Title	Functionals of a Wishart matrix and a normal vector and its application to linear discriminant analysis
Author(s)	米永, 航志朗
Degree Grantor	北海道大学
Degree Name	博士(経済学)
Dissertation Number	甲第14922号
Issue Date	2022-03-24
DOI	https://doi.org/10.14943/doctoral.k14922
Doc URL	https://hdl.handle.net/2115/85619
Type	doctoral thesis
File Information	Koshiro_Yonenaga.pdf

Functionals of a Wishart matrix and a normal vector and its application to linear discriminant analysis

Koshiro Yonenaga

Graduate School of Economics and Business

Hokkaido University

A thesis submitted for the degree of Doctor of Philosophy
March 2022

Abstract

In this dissertation, we investigate some functionals of a Wishart matrix and a normal vector and discuss the application to linear discriminant analysis in a Bayesian framework.

In section 2, we consider the distribution of the product of a Wishart matrix and a normal vector which are independently distributed. We derive the stochastic representation of the product which is used to derive the density function and higher order moments of the product. Based on the higher order moments of the product, we further present an Edgeworth type expansion for the product. In addition, it turns out that the obtained stochastic representation, density function and moments of the product remain valid for the product of a singular Wishart matrix and a normal vector.

In section 3, we consider the distribution of the product of a Wishart matrix and a conditional normal vector given a Wishart matrix. This type of the product plays an important role in Bayesian analysis of the optimal portfolio. We derive the novel stochastic representation for the product and observe from the stochastic representation that the distribution of the product is closed under conditioning, marginalization, and affine transformations. Moreover, the formulae for the first four moments, density function and an Edgeworth type expansion are explicitly presented.

In section 4, we consider discriminant analysis in the case of two multivariate normal populations with different means and common covariance matrices. We derive the posterior predictive density function and the first four moments of the population linear discriminant function under some prior distributions. Based on the derived posterior predictive density function, we consider the Bayesian estimation for the misclassification rate associated with a population linear discriminant function, referred to as the optimal error rate. We obtain an explicit expression of the Bayes estimator of the optimal error rate. Although the Bayes estimator of the optimal error rate is expressed by the infinite sums and special functions in general, it is simply expressed under some conditions. In addition, an Edgeworth type expansion for the Bayes estimator is suggested based on the approximate posterior predictive distribution of the population linear discriminant function.

Acknowledgements

First and foremost, I would like to thank my supervisor Prof. Akio Suzukawa. He has patiently guided me for 7 years since my undergraduate days. During my research consultations with him, I was able to learn not only how to write research papers and specialized knowledge in statistics, but also how to live as a researcher.

I would also like to express my gratitude to my associate advisor, Prof. Yoshihide Kakizawa. He always cared about me and gave me concrete advice on the direction of my research.

I would like to thank Prof. Shingo Takagi. Thank you very much for your kind attention to my unclear points in my research.

Thanks also go to Prof. Hikaru Hasegawa, and Prof. Shintaro Sono. Their valuable comments on my research and their teaching were a definite driving force for my research.

There are many other people I would like to thank as well, so I will name them here (in alphabetical order); Dias Tulepov, Eriko Kato, Katsuya Nakamori, Kazuya Nagaoke, Ken Inosaki, Masatoshi Sato, Miho Murakami, Nao Inomata, Natsuha Tsubaki, Shiho Sogahata, Shu Nemoto, Shunki Tokuchi, Yusuke Iwama.

This work was supported by JST SPRING, Grant Number JPMJSP2119.

Last but not least, I would like to express my deepest gratitude to my father, Akitoshi Yonenaga, and my mother, Asami Yonenaga. They have supported my education for the past 22 years. I could not have survived without the support of my parents at all times in my life.

List of Figures

List of Tables

Contents

1 Preface

The multivariate normal distribution lies at the heart of multivariate statistical analysis. Statisticians often discuss the multivariate statistical method under the assumption of multivariate normality. Central limit theorem—sum of the identically and independently distributed random variables is asymptotically multivariate normal distribution — is one of the reasons for assuming the multivariate normality for the population. Most of the books on the multivariate statistical analysis, for example, Anderson (2003), Muirhead (1982) and Siotani et al. (1985) discuss the distributional properties of the multivariate normal distribution in detail. Gupta and Nagar (2000) discusses the properties of the matrix variate normal distribution, which is an extension of the multivariate normal distribution. In particular, they present a number of results about the moments of the matrix variate normal distribution.

Another important topic on the multivariate normal distribution is Stein's identity. This identity gives a recurrence formula for functions of a random vector which is according to the multivariate normal distribution. Stein's identity is not only used for calculating the moments of the multivariate normal distribution, but also for shrinkage estimation for mean vector (Stein, 1981).

The Wishart distribution also plays an important role in multivariate statistical analysis, as well as the multivariate normal distribution. There are many studies on the Wishart matrix and related quantities. Theorem 3.2.10 of Muirhead (1982), for example, provided the fundamental properties of the partitioned Wishart matrix. Bodnar and Okhrin (2008) established an analog to Theorem 3.2.10 of Muirhead (1982) for the singular, inverse and generalized inverse partitioned Wishart matrices. A number of papers deal with the moments of a Wishart matrix. Readers may refer to Watamori (1990), von Rosen (1991a), Sultan and Tracy (1996), Letac and Massam (2004), Drton et al. (2008), and Hillier and Kan (2021).

Haff's identity— a recurrence formula for functions of a Wishart matrix— is also an important as well as the Stein's identity. This identity is not only used to derive some moments of a Wishart matrix but also to be applied to improved estimation of the covariance matrix and its precision for the multivariate normal distribution (e.g. Haff ; 1977, 1979a, 1979b, 1979c, 1980).

It was pointed out by Bodnar and his co-authors (2011, 2013, 2015, 2019) that the distribution of the product of a Wishart matrix and a normal vector is often important in multivariate statistical analysis and portfolio theory, but has not been studied very much. To motivate the research on the product, we will begin with the brief description of portfolio theory and discriminant analysis, and then summarize the previous researches on the product. The mean-variance portfolio theory of Markowitz (1952) has been a fundamental approach to asset allocation. The optimal portfolios are obtained by maximizing the expected return subject to given level of the risk or by minimizing by the variance of the portfolio return. Let $\mathbf{x}_t$ denote a random vector of returns on k assets taken at time point t, and assume that $\mathbf{x}_1, \mathbf{x}_2, ...$ are independently and identically distributed as a multivariate normal distribution with the mean vector $\boldsymbol{\mu}$ and covariance matrix $\boldsymbol{\Sigma}$, where $\boldsymbol{\Sigma}$ is positive definite. This assumption is denoted by $\mathbf{x}_t \sim N_k(\boldsymbol{\mu}, \boldsymbol{\Sigma})$. Denoting the vector of portfolio weights by $\mathbf{w}$, this optimization problem is expressed as

$$\text{min } \sigma_p^2 = \text{min } \mathbf{w}'\boldsymbol{\Sigma}\mathbf{w} \text{ s.t. } \mathbf{w}'\boldsymbol{\mu} = \mu_p \text{ and } \mathbf{w}'\mathbf{1} = 1,$$

where $\mathbf{1}$ is a vector of ones, and μ_p is a given level of expected return of the portfolio.

The set of optimal portfolios obtained for all μ_p is called as the efficient frontier, which is an upper part of a parabola in the mean–variance space (cf. Merton, 1972). According to Bodnar and Schmid (2009), the efficient frontier has the following three parameters:

$$R_{GMV} = \frac{\mathbf{1}'\boldsymbol{\Sigma}^{-1}\boldsymbol{\mu}}{\mathbf{1}'\boldsymbol{\Sigma}^{-1}\mathbf{1}}, \tag{1}$$

$$V_{GMV} = \frac{1}{\mathbf{1}'\boldsymbol{\Sigma}^{-1}\mathbf{1}}, \tag{2}$$

$$s = \boldsymbol{\mu}'\boldsymbol{\Sigma}^{-1}\boldsymbol{\mu} - \frac{(\mathbf{1}'\boldsymbol{\Sigma}^{-1}\boldsymbol{\mu})^2}{\mathbf{1}'\boldsymbol{\Sigma}^{-1}\mathbf{1}}, \tag{3}$$

which are referred to as the expected return of the global minimum variance portfolio (henceforth, GMV portfolio), variance of GMV portfolio and the slope parameter of the parabola, respectively. In addition, the efficient frontier satisfy the following equation:

$$(R - R_{GMV})^2 = s(V - V_{GMV}).$$

If short-selling is allowed and a risk-free asset with return r_f is available, there are two portfolio weights which are derived from the efficient frontier: one is the tangency portfolio weights, the other is the GMV portfolio weights. This is seen in Figure 1. We observe from Figure 1 that the tangency portfolio weights corresponds to the tangency point between the efficient frontier and a capital market line drawn from the point $(R, V) = (r_f, 0)$, where r_f stands for the rate of risk-free asset. Moreover, the GMV portfolio is the point which possesses the smallest variance among all portfolios on the efficient frontier.

The explicit form of the tangency and GMV portfolio weights is given by

$$\mathbf{w}_{TP} = \alpha^{-1}\boldsymbol{\Sigma}^{-1}(\boldsymbol{\mu} - r_f\mathbf{1}), \tag{4}$$

$$\mathbf{w}_{GMV} = \frac{\mathbf{1}'\boldsymbol{\Sigma}^{-1}}{\mathbf{1}'\boldsymbol{\Sigma}^{-1}\mathbf{1}}, \tag{5}$$

where α denotes the risk aversion of the investor.

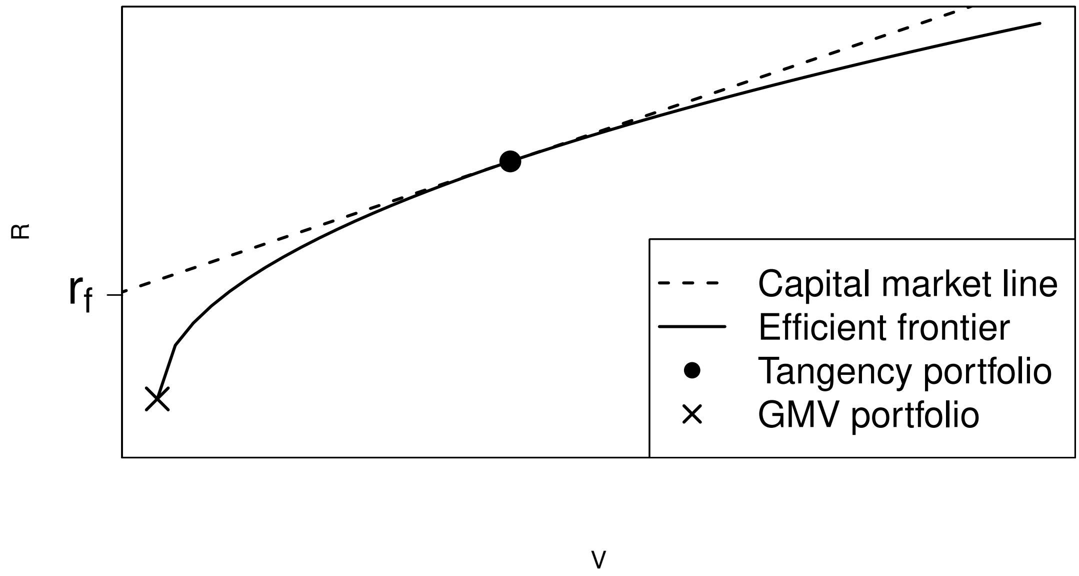


Figure 1: Illustration of the efficient frontier, capital market line, global minimum variance portfolio (GMV portfolio) and the tangency portfolio in the mean–variance space. r_f denotes the rate of risk-free asset.

