWStar))*(1+trace(thetaKW*KWSigma))
*exp(-0.5*trace((A)*inv(KWStar)));
if min(gsigmastar/gsigmaKW,1)>0.1*unifrnd(0,1),
 KWSigma=KWStar; end;

%Bayes estimates
if i>0.1*c,meanmuW=meanmuW+muW/c; end;
if i>0.1*c, meanSW=meanSW+SigmaW/c; end;
if i>0.1*c, meanHW=meanHW+HWSigma/c; end;
if i>0.1*c, meanKW=meanKW+KWSigma/c; end;

end;
```

```
end;

%Simulate samples based on estimates;
n=20;
XHW=mvnrnd(Xbar,meanHW,n);
XKW=mvnrnd(Xbar,meanKW,n);
XW=mvnrnd(Xbar,meanSW,n);
XIW=mvnrnd(Xbar,meanSIW,n);
XS=mvnrnd(Xbar,Sx,n);

%Sort the matrices
XS1=sortrows(XS,4);
XHW1=sortrows(XHW,4);
XKW1=sortrows(XKW,4);
XW1=sortrows(XW,4);
XIW1=sortrows(XIW,4);
X2=sortrows(X,4);

%Images of simulated samples

 figure;
 imagesc(X2); %Real dataset
 colormap;

 figure;
 imagesc(XS1); %MLE
 colormap;

 figure;
 imagesc(XHW1); %HW
 colormap;

 figure;
 imagesc(XKW1); %KW
 colormap;

 figure;
 imagesc(XW1); %W
 colormap;

 figure;
 imagesc(XIW1); %IW
 colormap;