If a risk-free asset is unavailable, we can use the portfolio weights which maximize the Sharp ratio defined by $SR = \mu_p/\sigma_p$. This portfolio weights are referred to as the Sharpe ratio optimal portfolio weights:

$$\mathbf{w}_{SR} = \frac{\mathbf{\Sigma}^{-1}\boldsymbol{\mu}}{\mathbf{1}'\mathbf{\Sigma}^{-1}\boldsymbol{\mu}}. \tag{6}$$

Secondly, we outline the theory of discriminant analysis for two multivariate normal populations. Let $\pi_i : N_p(\boldsymbol{\mu}_i, \mathbf{\Sigma}_i)$ be the multivariate normal populations with mean $\boldsymbol{\mu}_i$ and the covariance matrix $\mathbf{\Sigma}_i$ for $i = 1, 2$. Then the density function of the population π_i is

$$p_i(\mathbf{x}) = p_i(\mathbf{x}|\boldsymbol{\mu}_i, \mathbf{\Sigma}) = \frac{1}{(2\pi)^{p/2}|\Sigma|^{1/2}} \exp\left\{-\frac{1}{2}(\mathbf{x}-\boldsymbol{\mu}_i)'\mathbf{\Sigma}_i^{-1}(\mathbf{x}-\boldsymbol{\mu}_i)\right\}. \tag{7}$$

The log likelihood of $p_1(\mathbf{x})$ and $p_2(\mathbf{x})$ is given by

$$\begin{aligned} U_0 = \log\frac{p_1(\mathbf{x})}{p_2(\mathbf{x})} &= \frac{1}{2}\mathbf{x}'(\mathbf{\Sigma}_2^{-1} - \mathbf{\Sigma}_1^{-1})\mathbf{x} + \mathbf{x}'(\mathbf{\Sigma}_1^{-1}\boldsymbol{\mu}_1 - \mathbf{\Sigma}_2^{-1}\boldsymbol{\mu}_2) \\ &+ \frac{1}{2}(\boldsymbol{\mu}_2'\mathbf{\Sigma}_2^{-1}\boldsymbol{\mu}_2 - \boldsymbol{\mu}_1'\mathbf{\Sigma}_1^{-1}\boldsymbol{\mu}_1) + \frac{1}{2}\log\frac{|\mathbf{\Sigma}_2|}{|\mathbf{\Sigma}_1|}. \end{aligned} \tag{8}$$

Suppose we have a priori probabilities q_i for $i = 1, 2$ of drawing an observation from π_i and let $C(j|i)$ be the cost due to mis-discrimination where $C(i|i) = 0$. Then we obtain from Theorem 9.2.1 of Siotani et al. (1985) that the Bayes rule for given (q_1, q_2) is classifying a new observation into π_1 if $U_0 \geq c$, whereas into π_2 otherwise, where $c = \log[q_2 C(1|2)/q_1 C(2|1)]$. When $\mathbf{\Sigma}_1 = \mathbf{\Sigma}_2 = \mathbf{\Sigma}$, U_0 reduces to

$$U = (\boldsymbol{\mu}_1 - \boldsymbol{\mu}_2)'\mathbf{\Sigma}^{-1}\left\{\mathbf{x} - \frac{1}{2}(\boldsymbol{\mu}_1 + \boldsymbol{\mu}_2)\right\}, \tag{9}$$

which is known as the population linear discriminant function.

In a frequentist perspective, the population covariance matrix $\mathbf{\Sigma}$ and mean vector $\boldsymbol{\mu}$ are often estimated by the sample covariance matrix and sample mean vector, respectively. The distribution of the sample covariance matrix is a Wishart distribution and the distribution of the sample mean vector is the multivariate normal distribution. If we consider the product of the inverse of the sample covariance matrix and a sample mean vector, then we should discuss the product of an inverse Wishart matrix and a normal vector. Later on, we will look at the previous researches on the distribution of the product of an inverse Wishart matrix and a normal vector. Bodnar and Okhrin (2011) derived the density function and characteristic function of the product of an inverse Wishart matrix and a normal vector, which are represented by multiple integrals and special functions. These results are extension of Bodnar and Schmid (2006) where they dealt with the product of an inverse χ^2 and a normal variable. On the basis of the derived density function of the product of an inverse Wishart matrix and a normal vector, they also provided the stochastic representation of the linear combination of the product. In addition, Walds type test statistic was devised to test the general linear hypothesis for $\mathbf{\Sigma}^{-1}\boldsymbol{\mu}$, and its distribution was given. The suggested test can be directly applied to test the significance of the coefficients of the linear discriminant function, and the tangency and Sharp ratio optimal portfolio weights given by (9), (4) and (6), respectively. In the empirical analysis, they exemplify that the classical approach to the estimation of the portfolio weights could be very poor, and explain this phenomena in terms of the confidence region and test theory for the GMV portfolio weights (5). As for the GMV portfolio weights, Bodnar and Schmid (2008) showed that the distribution of the sample GMV portfolio weights are identical to the multivariate t distribution under an elliptically contoured population. In addition, they provided the statistical test for the general linear hypothesis of the GMV portfolio weights. Bodnar et al. (2017) considered the Bayesian estimation of the GMV portfolio under the multivariate normal population. They firstly derived the posterior distributions for $\boldsymbol{\mu}$ and $\mathbf{\Sigma}$ under the several prior distributions including non-informative and informative priors. In Theorem 1 of their paper, they derived the posterior distributions of

$$\boldsymbol{\theta} = \frac{\mathbf{L}\mathbf{\Sigma}^{-1}\mathbf{1}}{\mathbf{1}'\mathbf{\Sigma}^{-1}\mathbf{1}},$$

where $\mathbf{L}$ and $\mathbf{1}$ are the constant matrix and vector of ones, respectively. Meanwhile, they directly assumed prior distributions for $\boldsymbol{\theta}$, and constructed the Jeffreys' prior distribution for $\boldsymbol{\theta}$, which resulted in the multivariate t distribution. In Theorem 2, they evaluated the posterior distribution for $\boldsymbol{\theta}$ under the derived Jeffreys' prior distribution.

Bodnar et al. (2019) discussed the distribution of the product of a Moore-Penrose inverse of a (singular) Wishart matrix and a normal vector whose covariance matrix is not full rank. This situation can be considered when the strong correlations of asset returns can yield the near singularity of the population covariance, or the number of asset in a portfolio is greater than the number of observations. They derived the stochastic representation of the product and developed the statistical tests on the estimated tangency portfolio weights. In addition, the asymptotic normality was established under the high-dimensional regime. Kotsuiba and Mazur (2014) established the asymptotic normality of the product of an inverse Wishart matrix and a normal vector and suggested an integral approximation for the product, which is based on the third order Taylor expansion for the density function of the product. Bodnar et al. (2019) derived the stochastic representation for the linear combination of the product of an inverse Wishart matrix and a normal vector, and suggested the use of a test statistic to investigate whether two population linear discriminant function coefficients are equal as well as a coefficient in the linear discriminant function is larger than another one. Moreover, they established the asymptotic normality of the product based on the derived stochastic representation when the sample size and dimension goes to infinity, while Bodnar et al. (2016) proved the asymptotic normality of the product under the high-dimensional asymptotic regime when the population is a matrix variate family of skewed distributions. Javed et al. (2021) derived the higher order moments of the product of an inverse Wishart matrix and a normal vector, which involves a confluent hypergeometric function in general. In particular, the first four moments of the product were provided in closed-form. In an empirical illustration, they computed the first four moments of the estimated portfolio weights for the weekly log returns of four financial indices listed in NASDAQ stock exchange, and concluded that the distribution of estimated portfolio weights are well approximated by a normal distribution.

In Bayesian context, we often deal with the distribution of the product of a Wishart matrix and a normal vector, since the prior or posterior distribution for Σ^{-1} and $\boldsymbol{\mu}$ is a Wishart distribution and multivariate normal distribution, respectively.

Bodnar et al. (2013) derived the stochastic representation of the product of a Wishart matrix and a normal vector, which are independently distributed. Although the simulated value of the product is usually generated from $k(k+1)/2+k$ random variables, the derived representation enables us to obtain the simulated value from less than $2(k+1)$ random variables. This fact implies that the derived representation is very efficient for computational purposes. Based on the derived stochastic representation, they constructed the density function of the product which involves the multi-dimensional integrals. They also found an approximation for the density function of the product based on the Gaussian integral and the third-order Taylor series expansion, and documented a good performance. Bodnar et al. (2015) extended the results of Bodnar et al. (2013) by investing the distributional properties of the product of a singular Wishart matrix and a normal vector. Firstly, they presented the distributional properties of a partitioned singular Wishart matrix and characteristic function of the singular Wishart matrix. The former result is an extension of the results of Srivastava (2003) and Bodnar and Okhrin (2008). On the basis of the derived properties of the singular Wishart matrix, they proved that the stochastic representation of the product of the singular Wishart matrix and a normal vector has the same form as the representation of the product of the non-singular Wishar matrix and a normal vector. This is stated as follows:

Claim 1.1 (Bodnar and his co-authors; 2013, 2015)**.** *Let* $\mathbf{A} \sim W_k(n, \mathbf{\Sigma})$ *and* $\mathbf{z} \sim N_k(\boldsymbol{\mu}, \lambda\mathbf{\Sigma})$*, where* $\mathbf{\Sigma}$ *is positive definite. Assume that* $\mathbf{A}$ *and* $\mathbf{z}$ *are independent. Let* $\mathbf{L}$ *be a* $p \times k$ *constant matrix with rank* $p < k$ *and* $\mathbf{S}_1 = (\mathbf{L\Sigma L}')^{-1/2}\mathbf{L\Sigma}^{1/2}$ *and* $\mathbf{S}_2 = (\mathbf{I}_k - \mathbf{S}_1'\mathbf{S}_1)^{1/2}$*,*

$$\mathbf{LAz} \stackrel{\mathrm{d}}{=} \xi(\mathbf{L\Sigma L}')^{1/2}\mathbf{y}_1 + \sqrt{\xi}(\mathbf{L\Sigma L}')^{1/2}\left[\sqrt{\mathbf{y}_1'\mathbf{y}_1+\eta}\,\mathbf{I}_p - \frac{\sqrt{\mathbf{y}_1'\mathbf{y}_1+\eta}-\sqrt{\eta}}{\mathbf{y}_1'\mathbf{y}_1}\mathbf{y}_1\mathbf{y}_1'\right]\mathbf{z}_0,$$

where $\eta = \mathbf{y}_2'\mathbf{y}_2$*,* $\xi \sim \chi^2_n$*,* $\mathbf{z}_0 \sim N_p(\mathbf{0}, \mathbf{I}_p)$*,*

$$\mathbf{y} = \begin{pmatrix}\mathbf{y}_1\\ \mathbf{y}_2\end{pmatrix} \sim N_k\left(\begin{pmatrix}\mathbf{S}_1\mathbf{\Sigma}^{1/2}\boldsymbol{\mu}\\ \mathbf{S}_2\mathbf{\Sigma}^{1/2}\boldsymbol{\mu}\end{pmatrix}, \lambda\begin{pmatrix}\mathbf{S}_1\mathbf{\Sigma}\mathbf{S}_1' & \mathbf{S}_1\mathbf{\Sigma}\mathbf{S}_2'\\ \mathbf{S}_2\mathbf{\Sigma}\mathbf{S}_1' & \mathbf{S}_2\mathbf{\Sigma}\mathbf{S}_2'\end{pmatrix}\right).$$

In addition, ξ *and* $\mathbf{z}_0$ *are independent of* $\mathbf{y}$.

This result implies that all distributional properties on the product of a Wishart matrix and a normal vector remain valid for the product of a singular Wishart matrix and a normal vector.

Bodnar et al. (2019) dealt with the distribution of the product of a (singular) Wishart matrix and a singular normal vector, which are independently distributed. Additionally, they assumed that the rank of the covariance matrix of a (singular) Wishart matrix is not full rank, and provided the results on the distribution of a linear symmetric transformation of the Wishart matrix. On the basis of the derived distributional properties on the Wishart matrix, they provided the stochastic representation of the product, and the characteristic function of the product. Under the double asymptotic regime, i.e., when both the rank of the covariance matrix and degrees of freedom of a Wishart matrix denoted by r and n, respectively, tend to infinity such that $r/n \to c \in [0, +\infty)$, they established the asymptotic normality of the linear combination of the product.

In section 2, we extend the result of Bodnar and his co-authors (2013, 2015) with providing the stochastic representation of the product of a (singular) Wishart matrix and a normal vector with uncommon covariance matrix. This setting is seen in Bayesian statistics when we assume the semi-conjugate prior distribution for the parameters of multivariate normal distribution (cf. Hoff, 2009). The derived density function is expressed by multi-integrals of the complicated univariate density functions, which leads to the difficulties in investing the characteristic of the distribution of the product. To understand the characteristic of the distribution in detail, the first four moments are useful. In particular, the skewness and kurtosis inform us of how the distribution deviates from a normal distribution, so we study the higher order moments of the product. Finally, we propose a Edgeworth expansion for the product, where we explicitly use the result of moments.

Bauder and his co-authors (2018, 2020) considered the product of a Wishart matrix and a conditional normal vector given a Wishart matrix. This situation corresponds to the conjugate normal inverse Wishart prior or posterior distribution for $\mathbf{\Sigma}^{-1}$ and $\boldsymbol{\mu}$. They derived the stochastic representation of the product by employing the method detailed in Bodnar and his co-authors (2013, 2015, 2019). Moreover, the first two moments and the asymptotic normality were established. In the simulation study, they applied the obtained results to compute the coverage probabilities of credible intervals for the tangency portfolio weights.

In section 3, we consider the product of a Wishart matrix and a conditional normal vector given a Wishart matrix. We obtain the stochastic representation of the product which is highly computationally efficient than the stochastic representation which is given by Bauder and his co-authors (2018, 2020). Since the obtained stochastic representation shows that the distribution of the product belongs to an existing family of distributions: generalized hyperbolic distributions (cf. Blæsild and Jensen, 1981). This fact enables us to access the explicit expression for the density function and the first four moments of the product. In addition, it turns out that the distribution of the product is closed under conditioning, marginalization, and affine transformation. Since the density function of the product involves the special function in general, it is difficult to obtain the explicit expression of the cumulative distribution function. To evaluate the cumulative distribution function of the product, we provide a Edgeworth expansion for the product.

Although the distribution of the product of (inverse) Wishart matrix and a normal vecotor is an important statistic, more complicated functional form including the product appears in multivariate analysis and portfolio theory. In portfolio theory, some parameters in the efficient frontier are more complex functions in contrast to the product. Bodnar and Schmid (2009) considered the several statistical tests for the expected return R_{GMV} and the slope parameter s, which is defined by (1) and (3), respectively. They proved that a confidence region of the efficient frontier is bordered five parabolas. Other references which have studied the parameters of the efficient frontier are Jobson and Korkie (1980), Jobson (1991), Bodnar (2004), Kan and Smith (2008) and Bauder et al. (2019). In particular, Bauder et al. (2019) discussed the distributional properties of the efficient frontier from a Bayesian perspective. Under the diffuse and conjugate prior distributions, they derived the stochastic representation of the posterior distributions for the parameters of the efficient frontier, i.e. variance, expected return of the GMV portfolio and slope parameter. Based on the derived stochastic representation, they derived the first two moments and established the asymptotic normality of the parameters of the efficient frontier. In the appendix of section 3, we discuss the results given in Bauder et al. (2019) - in particular, the stochastic representations for the parameters of the efficient frontier.

In multivariate analysis, linear discriminant function is one of the important statistic in multivariate statistical analysis and its distribution from frequentist perspective are well-studied (cf. Okamoto, 1963; Siotani and Wang, 1977; Anderson, 1973). In this dissertation, we discuss the distribution of linear discriminant function from Bayesian perspective. There are some papers on the linear discriminant analysis in a Bayesian setting. We take a brief discussion on the Bayesian discriminant analysis. The posterior probability that the new observation $\mathbf{y}$ arises from either multivariate normal population $\pi_1 : N_p(\boldsymbol{\mu}_1, \boldsymbol{\Sigma}_1)$ or $\pi_2 : N_p(\boldsymbol{\mu}_2, \boldsymbol{\Sigma}_2)$ follows from Bayes' theorem that

$$Pr(\mathbf{y} \in \pi_i|\mathbf{y}) = \frac{q_i p_i(\mathbf{y})}{q_1 p_1(\mathbf{y}) + q_2 p_1(\mathbf{y})},$$

where q_i is a prior probability of drawing an observation from π_i and $p_i(\mathbf{y})$ is defined as (7). If the population parameters and prior probabilities are known, then we can place the new observations in the population where the posterior probability is maximized. However, in most cases, they are unknown and need to be estimated. According to Rigby (1982), the estimate of the posterior probability obtained by substituting the posterior predictive distribution of $\mathbf{y}$ for $p_i(\mathbf{y})$ is called the "predictive" estimate. Geisser (1964, 1982) calculated the posterior predictive distribution under various assumptions for cases where the prior probability is known or unknown, and obtained "predictive" estimates. In Rigby (1992), the "predictive" estimate and the classical estimate were compared, and it was shown through simulation that the "predictive" estimate performed better for the posterior probability $Pr(\mathbf{y} \in \pi_i|\mathbf{y})$ than the classical estimate. Finally, the credible interval for the posterior probability $Pr(\mathbf{y} \in \pi_i|\mathbf{y})$ was approximated by using Pearson curve systems. On the other hand, Rigby (1982) calculated the credible interval of

$Pr(\mathbf{y} \in \pi_i|\mathbf{y})$ from the posterior distribution of $z_i = p_i(\mathbf{y})$. Their method is as follows: they calculated the moments of $L = \log(z_1/z_2)$, approximated the distribution of L using the Peason curve systems, and calculated the credible interval for $Pr(\mathbf{y} \in \pi_i|\mathbf{y})$ using the transformation given by

$$Pr(\mathbf{y} \in \pi_i|\mathbf{y}) = \frac{1}{1 + (q_2/q_1)\exp(-L)}.$$

Fatti (1982) also discussed the "predictive" estimation of the posterior probability $Pr(\mathbf{y} \in \pi_i|\mathbf{y})$. They extended the discussion in Geisser (1964, 1966) and calculated the posterior predictive distribution of $\mathbf{y}$ under a random-effect model in which each of k normal populations is randomly selected. The random-effect model can be formulated as $\pi_i : N_p(\boldsymbol{\mu}_i, \boldsymbol{\Sigma})$ with $\boldsymbol{\mu}_i \sim N_p(\boldsymbol{\xi}, \mathbf{T})$ for $i = 1, \cdots, k$. They employed both the usual diffuse prior and Box-Tiao's non-informative prior distributions (cf. Box and Tiao, 1973) as the prior distributions for the unknown $\boldsymbol{\Sigma}$, $\boldsymbol{\xi}$, and $\mathbf{T}$. The resulting posterior predictive distributions of $\mathbf{y}$ under these prior distributions are expressed by a hypergeometric function of matrix argument.

Fatti (1983) discussed the linear discriminant analysis under the random-effects model. They derived the density function of the population-based and sample-based Mahalanobis distances between two different populations, as well as those between an observation and a randomly selected population. These results were used for evaluating the probability of the misclassification associated with population linear discriminant and sample linear discriminant under the random-effects model. Since all of the estimates of the probability of the misclassification include the eigenvalues of $\mathbf{T}\boldsymbol{\Sigma}^{-1}$, they estimated these eigenvalues by using empirical Bayes method.

Geisser (1967, 1982) considered the Baysian estimation of the two probabilities of misclassification: one is the optimal error rate, the other is the actual error rate. The optimal error rates are the probability of misclassification that occurs when we use the population linear discriminant (9) with cut-off point x, and defined as

$$\epsilon_1 = \epsilon_1(\boldsymbol{\mu}_1, \boldsymbol{\mu}_2, \boldsymbol{\Sigma}) = Pr[U < x|\boldsymbol{\mu}_1, \boldsymbol{\mu}_2, \boldsymbol{\Sigma}, \mathbf{y} \in \pi_1] = \Phi\left(\frac{x - \Delta^2/2}{\Delta}\right),$$
$$\epsilon_2 = \epsilon_2(\boldsymbol{\mu}_1, \boldsymbol{\mu}_2, \boldsymbol{\Sigma}) = Pr[U > x|\boldsymbol{\mu}_1, \boldsymbol{\mu}_2, \boldsymbol{\Sigma}^{-1}, \mathbf{y} \in \pi_2] = \Phi\left(\frac{x + \Delta^2/2}{\Delta}\right),$$

where $\Phi(\cdot)$ denotes the cumulative distribution function of the standard normal distribution, and Δ^2 denotes the squared Mahalanobis distance between π_1 and π_2, which is defined by $\Delta^2 = (\boldsymbol{\mu}_1 - \boldsymbol{\mu}_2)'\boldsymbol{\Sigma}^{-1}(\boldsymbol{\mu}_1 - \boldsymbol{\mu}_2)$.

Suppose we collect two mutually independent random samples $\mathbf{x}_{11}, \cdots, \mathbf{x}_{1N_1}$ and $\mathbf{x}_{21}, \cdots, \mathbf{x}_{2N_2}$ drawn from π_1 and π_2, respectively. Our estimate of $\boldsymbol{\mu}_1$ is $\bar{\mathbf{x}}_1 = \sum_{\alpha=1}^{N_1} \mathbf{x}_{1\alpha}/N_1$, of $\boldsymbol{\mu}_2$ is $\bar{\mathbf{x}}_2 = \sum_{\alpha=1}^{N_2} \mathbf{x}_{2\alpha}/N_2$, and of $\boldsymbol{\Sigma}$ is $\mathbf{S}$ defined by $n\mathbf{S} = \sum_{\alpha=1}^{N_1}(\mathbf{x}_{1\alpha} - \bar{\mathbf{x}}_1)(\mathbf{x}_{1\alpha} - \bar{\mathbf{x}}_1)' + \sum_{\alpha=1}^{N_2}(\mathbf{x}_{2\alpha} - \bar{\mathbf{x}}_2)(\mathbf{x}_{2\alpha} - \bar{\mathbf{x}}_2)'$ with $n = N_1 + N_2 - 2$.

The actual error rates are the probability of misclassification that occurs when we use the sample linear discriminant

$$V = (\bar{\mathbf{x}}_1 - \bar{\mathbf{x}}_2)'\mathbf{S}^{-1}\left\{\mathbf{y} - \frac{1}{2}(\bar{\mathbf{x}}_1 + \bar{\mathbf{x}}_2)\right\}.$$

The actual error rates are defined as

$$\begin{aligned}
\beta_1 = \beta_1(\boldsymbol{\mu}_1, \boldsymbol{\mu}_2, \boldsymbol{\Sigma}) &= Pr[V < x|\boldsymbol{\mu}_1, \boldsymbol{\mu}_2, \boldsymbol{\Sigma}^{-1}, \mathbf{y} \in \pi_1] \\
&= \Phi\left(\frac{(\bar{\mathbf{x}}_1 - \bar{\mathbf{x}}_2)'\mathbf{S}^{-1}\left[(\bar{\mathbf{x}}_1 + \bar{\mathbf{x}}_2)/2 - \boldsymbol{\mu}_1\right] + x}{\sqrt{(\bar{\mathbf{x}}_1 - \bar{\mathbf{x}}_2)'\mathbf{S}^{-1}\boldsymbol{\Sigma}\mathbf{S}^{-1}(\bar{\mathbf{x}}_1 - \bar{\mathbf{x}}_2)}}\right), \\
\beta_2 = \beta_2(\boldsymbol{\mu}_1, \boldsymbol{\mu}_2, \boldsymbol{\Sigma}) &= Pr[V > x|\boldsymbol{\mu}_1, \boldsymbol{\mu}_2, \boldsymbol{\Sigma}^{-1}, \mathbf{y} \in \pi_2] \\
&= \Phi\left(\frac{(\bar{\mathbf{x}}_1 - \bar{\mathbf{x}}_2)'\mathbf{S}^{-1}\left[(\bar{\mathbf{x}}_1 + \bar{\mathbf{x}}_2)/2 - \boldsymbol{\mu}_2\right] + x}{\sqrt{(\bar{\mathbf{x}}_1 - \bar{\mathbf{x}}_2)'\mathbf{S}^{-1}\boldsymbol{\Sigma}\mathbf{S}^{-1}(\bar{\mathbf{x}}_1 - \bar{\mathbf{x}}_2)}}\right),
\end{aligned}$$

It is noted that β_1 and β_2 are random variables that are functions of $\boldsymbol{\mu}_1$, $\boldsymbol{\mu}_2$ and $\boldsymbol{\Sigma}$, whereas $\bar{\mathbf{x}}_1$, $\bar{\mathbf{x}}_2$ and $\mathbf{S}$ are constants.