%Mahalanobis distance
maHW=mahal(XHW,X); %HW and real data
maKW=mahal(XKW,X); %KW and real data
maW=mahal(XW,X); %W and real data
maIW=mahal(XIW,X); %IW and real data
maS=mahal(XS,X); %MLE and real data
mahalav=[mean(maHW) mean(maKW) mean(maW) mean(maIW) mean(maS)];
```

List of Figures

List of Tables

Bibliography

Adhikari, S. (2010). "Generalized Wishart distribution for probabilistic structural dynamics". In: *Computational Mechanics* 45.5, 495–511.

Anderson, T.W. (2003). *An Introduction to Multivariate Statistical Analysis*. Wiley Series in Probability and Statistics. Wiley, 2003. ISBN: 9780471360919. https://books.google.co.za/books?id=Cmm9QgAACAAJ.

Arashi, M., Bekker, A., and Van Niekerk, J. (2017). "Weighted-type Wishart distributions with application". In: *Revstat* 15(2), 205–222.

Arashi, M., Iranmanesh, A., and Salarzadeh Jenatabadi, H. (2013). "Bayesian analysis in multivariate regression models with conjugate priors". In: *Statistics* Accepted.

Arashi, M., Saleh, A.K.Md.E., and Tabatabaey, S.M.M. (2013). "Regression model with elliptically contoured errors". In: *Statistics* 47.6, 1266–1284.

Bekker, A. and Roux, J.J.J. (1995). "Bayesian multivariate normal analysis with a Wishart prior". In: *Communications in Statistics - Theory and Methods* 24.10, 2485–2497.

Bekker, A., Van Niekerk, J., and Arashi, M. (2017). "Wishart distributions - Advances in Theory with Bayesian application". In: *Journal of Multivariate Analysis* 155, 272–283.

Berger, J.O. (2013). *Statistical decision theory and Bayesian analysis*. Springer Science & Business Media, 2013.

Bhattacharya, S.K. and Saxena, A.K. (1989). "Multivariate modified Bessel distribution". In: *Estadistica* 41.136.

Bolfarine, H. and Arellano-Valle, R.B. (2005). "Elliptical Measurement Error Models–A Bayesian Approach". In: *Handbook of Statistics* 25, 669–688.

Bryc, W. (2008). "Compound real Wishart and q-Wishart matrices". In: *International Mathematics Research Notices* 2008, rnn079.

Caro-Lopera, F.J., Gonzàlez-Farìas, G., and Balakrishnan, N. (2014). "On Generalized Wishart Distributions-I: Likelihood Ratio Test for Homogeneity of Covariance Matrices". In: *Sankhyā: The Indian Journal of Statistics, Series A* 76.2, 179–194.

Casella, G. and George, E.I. (1992). "Explaining the Gibbs sampler". In: *The American Statistician* 46.3, 167–174.

Chakrabarty, D., Biswas, M., and Bhattacharya, S. (2015). "Bayesian nonparametric estimation of Milky Way parameters using matrix-variate data, in a new Gaussian Process based method". In: *Electronic Journal of Statistics* 9.1, 1378–1403.

Chu, K. C. (1973). "Estimation and decision for linear systems with elliptical random processes". In: *IEEE Transactions on Automatic Control* Ac18, 499–505.

Das, S. and Dey, D. K. (2010). "On Bayesian inference for multivariate generalized gamma distribution". In: *Statistics and Probability letters* 80.19-20, 1492–1499.

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

Davis, A.W. (1980). "Invariant polynomials with two matrix arguments, extending the zonal polynomials". In: *Multivariate Analysis-V*, 287–299.

Dawid, A.P. and Lauritzen, S.L. (1993). “Hyper Markov laws in the statistical analysis of decomposable graphical models”. In: *The Annals of Statistics*, 1272–1317.

DeGroot, M.H. (1970). *Optimal statistical decisions*. United States of America: McGraw-Hill Inc., 1970.

Dìaz-Garcìa, J.A. (2009). “Special functions: Integral properties of Jack polynomials, hypergeometric functions and Invariant polynomials”. In: *arXiv preprint ArXiv:0909.1988*.

Dìaz-Garcìa, J.A. and Jàimez, R.G. (2006). “Wishart and Pseudo-Wishart distributions under elliptical laws and related distributions in the shape theory context”. In: *Journal of Statistical Planning and Inference* 136.12, 4176–4193.

Dìaz-Garcìa, J.A., Gutièrrez-Jàimez, R., et al. (2011). “On Wishart distribution: some extensions”. In: *Linear Algebra and its Applications* 435.6, 1296–1310.

Dirac, P. (1958). *The priciples of Quantum Mechanics*. Vol. 4. Clarendon Press, Oxford, 1958.

Duran, B.S. and Booker, J.M. (1988). “A Bayes sensitivity analysis when using the beta distribution as a prior”. In: *IEEE Transactions on Reliability* 37.2, 239–247.

Erdelyi, A., Magnus, W., Oberhettinger, F., and Tricomi, F.G. (1953). *Higher transcendental functions I*. United States of America: McGraw-Hill Inc., 1953.

Fang, K.T. and Li, R.Z. (1999). “Bayesian statistical inference on elliptical matrix distributions”. In: *Journal of Multivariate Analysis* 70, 66–85.

Fisher, R.A. (1915). “Frequency distribution of the values of the correlation coefficient in samples from an indefinitely large population”. In: *Biometrika*, 507–521.

Funke, D. (2014). “Length-biased distributions and their properties”. Dissertation. University of Pretoria, 2014.

Furlotte, N.A. and Eskin, E. (2015). “Efficient multiple-trait association and estimation of genetic correlation using the matrix-variate linear mixed model”. In: *Genetics* 200.1, 59–68.

Gelman, A., Carlin, J.B., Stern, H.S., and Rubin, D.B. (1995). *Bayesian Data Analysis*. London: Chapman / Hall, 1995.

Gelman, A. and Rubin, D.B. (1992). “Inference from iterative simulation using multiple sequences”. In: *Statistical science*, 457–472.

Golub, G.H. and Loan, C.F. Van (1996). “Matrix computations”. In: *Johns Hopkins University, Press, Baltimore, MD, USA*, 374–426.

Gradshteyn, I.S. and Ryzhik, I.M. (2007). *Table of integrals, series and products*. United States of America: Academic Press, 2007.

Grafen, A. and Hails, R. (2002). *Modern statistics for the life sciences*. Vol. 123. Oxford University Press Oxford, 2002.

Griffin, J.E. and Brown, P.J. (2010). “Inference with normal-gamma prior distributions in regression problems”. In: *Bayesian Anal.* 5.1, 171–188. doi:10.1214/10-BA507. http://dx.doi.org/10.1214/10-BA507.

Gross, K.I. and Richards, D.P. (1987). “Special functions of matrix argument. I. Algebraic induction, zonal polynomials, and hypergeometric functions”. In: *Transactions of the American Mathematical Society* 301.2, 781–811.