Let Q be the sample squared Mahalanobis distance defined by $(\bar{\mathbf{x}}_1 - \bar{\mathbf{x}}_2)'\mathbf{S}^{-1}(\bar{\mathbf{x}}_1 - \bar{\mathbf{x}}_2)$. Although the exact posterior distributions for β_1 and β_2 are found in Geisser (1967), the following approximations are also available:

$$Pr[\beta_1 \leq b] \approx \Phi\left(\frac{\Phi^{-1}(b) - A_1}{(N_1^{-1} + B_1)^{1/2}}\right),$$

where

$$A_1 = \left(\frac{[n-p+1]/2}{Qn}\right)^{1/2}(x - Q/2), B_1 = [x - Q/2]^2/(2nQ),$$

and

$$Pr[\beta_2 \leq b] \approx 1 - \Phi\left(\frac{\Phi^{-1}(1-b) - A_2}{(N_2^{-1} + B_2)^{1/2}}\right),$$

where

$$A_2 = \left(\frac{[n-p+1]/2}{Qn}\right)^{1/2}(x + Q/2), B_2 = [x + Q/2]^2/(2nQ).$$

The expected value of β_i is described by the posterior predictive density of V, $f(V|\pi_1) = f(V|\mathbf{y} \in \pi_1, \bar{\mathbf{x}}_1, \bar{\mathbf{x}}_2, \mathbf{S})$, as follows:

$$\mathrm{E}(\beta_1) = \int_{-\infty}^{x} f(V|\pi_1)dV$$

$$
\begin{aligned}
&= Pr[t_{n+1-p} \leq (x - Q/2)[n(N_1+1)Q/(n+1-p)N_1]^{-1/2}], \\
\mathrm{E}(\beta_2) &= \int_x^\infty f(V|\pi_2)dV \\
&= Pr[t_{n+1-p} > (x + Q/2)[n(N_2+1)Q/(n+1-p)N_2]^{-1/2}],
\end{aligned}
$$

where t_s is a t-distributed random variable with s degrees of freedom.

As for the posterior distribution of the optimal error rate ϵ_i, Geisser (1964) constructed a $(1-\alpha)\%$ confidence interval based on the posterior distribution of the squared Mahalanobis distance. In addition, Geisser (1964, 1982) also gave the following approximation for the posterior distribution of ϵ_1:

$$
Pr[\epsilon_1 \leq b] \approx 1 - G_s\left[\frac{4c^{-1}(p + c^{-1}Q)(\Phi^{-1}(b))^2}{p + c^{-1}Q + n^{-1}(cQ)^2}\right],
$$

where $G_s(\cdot)$ is the cumulative distribution function of χ^2_s random variable with $s = (p+c^{-1}Q)^2/(p+c^{-1}Q+n^{-1}(cQ)^2)$. In addition, they provided the normal approximation based on the first two moments of the posterior predictive distribution of population linear discriminant as follows:

$$
\mathrm{E}(\epsilon_1) \approx \Phi\left(\frac{x - (pc + Q)/2}{[pc + (1 + pc)Q]^{1/2}}\right). \tag{10}
$$

In section 4, we propose an Edgeworth expansion for $\mathrm{E}(\epsilon_1)$ to improve the accuracy of the approximation given by (10). We also obtain an exact representation for $\mathrm{E}(\epsilon_1)$ in the form involving infinite series and special functions. It will be seen that under special conditions this representation can be expressed only in terms of finite sums and elementary functions. In Geisser (1964), the expression of the posterior predictive distribution for the population linear discriminant was not given explicitly, and only moments up to the second order were given. Therefore, in this paper, the expression of the posterior predictive distribution for the population linear discriminant is given explicitly, and a precise expression of the moments up to the fourth order is derived. These results are useful for Bayesian estimation of the optimal error rate in the paper.

2 Distribution of the product of a Wishart matrix and a normal vector

2.1 Introduction

Functions of a Wishart matrix and a normal vector appear in a variety of multivariate statistical methods, such as principal components analysis, multiple comparison, and discriminant analysis under multivariate normality. Over the past ten years, there has been an increase in the number of studies on the product of a Wishart matrix and a normal vector. Bodnar and Okhrin (2011), for example, derived the density function of the product of an inverse Wishart matrix and a normal vector, which is expressed by the four dimensional integrals and the modified Bessel function of the first kind (cf. Abramowitz and Stegun, 1965). In addition, they discussed potential applications to portfolio theory and discriminant analysis. Meanwhile, Bodnar et al. (2013) presented the stochastic representation and density function of the product of a Wishart matrix and a normal vector. They also found an approximation for the density function based on the Gaussian integral and the third-order Taylor series expansion. Bodnar et al. (2014) investigated the distributional properties of the product of a singular Wishart matrix and a normal vector. The distributional properties of (inverse) singular Wishart distribution are also well studied by Díaz-García (1997), Srivastava (2003), Bodnar and Okhrin (2008), Bodnar and his co-authors (2016, 2018), among others. Kotsuiba and Mazur (2015) extended the results in Bodnar and Okhrin (2011) by proving the asymptotic and approximate density functions of the product of an inverse Wishart matrix and a normal vector. Bodnar et al. (2018) derived the stochastic representations of the product of a singular Wishart matrix and a singular normal vector, in addition to proving the asymptotic normality of the product under the high-dimensional asymptotic regime. Bodnar et al. (2019) derived the central limit theorems for functionals of large sample covariance matrix and mean vector when the population is a matrix-variate location mixture of a normal distribution. Javed et al. (2021) derived the higher order moments of the product of an inverse Wishart matrix and a normal vector.

The product of a Wishart matrix and a normal vector plays an important role in discriminant analysis and in portfolio theory. The coefficients of the sample linear discriminant function, for example, is the product of an inverse Wishart matrix and a normal vector, whereas the optimal portfolio weights are expressed as the product of a Wishart matrix and a normal vector from Bayesian perspective. Bodnar et al. (2020) derived the stochastic representations for the coefficients, and developed the statistical test for the coefficients. Bauder and his co-authors. (2018, 2020) provided the stochastic representations for the posterior distribution of the weights, and established asymptotic normality based on the derived stochastic representations.

Previous studies on the product of a Wishart matrix and a normal vector have assumed the covariance matrix of a normal vector is a multiple of that of a Wishart matrix. However, assuming an uncommon covariance structure (in other words, assuming the covariance matrix of a normal vector is different from that of a Wishart matrix), seems more appropriate when we consider that one can choose the semi-conjugate priors for the parameters of multivariate normal population, as discussed in Hoff (2009). In addition, there has been little research on moments of the product. Thus, the purpose of this study is to present the stochastic representation, density function and exact moments of the product under the uncommon covariance structure. The rest of this paper is structured as follows. Under the uncommon covariance structure, the stochastic representation and density function of the product are naturally derived from the result of Bodnar et al. (2013). In addition, we also derive the explicit moment formulae of the product in Section 2.2. The results of numerical studies are given in Section 2.3, while Section 2.5 summarizes the paper. The Appendix contains the result of the higher order expectations of the elements of a Wishart matrix, which are used to prove the main results presented in Section 2.2.

2.2 Main Results

In this section we present the main findings. In particular, we derive the moments of the product of a Wishart matrix and a normal vector, which are independent. Let $\mathbf{A}$ be a k-dimensional Wishart matrix with n degrees of freedom and covariance matrix $\mathbf{\Sigma}$. In this paper, we permit the singularity of a Wishart matrix, i.e. $\mathbf{A} \sim W_k(n, \mathbf{\Sigma})$, $k > n$. Let $\mathbf{z} \sim N_k(\boldsymbol{\mu}, \mathbf{\Omega})$; i.e., it follows a k-dimensional multivariate normal distribution with the mean vector $\boldsymbol{\mu}$ and covariance matrix $\mathbf{\Omega}$. Throughout the paper, we assume $\mathbf{\Sigma}$ and $\mathbf{\Omega}$ are positive definite, which is denoted by $\mathbf{\Sigma} > 0$ and $\mathbf{\Omega} > 0$.

2.2.1 Stochastic representation and density function

Stochastic representation is not only useful for Monte Carlo simulations but for the theory of elliptically contoured distributions (cf. Gupta et al., 2013). In what follows, we present the stochastic representation of $\mathbf{LAz}$ where $\mathbf{L}$ $(p \times k)$ is a constant matrix with rank $p < k$. If we take $\mathbf{L}$ as $\mathbf{l}' = (1, 0, \cdots, 0)$, we can investigate the distribution of the first element of the product, whereas vector of ones, we can investigate the distribution of the sum of the elements. In addition, the derived stochastic representation can be applied to construct the density function and to evaluate the general moment formulae of the product. The symbol $\overset{\mathrm{d}}{=}$ denotes the equality in distribution. In Theorem 2.1, we present a stochastic representation for linear combinations of the elements of the random vector $\mathbf{Az}$, which is represented by independent random variables: χ^2, and multivariate normal distributions.

Theorem 2.1. *Let* $\mathbf{A} \sim W_k(n, \mathbf{\Sigma})$ *and* $\mathbf{z} \sim N_k(\boldsymbol{\mu}, \mathbf{\Omega})$*, where* $\mathbf{\Sigma} > 0$ *and* $\mathbf{\Omega} > 0$*. Assume that* $\mathbf{A}$ *and* $\mathbf{z}$ *are independent. Let* $\mathbf{L}$ *be a* $p \times k$ *constant matrix with rank* $p < k$ *and* $\mathbf{S}_1 = (\mathbf{L\Sigma L}')^{-1/2}\tilde{\mathbf{L}}\tilde{\mathbf{\Sigma}}^{1/2}$ *and* $\mathbf{S}_2 = (\mathbf{I}_k - \mathbf{S}_1'\mathbf{S}_1)^{1/2}$*, where* $\tilde{\mathbf{L}} = \mathbf{L\Omega}^{-1/2}$ *and* $\tilde{\mathbf{\Sigma}} = \mathbf{\Omega}^{1/2}\mathbf{\Sigma}\mathbf{\Omega}^{1/2}$*. Then the stochastic representation of* $\mathbf{LAz}$ *is given by*

$$\mathbf{LAz} \overset{\mathrm{d}}{=} \xi(\mathbf{L\Sigma L}')^{1/2}\mathbf{y}_1 + \sqrt{\xi}(\mathbf{L\Sigma L}')^{1/2}\left[\sqrt{\mathbf{y}_1'\mathbf{y}_1 + \eta}\,\mathbf{I}_p - \frac{\sqrt{\mathbf{y}_1'\mathbf{y}_1 + \eta} - \sqrt{\eta}}{\mathbf{y}_1'\mathbf{y}_1}\mathbf{y}_1\mathbf{y}_1'\right]\mathbf{z}_0,$$

where $\eta = \mathbf{y}_2'\mathbf{y}_2$, $\xi \sim \chi^2_n$, $\mathbf{z}_0 \sim N_p(\mathbf{0}, \mathbf{I}_p)$,

$$\mathbf{y} = \begin{pmatrix} \mathbf{y}_1 \\ \mathbf{y}_2 \end{pmatrix} \sim N_k\left(\begin{pmatrix} \mathbf{S}_1\tilde{\mathbf{\Sigma}}^{1/2}\tilde{\boldsymbol{\mu}} \\ \mathbf{S}_2\tilde{\mathbf{\Sigma}}^{1/2}\tilde{\boldsymbol{\mu}} \end{pmatrix}, \begin{pmatrix} \mathbf{S}_1\tilde{\mathbf{\Sigma}}\mathbf{S}_1' & \mathbf{S}_1\tilde{\mathbf{\Sigma}}\mathbf{S}_2' \\ \mathbf{S}_2\tilde{\mathbf{\Sigma}}\mathbf{S}_1' & \mathbf{S}_2\tilde{\mathbf{\Sigma}}\mathbf{S}_2' \end{pmatrix}\right) \text{ with } \tilde{\boldsymbol{\mu}} = \mathbf{\Omega}^{-1/2}\boldsymbol{\mu}.$$

In addition, ξ *and* $\mathbf{z}_0$ *are independent of* $\mathbf{y}$.

Proof. If we assume that $\mathbf{A}$ is a non-singular Wishart matrix, then the stochastic representation of $\mathbf{LAz}$ can be written by

$$\mathbf{LAz} \overset{\mathrm{d}}{=} \tilde{\mathbf{L}}\mathbf{\Omega}^{1/2}\mathbf{A}\mathbf{\Omega}^{1/2}\mathbf{z}_1,$$

where $\mathbf{z}_1 \sim N_k(\tilde{\boldsymbol{\mu}}, \mathbf{I}_k)$. Using Theorem 3.2.5 of Muirhead (1982),

$$\mathbf{\Omega}^{1/2}\mathbf{A}\mathbf{\Omega}^{1/2} \sim W_k(n, \tilde{\mathbf{\Sigma}}).$$

If we decompose $\tilde{\mathbf{\Sigma}}$ as $\tilde{\mathbf{\Sigma}} = \tilde{\mathbf{\Sigma}}^{1/4}\tilde{\mathbf{\Sigma}}^{1/2}\tilde{\mathbf{\Sigma}}^{1/4}$, then

$$\mathbf{LAz} \overset{\mathrm{d}}{=} \tilde{\mathbf{L}}\tilde{\mathbf{\Sigma}}^{1/4}\mathbf{A}_1\tilde{\mathbf{\Sigma}}^{1/4}\mathbf{z}_1, \tag{11}$$

where $\mathbf{A}_1 \sim W_k(n, \tilde{\mathbf{\Sigma}}^{1/2})$. Since $\tilde{\mathbf{\Sigma}}^{1/4}\mathbf{z}_1 \sim N_k(\tilde{\mathbf{\Sigma}}^{1/4}\tilde{\boldsymbol{\mu}}, \tilde{\mathbf{\Sigma}}^{1/2})$, the covariance matrix of $\tilde{\mathbf{\Sigma}}^{1/4}\mathbf{z}_1$ is the same as that of $\mathbf{A}_1$. This implies that if we apply Theorem 1 of Bodnar et al. (2013) to (11), the desired result follows immediately.

Since Theorem 4 of Bodnar et al. (2014) showed that the stochastic representation of the product of a non-singular Wishart matrix and a normal vector which is given by Theorem 1 of Bodnar et al. (2013) remains valid for the product of a singular Wishart matrix and a normal vector. This fact implies that the stochastic representation presented in Theorem 2.1 works for the product of a singular Wishart matrix and a normal vector. □

Since the obtained stochastic representation allows us to simulate $p + k + 2$ random variables to obtain the realizations of the product, this representation is efficient for computational purposes, especially for larger values of k. Without this representation, we would have to simulate $k + k(k+1)/2$ random variables.

In Corollary 2.2, we present the exact density function of $\mathbf{LAz}$.

Corollary 2.2. *Assume the same conditions as Theorem* 2.2*. Let* $f_{N_p(\boldsymbol{\theta},\mathbf{T})}(\cdot)$ *be the density function of* p*-dimensional normal with mean* $\boldsymbol{\theta}$ *and covariance matrix* $\mathbf{T}$ *and* $f_{\chi^2_n}(\cdot)$ *be the density function of* χ^2 *variable with degrees of freedom* n*. The density function of* $\mathbf{LAz}$ *is given by*

$$f_{\mathbf{LAz}}(\mathbf{x}) = \int_0^\infty \int_0^\infty \int_{R^p} f_{N_p(\bar{\boldsymbol{\mu}},\bar{\boldsymbol{\Sigma}})}(\mathbf{x}|\xi = v, \mathbf{y}_1 = \mathbf{z}_1, \eta = z_2) f_{N_p(\mathbf{S}_1\tilde{\boldsymbol{\Sigma}}^{1/2}\tilde{\boldsymbol{\mu}},\mathbf{S}_1\tilde{\boldsymbol{\Sigma}}\mathbf{S}_1')}(\mathbf{z}_1) f_{\chi^2_n}(v) \\ \times f_{\eta|\mathbf{y}_1}(z_2|\mathbf{y}_1 = \mathbf{z}_1) d\mathbf{z}_1 dz_2 dv,$$

where $\bar{\boldsymbol{\mu}} = \xi(\mathbf{L\Sigma L}')^{1/2}\mathbf{y}_1$ *and* $\bar{\boldsymbol{\Sigma}} = \xi(\mathbf{L\Sigma L}')^{1/2}[(\mathbf{y}_1'\mathbf{y}_1 + \eta)\mathbf{I}_p - \mathbf{y}_1\mathbf{y}_1'](\mathbf{L\Sigma L}')^{1/2}$.

Proof. From the stochastic representation in Theorem 2.1, the conditional distribution of $\mathbf{LAz}$ given ξ, $\mathbf{y}_1$ and η is $N_p(\bar{\boldsymbol{\mu}}, \bar{\boldsymbol{\Sigma}})$, where $\bar{\boldsymbol{\mu}}$ and $\bar{\boldsymbol{\Sigma}}$ are defined in Corollary 2.2. To obtain the unconditional density function of $\mathbf{LAz}$, we firstly construct the joint density function of $\mathbf{LAz}$, ξ, $\mathbf{y}_1$ and η, and then compute the marginal density function of $\mathbf{LAz}$ by integrating out the two random variables ξ and η as well as the random vector $\mathbf{y}_1$. □

The conditional distribution of $\mathbf{y}_2$ given $\mathbf{y}_1$ is the $k-p$ dimensional normal distribution with the mean vector $\boldsymbol{\nu}$ and covariance matrix $\boldsymbol{\Psi}$, where

$$\boldsymbol{\nu} = \mathbf{S}_2\tilde{\boldsymbol{\Sigma}}^{1/2}\tilde{\boldsymbol{\mu}} + \mathbf{S}_2\tilde{\boldsymbol{\Sigma}}\mathbf{S}_1'(\mathbf{S}_1\tilde{\boldsymbol{\Sigma}}\mathbf{S}_1')^{-1}(\boldsymbol{y}_1 - \mathbf{S}_1\tilde{\boldsymbol{\Sigma}}^{1/2}\tilde{\boldsymbol{\mu}}),$$
$$\boldsymbol{\Psi} = \mathbf{S}_2\tilde{\boldsymbol{\Sigma}}\mathbf{S}_2' - \mathbf{S}_2\tilde{\boldsymbol{\Sigma}}\mathbf{S}_1'(\mathbf{S}_1\tilde{\boldsymbol{\Sigma}}\mathbf{S}_1')^{-1}\mathbf{S}_1\tilde{\boldsymbol{\Sigma}}\mathbf{S}_2'.$$

Let ${}_pF_q(a_1, \cdots, a_p : b_1, \cdots, b_q; z)$ is the generalized hypergeometric function defined as

$${}_pF_q(a_1, \cdots, a_p : b_1, \cdots, b_q; z) = \sum_{k=0}^\infty \frac{(a_1)_k \cdots (a_p)_k}{(b_1)_k \cdots (b_q)_k} \frac{z^k}{k!},$$

where $(a)_k = a(a+1)\cdots(a+k-1)$. It is noted that z can be a complex variable, whereas b_i for $i = 1, \cdots q$ are not allowed to take zero or a negative integer. In addition, if any numerator parameters is zero or a negative integer, the generalized hypergeometric function is a finite polynomials. The convergence of this series is documented for all finite z if $p \le q$, while for $|z| < 1$ if $p = q+1$. Conversely, this series diverges for $|z| > 1$ if $p = q+1$ and all $z \ne 0$ if $p > q+1$.

The density function of the product involves the conditional density function of the quadratic form η given $\mathbf{y}_1$, which is of the form

$$f_{\eta|\boldsymbol{y}_1}(y) = \sum_{i=0}^\infty c_i \frac{i!}{2\beta\Gamma\left(\frac{k-p}{2} + i\right)} \left(\frac{y}{2\beta}\right)^{\frac{k-p}{2}-1} \exp\left(-\frac{y}{2\beta}\right) L_i^{\left(\frac{k-p}{2}-1\right)}\left(\frac{y}{2\beta}\right),$$

where β is an arbitrary positive constant, and $L_i^{(\alpha)}(x)$ is the generalized Laguerre polynomial defined as

$$L_i^{(\alpha)}(x) = \sum_{r=0}^i \frac{(-x)^r}{r!(k-r)!}(\alpha + r + 1)(\alpha + r + 2)\cdots(\alpha + i)$$

$$= \frac{(\alpha+1)_k}{k!} {}_1F_1(-k, \alpha+1; x)$$
$$= \frac{(-x)^i}{i!} {}_2F_0(-i, -\alpha - i; -1/x).$$

In addition, the coefficients c_k for $k = 1, 2, 3, \cdots$ are available from the formula

$$c_0 = 1,\ c_i = \frac{1}{i!}\sum_{r=0}^{i-1} d_{i-r}c_r,\ d_j = \frac{1}{2}\sum_{j_1=1}^{k-p}(1 - jb_{j_1}^2)(2\alpha_{j_1})^{-j},\ j \geq 1,$$

where

$$\mathbf{H}'\mathbf{\Psi}\mathbf{H} = \mathrm{diag}(\alpha_1, \cdots, \alpha_{k-p}),\ \mathbf{H}\mathbf{H}' = \mathbf{I}_{k-p}$$

and $\alpha_1, \cdots, \alpha_{k-p}$ are the eigenvalues of $\mathbf{\Psi}$ and $\boldsymbol{b} = (b_1, \cdots, b_{k-p})' = \mathbf{H}'\mathbf{\Psi}^{-\frac{1}{2}}\boldsymbol{\nu}$.

If we put $\mathbf{\Sigma} = \mathbf{I}_k$ and $\mathbf{\Omega} = \lambda\mathbf{I}_k$ in Corollary 2.2, then the density function of $\mathbf{LAz}$ is given by

$$f_{\mathbf{LAz}}(\mathbf{x}) = \int_0^\infty \int_0^\infty \int_{R^p} f_{N_p(\bar{\boldsymbol{\mu}}, \bar{\mathbf{\Sigma}})}(\mathbf{x}|\xi = v, \mathbf{y}_1 = \mathbf{z}_1, \eta = z_2) f_{N_p(\mathbf{S}_1\tilde{\boldsymbol{\mu}}, \lambda\mathbf{S}_1\mathbf{S}_1')}(\mathbf{z}_1) f_{\chi_n^2}(v)$$
$$\times f_{\chi^2_{k-p};\delta^2}(\lambda^{-1}z_2|\mathbf{y}_1 = \mathbf{z}_1) d\mathbf{z}_1 dz_2 dv,$$

where $\bar{\boldsymbol{\mu}} = \xi(\mathbf{LL}')^{1/2}\mathbf{y}_1$, $\bar{\mathbf{\Sigma}} = \xi(\mathbf{LL}')^{1/2}[(\mathbf{y}_1'\mathbf{y}_1 + \eta)\mathbf{I}_p - \mathbf{y}_1\mathbf{y}_1'](\mathbf{LL}')^{1/2}$ and $\delta^2 = \lambda^{-1}\boldsymbol{\mu}'\mathbf{S}_2'\mathbf{S}_2\boldsymbol{\mu}$. Although the proof of this result is in line with the proof of Corollary 2 in Bodnar et al. (2013), the following explanation is omitted in their paper. Since $\mathbf{Q}(\mathbf{I}_k - \mathbf{Q}) = \mathbf{S}_1'\mathbf{S}_1\mathbf{S}_2'\mathbf{S}_2 = 0$, we can obtain $\mathrm{tr}(\mathbf{S}_1\mathbf{S}_2'\mathbf{S}_2\mathbf{S}_1') = 0$. If we use Lemma 5.3.1 of Harville (1997) with $A = \mathbf{S}_1\mathbf{S}_2$, then $\mathbf{S}_1\mathbf{S}_2 = \mathbf{0}$.

2.2.2 Exact moments

To understand the characteristics of the distribution of the product, it is important to compute the moments. In particular, the skewness and kurtosis play important roles in measuring the deviation from a normal distribution.