Gupta, A.K. and Nagar, D.K. (2000). *Matrix variate distributions*. United States of America: Chapman / Hall, 2000.

Gupta, A.K., Varga, T., and Bodnar, T. (2013). *Elliptically contoured models in statistics and portfolio theory*. Springer, 2013.

Herz, C.Z. (1955). “Bessel functions of matrix argument”. In: *Annals of Mathematics* 61, 474–523.

Hung, H. and Wang, C. (2013). “Matrix variate logistic regression model with application to EEG data”. In: *Biostatistics* 14.1, 189–202.

Jain, K. and Nanda, A.K. (1995). “On multivariate weighted distributions”. In: *Communications in Statistics - Theory and Methods* 24.10, 2517–2539.

James, A.T. (1964). “Distributions of matrix variates and latent roots derived from normal samples”. In: *The Annals of Mathematical Statistics*, 475–501.

Lesaffre, E. and Lawson, A.B. (2012). *Bayesian biostatistics*. John Wiley & Sons, 2012.

Lichman, M. (2013). *(UCI) Machine Learning Repository*. 2013. http://archive.ics.uci.edu/ml.

Mahalanobis, P.C. (1936). “On the generalized distance in statistics”. In: *Proceedings of the National Institute of Sciences (Calcutta)* 2, 49–55.

Mathai, A.M. (1993). *A handbook of generalized special functions for statistical and physical sciences*. Oxford University Press, USA, 1993.

McElreath, R. (2016). *Statistical Rethinking: A Bayesian course with examples in R and Stan*. Vol. 122. CRC Press, 2016.

Metropolis, N., Rosenbluth, A., Rosenbluth, M., Teller, A., and Teller, E. (1953). “Simulated annealing”. In: *Journal of Chemical Physics* 21, 1087–1092.

Miller Jr, R.G. (1980). “Kanamycin levels in premature babies”. In: *Biostatistics Casebook* 3, 127–142.

Muirhead, R.J. (1982). *Aspects of Multivariate Statistical theory*. Hoboken, New Jersey: John Wiley / Sons, Inc., 1982.

Munilla, S. and Cantet, R.J.C. (2012). “Bayesian conjugate analysis using a generalized inverted Wishart distribution accounts for differential uncertainty among the genetic parameters–an application to the maternal animal model”. In: *Journal of Animal Breeding and Genetics* 129.3, 173–187.

Murata, N. (2016). “Matrix Factorization for Image Processing”. In: *Applied Matrix and Tensor Variate Data Analysis*. Springer, 2016, 73–92.

Nanda, A.K. and Jain, K. (1999). “Some weighted distribution results on univariate and bivariate cases”. In: *Journal of Statistical Planning and Inference* 77.2, 169–180.

Navarro, J., Ruiz, J.M., and Aguila, Y. Del (2006). “Multivariate weighted distributions: a review and some extensions”. In: *Statistics* 40.1, 51–64.

Patil, G.P. and Ord, J.K. (1976). “On size-biased sampling and related form-invariant weighted distributions”. In: *Sankhyā: The Indian Journal of Statistics, Series B*, 48–61.

Pauw, J., Bekker, A., and Roux, J.J.J. (2010). “Densities of composite weibullized generalized gamma variables”. In: *South African Statistical Journal* 44.1, 17–42.

Press, S.J. (1982). “Applied multivariate analysis: Including Bayesian and frequentist methods of inference”. In: *Melbourne, Flordia: Krieger Publishing Company*.

Provost, S.B. and Cheong, Y. (2002). “The distribution of Hermitian quadratic forms in elliptically contoured random vectors”. In: *Journal of Statistical Planning and Inference* 102.2, 303–316.

Raìffa, H. and Schlaifer, R. (1961). *Applied statistical decision theory*. Studies in managerial economics. Division of Research, Graduate School of Business Adminitration, Harvard University, 1961. ISBN: 9780875840178. https://books.google.co.za/books?id=wPBLAAAAMAAJ.

Rao, B.L.S.P. (1990). “Remarks on univariate elliptical distributions”. In: *Statistics & Probability Letters* 10.4, 307–315.

Rao, C.R. (1965). “On discrete distributions arising out of methods of ascertainment”. In: *Sankhyā: The Indian Journal of Statistics, Series A*, 311–324.

Roux, J.J.J. (1971). “On generalized multivariate distributions”. In: *South African Statistical Journal* 5, 91–100.

Roverato, A. (2002). “Hyper Inverse Wishart Distribution for Non-decomposable Graphs and its Application to Bayesian Inference for Gaussian Graphical Models”. In: *Scandinavian Journal of Statistics* 29.3, 391–411.

Royston, J.P. (1983). “Some techniques for assessing multivarate normality based on the shapiro-wilk w”. In: *Applied Statistics*, 121–133.

Rue, H., Martino, S., and Chopin, N. (2009). “Approximate Bayesian inference for latent Gaussian models by using integrated nested Laplace approximations”. In: *Journal of the Royal Statistical Society: Series B* 71.2, 319–392.

Teng, C.H., Fang, H., and Deng, W. (1989). “The generalized noncentral Wishart distribution”. In: *Journal of Mathematical Research and Exposition*. Vol. 9. 4. 1989, 479–488.

Tourneret, J., Ferrari, A., and Letac, G. (2005). “The noncentral Wishart distribution: Properties and application to speckle imaging”. In: *Statistical Signal Processing, 2005 IEEE/SP 13th Workshop on*. IEEE. 2005, 924–929.

Van der Merwe, G.J. and Roux, J.J.J. (1974). “On generalized matrix-variate hypergeometric distribution”. In: *South African Statistical Journal* 8, 49–58.

Van Niekerk, J. (2012). “Subjective Bayesian analysis of the elliptical model”. Dissertation. University of Pretoria, 2012.

Van Niekerk, J., Bekker, A., and Arashi, M. (2016). “A gamma-mixture class of distributions with Bayesian application”. In: *Communications in Statistics - Simulation and Computation* (Accepted).

Van Niekerk, J., Bekker, A., Arashi, M., and De Waal, D.J. (2016). “Estimation under the matrix variate elliptical model”. In: *South African Statistical Journal* 50.1, 149–171.

Van Niekerk, J., Bekker, A., Arashi, M., and Roux, J.J.J. (2015). “Subjective Bayesian analysis of the elliptical model”. In: *Communications in Statistics - Theory and Methods* 44.17, 3738–3753.

Wang, H. and West, M. (2009). “Bayesian analysis of matrix normal graphical models”. In: *Biometrika* 96.4, 821–834.

Wishart, J. (1928). “The generalised product moment distribution in samples from a normal multivariate population”. In: *Biometrika*, 32–52.