By using the similar approach which is detailed in Javed et al. (2021), we can compute the higher order moments of the product as follows:

Theorem 2.3. *Assume the same conditions as Theorem* 2.1. *Let* $\mathbf{l}$ *be a constant vector. The higher order moments of* $\mathbf{l}'\mathbf{Az}$ *are expressed by*

$$\mathrm{E}[(\mathbf{l}'\mathbf{Az})^r] = (\mathbf{l}'\mathbf{\Sigma}\mathbf{l})^{r/2}\sum_{j=0}^{\lfloor r/2 \rfloor}\binom{r}{2j}\frac{(2j)!}{2^j j!}\frac{2^{r-j}\Gamma(n/2 + r - j)}{\Gamma(n/2)}$$
$$\times \frac{1}{(r-j)!}\sum_{\nu_1=0}^{j}\sum_{\nu_2=0}^{r-2j}(-1)^{\nu_1+\nu_2}\binom{j}{\nu_1}\binom{r-2j}{\nu_2}\mathrm{E}\left[(h_1\mathbf{y}'\mathbf{B}\mathbf{y} + h_2\mathbf{b}'\mathbf{y})^{r-j}\right], \qquad (12)$$

where $h_1 = j/2 - \nu_1$, $h_2 = r/2 - j - \nu_2$, $\mathbf{B} = \mathrm{diag}(0, 1, \cdots, 1)$ *and* $\mathbf{b} = (1, 0, \cdots, 0)'$.

Proof. If we put $p = 1$ at Theorem 2.1, then

$$\mathbf{l}'\mathbf{A}\mathbf{z} \stackrel{\mathrm{d}}{=} (\mathbf{l}'\boldsymbol{\Sigma}\mathbf{l})^{1/2}(\xi y_1 + \sqrt{\xi\eta} z_0),$$

where $\eta = \mathbf{y}_2'\mathbf{y}_2$, $\xi \sim \chi_n^2$, $z_0 \sim N(0,1)$ and

$$\mathbf{y} = \begin{pmatrix} y_1 \\ \mathbf{y}_2 \end{pmatrix} \sim N_k\left(\begin{pmatrix} \mathbf{S}_1\tilde{\boldsymbol{\Sigma}}^{1/2}\tilde{\boldsymbol{\mu}} \\ \mathbf{S}_2\tilde{\boldsymbol{\Sigma}}^{1/2}\tilde{\boldsymbol{\mu}} \end{pmatrix}, \begin{pmatrix} \mathbf{S}_1\tilde{\boldsymbol{\Sigma}}\mathbf{S}_1' & \mathbf{S}_1\tilde{\boldsymbol{\Sigma}}\mathbf{S}_2' \\ \mathbf{S}_2\tilde{\boldsymbol{\Sigma}}\mathbf{S}_1' & \mathbf{S}_2\tilde{\boldsymbol{\Sigma}}\mathbf{S}_2' \end{pmatrix} \right),$$

where $\tilde{\boldsymbol{\mu}} = \boldsymbol{\Omega}^{-1/2}\boldsymbol{\mu}$, $\mathbf{S}_1 = (\mathbf{l}'\boldsymbol{\Sigma}\mathbf{l})^{-1/2}\mathbf{l}'\boldsymbol{\Omega}^{-1/2}\tilde{\boldsymbol{\Sigma}}^{1/2}$, and $\mathbf{S}_2 = (\mathbf{I}_k - \mathbf{S}_1'\mathbf{S}_1)^{1/2}$. In addition, ξ and $\mathbf{z}_0$ are independent of $\mathbf{y}$. The r-th moment of $\mathbf{l}'\mathbf{A}\mathbf{z}$ is computed by

$$\begin{aligned}
\mathrm{E}[(\mathbf{l}'\mathbf{A}\mathbf{z})^r] &= (\mathbf{l}'\boldsymbol{\Sigma}\mathbf{l})^{r/2}\mathrm{E}\left[(\xi y_1 + \sqrt{\xi\eta} z_0)^r\right] = (\mathbf{l}'\boldsymbol{\Sigma}\mathbf{l})^{r/2}\mathrm{E}\left[\sum_{i=0}^{r}\binom{r}{i}(\xi y_1)^{r-i}(\sqrt{\xi\eta} z_0)^i\right] \\
&= (\mathbf{l}'\boldsymbol{\Sigma}\mathbf{l})^{r/2}\sum_{i=0}^{r}\binom{r}{i}\mathrm{E}\left[\xi^{r-i/2}\right]\mathrm{E}\left[y_1^{r-i}\eta^{i/2}\right]\mathrm{E}\left[z_0^i\right], \qquad (13)
\end{aligned}$$

where the second line follows from the fact that ξ and z_0 are independent of $\mathbf{y}$. According to Javed et al. (2021), the even moments of z_0 are given by

$$\mathrm{E}[z_0^{2j}] = \frac{(2j)!}{2^j j!} \text{ for } j \geq 1. \qquad (14)$$

In addition, the s-th moment of a chi-squared variable is given by

$$\mathrm{E}[\xi^s] = \frac{2^s\Gamma(n/2+s)}{\Gamma(n/2)}. \qquad (15)$$

If we put $i = 2j$ in (13), and apply (14) and (15) to (13), then we get

$$\mathrm{E}[(\mathbf{l}'\mathbf{A}\mathbf{z})^r] = (\mathbf{l}'\boldsymbol{\Sigma}\mathbf{l})^{r/2}\sum_{j=0}^{\lfloor r/2\rfloor}\binom{r}{2j}\frac{(2j)!}{2^j j!}\frac{2^{r-j}\Gamma(n/2+r-j)}{\Gamma(n/2)}\mathrm{E}\left[\eta^j y_1^{r-2j}\right], \qquad (16)$$

where the expectation of the right side is taken over the joint distribution of y_1 and $\mathbf{y}_2$. If we apply the identity (2) given by Kan (2008) to (16), then

$$\begin{aligned}
\mathrm{E}\left[\eta^j y_1^{r-2j}\right] &= \frac{1}{(r-j)!}\sum_{\nu_1=0}^{j}\sum_{\nu_2=0}^{r-2j}(-1)^{\nu_1+\nu_2}\binom{j}{\nu_1}\binom{r-2j}{\nu_2}\mathrm{E}\left[(h_1\eta + h_2 y_1)^{r-j}\right], \\
&= \frac{1}{(r-j)!}\sum_{\nu_1=0}^{j}\sum_{\nu_2=0}^{r-2j}(-1)^{\nu_1+\nu_2}\binom{j}{\nu_1}\binom{r-2j}{\nu_2}\mathrm{E}\left[(h_1\mathbf{y}'\mathbf{B}\mathbf{y} + h_2\mathbf{b}'\mathbf{y})^{r-j}\right],
\end{aligned}$$

where $h_1 = j/2 - \nu_1$, $h_2 = r/2 - j - \nu_2$, $\mathbf{B} = \mathrm{diag}(0, 1, \cdots, 1)$ and $\mathbf{b} = (1, 0, \cdots, 0)'$. □

If we put

$$\boldsymbol{\theta} = \begin{pmatrix} \mathbf{S}_1\tilde{\boldsymbol{\Sigma}}^{1/2}\tilde{\boldsymbol{\mu}} \\ \mathbf{S}_2\tilde{\boldsymbol{\Sigma}}^{1/2}\tilde{\boldsymbol{\mu}} \end{pmatrix}, \mathbf{T} = \begin{pmatrix} \mathbf{S}_1\tilde{\boldsymbol{\Sigma}}\mathbf{S}_1' & \mathbf{S}_1\tilde{\boldsymbol{\Sigma}}\mathbf{S}_2' \\ \mathbf{S}_2\tilde{\boldsymbol{\Sigma}}\mathbf{S}_1' & \mathbf{S}_2\tilde{\boldsymbol{\Sigma}}\mathbf{S}_2' \end{pmatrix},$$

then Theorem 3.2b.3 of Mathai and Provost (1992) shows

$$\begin{aligned}
&\mathrm{E}[(h_1\mathbf{y}'\mathbf{B}\mathbf{y}+h_2\mathbf{b}'\mathbf{x})^{r-j}] \\
&=\sum_{t_1=0}^{r-j-1}\binom{r-j-1}{t_1}g^{(t-1-t_1)}\sum_{t_2=0}^{t_1-1}\binom{t_1-1}{t_2}g^{(t_1-1-t_2)}\cdots
\end{aligned}\tag{17}$$

with

$$g^{(m)}=\begin{cases}\dfrac{1}{2}m!\displaystyle\sum_{i=1}^{k}(2h_1\lambda_i)^{m+1}+\dfrac{(m+1)!}{2}\sum_{i=1}^{k}b_i^{*2}(2h_1\lambda_i)^{m-1}, & m\geq 1,\\ h_1\displaystyle\sum_{i=1}^{k}\lambda_i+(h_2\mathbf{b}'\boldsymbol{\theta}+h_1\boldsymbol{\theta}'\mathbf{B}\boldsymbol{\theta}) & m=0,\end{cases}\tag{18}$$

where $\mathbf{P}'\mathbf{T}^{1/2}\mathbf{B}\mathbf{T}^{1/2}\mathbf{P}=\mathrm{diag}(\lambda_1,\cdots,\lambda_k)$ for an orthogonal matrix $\mathbf{P}$, and $\mathbf{P}'(\mathbf{T}^{1/2}\mathbf{b}h_2+2\mathbf{T}^{1/2}\mathbf{B}\boldsymbol{\theta}h_1)=\mathbf{b}^*=(b_1^*,\cdots,b_k^*)'$.

Using (12) and (17), we can deliver the first four moments of $\mathbf{l}'\mathbf{A}\mathbf{z}$ in the following corollary.

Corollary 2.4. *Assume the same conditions as Theorem* 2.3. *Then, the mean, variance, skewness and kurtosis of* $\mathbf{l}'\mathbf{A}\mathbf{z}$ *are given by*

$$\begin{aligned}
\mathrm{E}[\mathbf{l}'\mathbf{A}\mathbf{z}]&=na,\\
\mathrm{V}[\mathbf{l}'\mathbf{A}\mathbf{z}]&=n[a^2+(n+1)f+bs],\\
\mathrm{Skewness}[\mathbf{l}'\mathbf{A}\mathbf{z}]&=\frac{2\{a^3+3[bs+(n+1)f]a+3bd(n+2)\}}{\sqrt{n}[a^2+(n+1)f+bs]^{3/2}},\\
\mathrm{Kurtosis}[\mathbf{l}'\mathbf{A}\mathbf{z}]&=\frac{3}{n[a^2+(n+1)f+bs]^2}\\
&\quad\times\{(n+2)a^4+2[bs+(n+1)f](n+6)a^2+24bd(n+2)a\\
&\quad+(n+2)[(s^2+4g+2t)b^2+2(n+3)(fs+2h)b+(n+1)(n+3)f^2]\},
\end{aligned}$$

where $a=\mathbf{l}'\boldsymbol{\Sigma}\boldsymbol{\mu}$, $b=\mathbf{l}'\boldsymbol{\Sigma}\mathbf{l}$, $d=\mathbf{l}'\boldsymbol{\Sigma}\boldsymbol{\Omega}\boldsymbol{\Sigma}\boldsymbol{\mu}$, $f=\mathbf{l}'\boldsymbol{\Sigma}\boldsymbol{\Omega}\boldsymbol{\Sigma}\mathbf{l}$, $g=\boldsymbol{\mu}'\boldsymbol{\Sigma}\boldsymbol{\Omega}\boldsymbol{\Sigma}\boldsymbol{\mu}$, $h=\mathbf{l}'\boldsymbol{\Sigma}\boldsymbol{\Omega}\boldsymbol{\Sigma}\boldsymbol{\Omega}\boldsymbol{\Sigma}\mathbf{l}$, $s=\boldsymbol{\mu}'\boldsymbol{\Sigma}\boldsymbol{\mu}+\mathrm{tr}(\boldsymbol{\Sigma}\boldsymbol{\Omega})$ *and* $t=\mathrm{tr}(\boldsymbol{\Sigma}\boldsymbol{\Omega}\boldsymbol{\Sigma}\boldsymbol{\Omega})$.

The proof of this corollary is given in Appendix.

If we put $\boldsymbol{\Sigma}=\mathbf{I}_k$, $\boldsymbol{\Omega}=\lambda\mathbf{I}_k$ with $\lambda>0$, and $\boldsymbol{\mu}=\mathbf{0}$ at Theorem 2.3, the general moment of $\mathbf{l}'\mathbf{A}\mathbf{z}$ is simply expressed.

Corollary 2.5. *Let* $\mathbf{A}\sim W_k(n,\mathbf{I}_k)$ *and* $\mathbf{z}\sim N_k(\mathbf{0},\lambda\mathbf{I}_k)$ *with* $\lambda>0$, *where* $\mathbf{A}$ *and* $\mathbf{z}$ *are independent. Then, the higher order moment of* $\mathbf{l}'\mathbf{A}\mathbf{z}$ *is given by*

$$\mathrm{E}[(\mathbf{l}'\mathbf{A}\mathbf{z})^{2m}]=(\mathbf{l}'\mathbf{l})^m(2\lambda)^m\sum_{j=0}^{m}\frac{(2m)!}{j!(m-j)!}\frac{\Gamma(n/2+2m-j)}{\Gamma(n/2)}\frac{\Gamma((k-1)/2+j)}{\Gamma((k-1)/2)}.$$

Proof. If we put $\mathbf{\Sigma} = \mathbf{I}_k$, $\mathbf{\Omega} = \lambda \mathbf{I}_k$ and $\boldsymbol{\mu} = \mathbf{0}$, (16) becomes

$$\mathrm{E}[(\mathbf{l}'\mathbf{Az})^r] = (\mathbf{l}'\mathbf{l})^{r/2} \sum_{j=0}^{\lfloor r/2 \rfloor} \binom{r}{2j} \frac{(2j)!}{2^j j!} \frac{2^{r-j}\Gamma(n/2+r-j)}{\Gamma(n/2)} \mathrm{E}[\eta^j]\mathrm{E}[y_1^{r-2j}]. \tag{19}$$

Since $\mathrm{E}[y_1^{r-2j}] = 0$ when r is odd, we put $r = 2m$ in (19). Then,

$$\mathrm{E}[(\mathbf{l}'\mathbf{Az})^{2m}] = (\mathbf{l}'\mathbf{l})^m \sum_{j=0}^{m} \binom{2m}{2j} \frac{(2j)!}{2^j j!} \frac{2^{2m-j}\Gamma(n/2+2m-j)}{\Gamma(n/2)} \mathrm{E}[\eta^j]\mathrm{E}[y_1^{2m-2j}],$$

Since (10) given by Bodnar et al. (2013) shows that $\lambda^{-1}\eta \sim \chi^2_{k-1}$ and $y_1 \sim N(0,1)$,

$$\begin{aligned}
\mathrm{E}[(\mathbf{l}'\mathbf{Az})^{2m}] &= (\mathbf{l}'\mathbf{l})^m \sum_{j=0}^{m} \binom{2m}{2j} \frac{(2j)!}{2^j j!} \frac{2^{2m-j}\Gamma(n/2+2m-j)}{\Gamma(n/2)} \\
&\quad \times \lambda^j \frac{2^j\Gamma((k-1)/2+j)}{\Gamma((k-1)/2)} \lambda^{m-j} \frac{(2(m-j))!}{2^{m-j}(m-j)!} \\
&= (\mathbf{l}'\mathbf{l})^m \sum_{j=0}^{m} \binom{2m}{2j} \frac{(2j)!}{j!} \frac{(2\lambda)^m\Gamma(n/2+2m-j)}{\Gamma(n/2)} \\
&\quad \times \frac{\Gamma((k-1)/2+j)}{\Gamma((k-1)/2)} \frac{(2(m-j))!}{(m-j)!} \\
&= (\mathbf{l}'\mathbf{l})^m \sum_{j=0}^{m} \frac{(2m)!}{(2j)!(2m-2j)!} \frac{(2j)!}{j!} \frac{(2\lambda)^m\Gamma(n/2+2m-j)}{\Gamma(n/2)} \\
&\quad \times \frac{\Gamma((k-1)/2+j)}{\Gamma((k-1)/2)} \frac{(2(m-j))!}{(m-j)!} \\
&= (\mathbf{l}'\mathbf{l})^m \sum_{j=0}^{m} \frac{(2m)!}{j!} \frac{(2\lambda)^m\Gamma(n/2+2m-j)}{\Gamma(n/2)} \frac{\Gamma((k-1)/2+j)}{\Gamma((k-1)/2)} \frac{1}{(m-j)!} \\
&= (\mathbf{l}'\mathbf{l})^m (2\lambda)^m \sum_{j=0}^{m} \frac{(2m)!}{j!(m-j)!} \frac{\Gamma(n/2+2m-j)}{\Gamma(n/2)} \frac{\Gamma((k-1)/2+j)}{\Gamma((k-1)/2)}.
\end{aligned}$$

□

In particular, all odd moments are 0, and

$$\begin{aligned}
\mathrm{V}[\mathbf{l}'\mathbf{Az}] &= n(n+k+1)\mathbf{l}'\mathbf{l}, \\
\mathrm{Kurtosis}[\mathbf{l}'\mathbf{Az}] &= 3\left(1+\frac{2}{n}\right)\frac{k(k+1)+(n+3)(n+2k+5)}{(n+k+1)^2}.
\end{aligned}$$

These results suggest that if $\mathbf{A} \sim W_k(n, \mathbf{I}_k)$ and $\mathbf{z} \sim N_k(\mathbf{0}, \lambda\mathbf{I}_k)$, then the distribution of any linear combination of the elements of $\mathbf{Az}$ is not skewed and has the same kurtosis.

In the next corollary, we also provide the first and second moments of $\mathbf{Az}$.

Corollary 2.6. *Assume the same conditions as Theorem* 2.1. *The mean vector and covariance matrix of* $\mathbf{Az}$ *are given by*

$$\mathrm{E}[\mathbf{Az}] = n\boldsymbol{\Sigma}\boldsymbol{\mu},$$
$$\mathrm{E}[(\mathbf{Az} - \mathrm{E}(\mathbf{Az}))(\mathbf{Az} - \mathrm{E}(\mathbf{Az}))'] = n\{\boldsymbol{\Sigma}\boldsymbol{\mu}\boldsymbol{\mu}'\boldsymbol{\Sigma} + (n+1)\boldsymbol{\Sigma}\boldsymbol{\Omega}\boldsymbol{\Sigma} + (\boldsymbol{\mu}'\boldsymbol{\Sigma}\boldsymbol{\mu} + \mathrm{tr}(\boldsymbol{\Sigma}\boldsymbol{\Omega}))\boldsymbol{\Sigma}\}.$$

Proof. Since the variance of $\mathbf{l}'\mathbf{Az}$ is written by

$$n\mathbf{l}'[\boldsymbol{\Sigma}\boldsymbol{\mu}\boldsymbol{\mu}'\boldsymbol{\Sigma} + (n+1)\boldsymbol{\Sigma}\boldsymbol{\Omega}\boldsymbol{\Sigma} + (\boldsymbol{\mu}'\boldsymbol{\Sigma}\boldsymbol{\mu} + \mathrm{tr}(\boldsymbol{\Sigma}\boldsymbol{\Omega}))\boldsymbol{\Sigma}]\mathbf{l},$$

the desired results follow immediately. □

Remark 2.7. The derivation of the higher order moments of $\mathbf{l}'\mathbf{Az}$ is based on the stochastic representation given in Theorem 2.1. Since $\mathbf{A}$ and $\mathbf{z}$ are independent, the conditional distribution of $\mathbf{l}'\mathbf{Az}$ given $\mathbf{A}$ is $N(\mathbf{l}'\mathbf{A}\boldsymbol{\mu}, \mathbf{l}'\mathbf{A}\boldsymbol{\Omega}\mathbf{A}\mathbf{l})$. If we can evaluate the higher order moments of $\mathbf{A}$, we can also evaluate the higher order moments of $\mathbf{l}'\mathbf{Az}$. In Appendix, we derive the general formulae for the higher order moments of a Wishart matrix, and apply the formulae to obtain the first four moments of $\mathbf{l}'\mathbf{Az}$.

2.3 Numerical Illustration

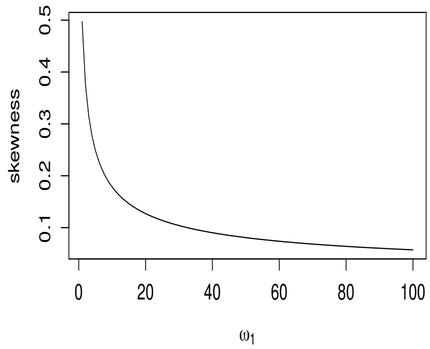
In the previous section, we obtain the stochastic representation and density function of the product of a Wishart matrix and a normal vector. Since the density function generally involves the multi-dimensional integrals, it is difficult to understand the characteristics of the distribution of the product. To capture the characteristics of the distribution of the product numerically, in this section, we compute the moments of $\mathbf{l}'\mathbf{Az}$ where $\mathbf{A} \sim W_k(n, \boldsymbol{\Sigma})$ and $\mathbf{z} \sim N_k(\boldsymbol{\mu}, \boldsymbol{\Omega})$. Throughout this section, the singularity of a Wishart matrix is permitted. Using Theorem 2 of Bodnar et al. (2014) and Theorem 3.2.5 of Muirhead (1982), the stochastic representation of $\mathbf{l}'\mathbf{Az}$ becomes

$$\mathbf{l}'\mathbf{Az} \overset{\mathrm{d}}{=} (\boldsymbol{\Sigma}^{1/2}\mathbf{l})'\mathbf{A}_1\mathbf{z}_1, \tag{20}$$

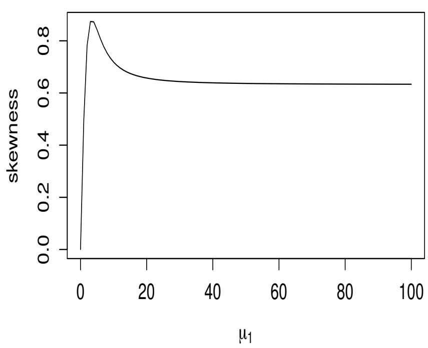
where $\mathbf{A}_1 \sim W_k(n, \mathbf{I}_k)$ and $\mathbf{z}_1 \sim N_k(\boldsymbol{\Sigma}^{1/2}\boldsymbol{\mu}, \boldsymbol{\Sigma}^{1/2}\boldsymbol{\Omega}\boldsymbol{\Sigma}^{1/2})$. If we diagonalize $\boldsymbol{\Sigma}^{1/2}\boldsymbol{\Omega}\boldsymbol{\Sigma}^{1/2}$ as $\mathbf{H}'\boldsymbol{\Sigma}^{1/2}\boldsymbol{\Omega}\boldsymbol{\Sigma}^{1/2}\mathbf{H} = \boldsymbol{\Lambda}$, where $\boldsymbol{\Lambda}$ is the diagonal matrix with positive eigenvalues in diagonal elements and $\mathbf{H}$ is an orthogonal matrix, then the right side of (20) becomes

$$(\boldsymbol{\Sigma}^{1/2}\mathbf{l})'\mathbf{A}_1\mathbf{z}_1 \overset{\mathrm{d}}{=} (\mathbf{H}'\boldsymbol{\Sigma}^{1/2}\mathbf{l})'\mathbf{H}'\mathbf{A}_1\mathbf{H}\mathbf{z}_2,$$

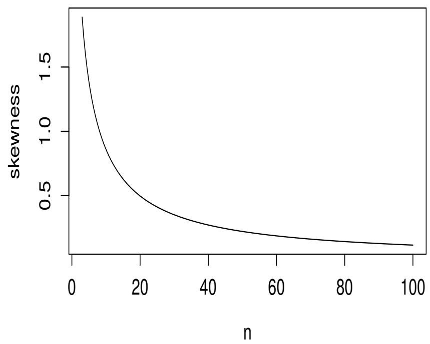
where $\mathbf{z}_2 \sim N_k(\mathbf{H}'\boldsymbol{\Sigma}^{1/2}\boldsymbol{\mu}, \boldsymbol{\Lambda})$. Since $\mathbf{H}'\mathbf{A}_1\mathbf{H} \sim W_k(n, \mathbf{I}_k)$, the distribution of $\mathbf{l}'\mathbf{Az}$ is identical to that of $\mathbf{l}_1'\mathbf{A}_2\mathbf{z}_2$, where $\mathbf{l}_1 = \mathbf{H}'\boldsymbol{\Sigma}^{1/2}\mathbf{l}$, $\mathbf{A}_2 \sim W_k(n, \mathbf{I}_k)$ and $\mathbf{z}_2 \sim N_k(\mathbf{H}'\boldsymbol{\Sigma}^{1/2}\boldsymbol{\mu}, \boldsymbol{\Lambda})$. Here, we assume that $\mathbf{A} \sim W_k(n, \mathbf{I}_k)$ and $\mathbf{z} \sim N_k(\boldsymbol{\mu}, \boldsymbol{\Omega})$, where $\boldsymbol{\mu} = (\mu_1, 0, \cdots, 0)'$ and $\boldsymbol{\Omega} = \mathrm{diag}(\omega_1, 1, \cdots, 1)$.



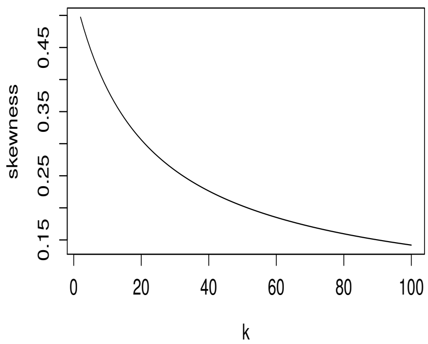
(a) $n = 20$, $k = 2$, $\mu_1 = 1$



(b) $n = 20$, $k = 2$, $\omega_1 = 1$



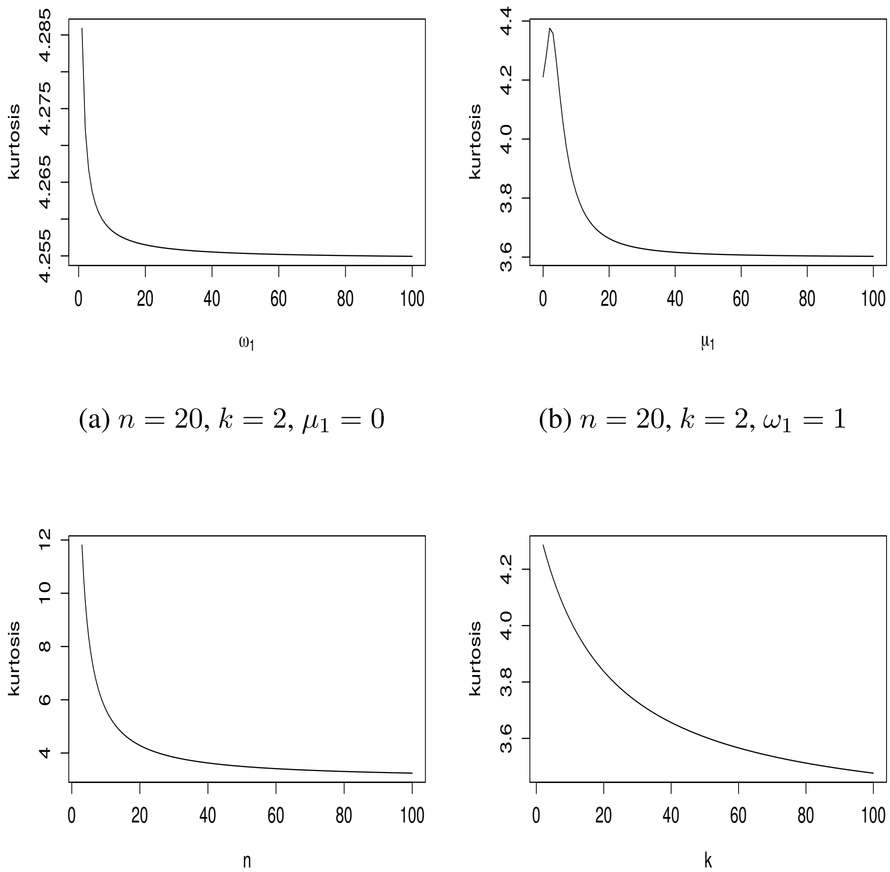
(c) $k = 2$, $\mu_1 = 1, \omega_1 = 1$



(d) $n = 20$, $\mu_1 = 1, \omega_1 = 1$

Figure 2: Plots of the skewness of the first element of $\mathbf{Az}$ against the each parameter of $\mathbf{A}$ and $\mathbf{z}$, where $\mathbf{A} \sim W_k(n, \mathbf{I}_k)$ and $\mathbf{z} \sim N_k(\boldsymbol{\mu}, \boldsymbol{\Omega})$ with $\boldsymbol{\mu} = (\mu_1, 0, \cdots, 0)'$ and $\boldsymbol{\Omega} = \mathrm{diag}(\omega_1, 1, \cdots, 1)$.

Figure 2 and 3 plot the skewness and kurtosis, respectively, of $\mathbf{l}'\mathbf{Az}$ as the function of the each parameter of $\mathbf{A}$ and $\mathbf{z}$; i.e., the first element of $\boldsymbol{\mu}$, the first diagonal element of $\boldsymbol{\Omega}$, degrees of freedom n and dimension k. The computation of the moments of $\mathbf{l}'\mathbf{Az}$ is based on Theorem 2.4. We observe from Figure 2 and 3 that the distribution of $\mathbf{l}'\mathbf{Az}$ is generally skewed and heavy tailed. Whereas the skewness decreases with an increase in the first diagonal element of $\boldsymbol{\Omega}$, the kurtosis may not be changed very much. For a large value of n or k, the skewness and kurtosis are close to 0 and 3, respectively. These results for a large n seem natural when it is noted that the $\mathbf{A}/n$ converges to $\boldsymbol{\Sigma}$. If the first element of $\boldsymbol{\mu}$ becomes large, it appears that the skewness and kurtosis converge to a constant value at around 0.65 and 3.6, respectively.



(a) $n = 20$, $k = 2$, $\mu_1 = 0$ (b) $n = 20$, $k = 2$, $\omega_1 = 1$

(c) $k = 2$, $\mu_1 = 0, \omega_1 = 1$ (d) $n = 20$, $\mu_1 = 0, \omega_1 = 1$

Figure 3: Plots of the kurtosis of the first element of $\mathbf{Az}$ against the each parameter of $\mathbf{A}$ and $\mathbf{z}$ where $\mathbf{A} \sim W_k(n, \mathbf{I}_k)$ and $\mathbf{z} \sim N_k(\boldsymbol{\mu}, \boldsymbol{\Omega})$ with $\boldsymbol{\mu} = (\mu_1, 0, \cdots, 0)'$ and $\boldsymbol{\Omega} = \mathrm{diag}(\omega_1, 1, \cdots, 1)$.

2.4 Finite Sample Performance

In the previous section, we have derived the general expression of the higher order moment of the product of a (singular) Wishart matrix and a normal vector. The aim of this section is to investigate the applicability of the Edgeworth type expansion provided in Theorem 3.2.2 of Kollo and von Rosen (2005). Concerning Edgeworth expansions, Javed et al. (2021) also discussed Edgeworth expansion of random sum of independent and identically distributed random vectors. To achieve this aim, we compare the kernel density estimator of the product with the Edgeworth type expansion of the product. In the statistical software R, kernel density estimation can be implemented by the command "`density()`". The comparison is made for $p = 1$, $\mathbf{\Sigma} = \mathbf{\Omega} = \mathbf{I}_k$, $\mathbf{l} = (1/n, 0, \cdots, 0)'$, and $\boldsymbol{\mu} = (1, \cdots, 1)'$. The kernel density estimator of $\mathbf{l}'\mathbf{Az}$ is computed based on $N = 10^4$ independent and standardized realizations from the stochastic representation (8) of Bodnar et al. (2013). To obtain $N = 10^4$ realizations of $\mathbf{l}'\mathbf{Az}$, we use the following algorithm:

(a) generate independently $z_0 \sim N(0,1)$, $\xi \sim \chi^2_n$, $y_1 \sim N(\mathbf{l}'\boldsymbol{\mu}/\sqrt{\mathbf{l}'\mathbf{l}}, 1)$, and $\eta \sim \chi^2_{k-1,\delta^2}$ with $\delta^2 = \boldsymbol{\mu}'\boldsymbol{\mu} - \mathbf{l}'\boldsymbol{\mu}\boldsymbol{\mu}'\mathbf{l}/\mathbf{l}'\mathbf{l}$;

(b) compute $\mathbf{l}'\mathbf{Az} = \sqrt{\mathbf{l}'\mathbf{l}}(\xi y_1 + \sqrt{\xi\eta} z_0)$;

(c) repeat (a)-(b) 10^4 times.

From Theorem 3.2.2 of Kollo and von Rosen (2005), the Edgeworth type expansion of the standardized $\mathbf{l}'\mathbf{Az}$ is given by

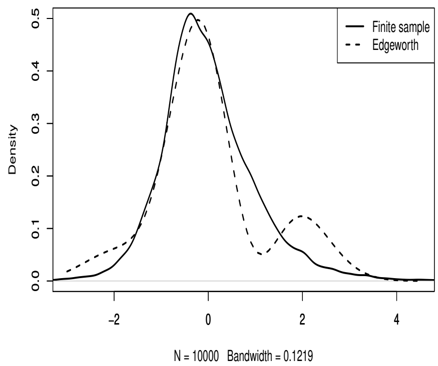
$$f_{\mathbf{l}'\mathbf{Az}}(x) = \phi(x)\left\{1 + c_3\frac{x^3 - 3x}{6} + (c_4 - 3)\frac{x^4 - 4x^2 + 1}{24} + \cdots\right\}, \tag{21}$$

where $\phi(x)$ is the standard normal density, $c_3 = \text{Skewness}[\mathbf{l}'\mathbf{Az}]$ and $c_4 = \text{Kurtosis}[\mathbf{l}'\mathbf{Az}]$.

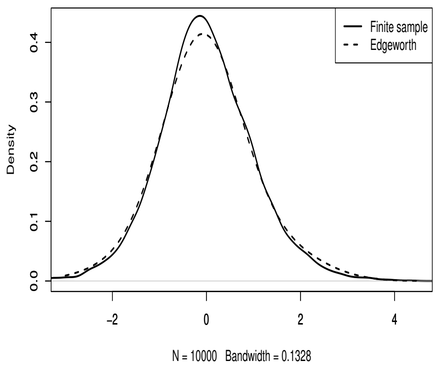
In Figure 4, we present the kernel density estimators and Edgeworth type expansions of $\mathbf{l}'\mathbf{Az}$ for several values of k and n. The finite sample distributions are shown as solid lines, while the Edgeworth type expansions are dashed lines. For small n and k, the performance of the Edgeworth type expansions could be poor, which indicates that we should include the moments of order higher than four in the Edgeworth type expansions given by (21). However, if we increase the dimension k, the approximation for small n is greatly improved. This is clear by comparing Figure 4a and 4b. In addition, we observe from Figure 4c and 4d that the Edgeworth type expansions provide a good approximation for moderately large n.

2.5 Summary

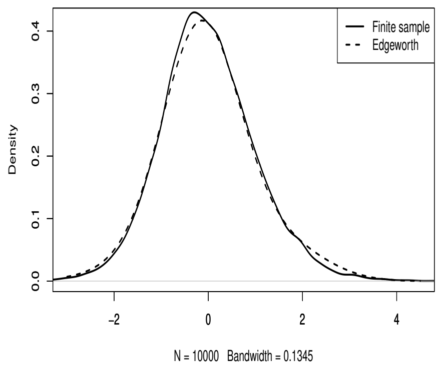
Functions of a Wishart matrix and a normal vector appear in a variety of multivariate statistical methods, such as principal components analysis, multiple comparison, and



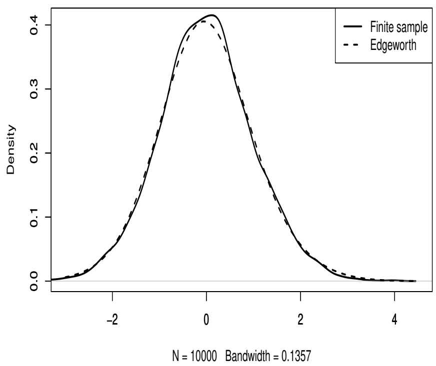
(a) $n = 10$, $k = 2$



(b) $n = 10$, $k = 50$



(c) $n = 30$, $k = 2$



(d) $n = 30$, $k = 50$

Figure 4: Edgeworth type expansion and the kernel density estimator based on the finite sample for the standardized $\mathbf{l}'\mathbf{A}\mathbf{z}$, where $\mathbf{l}' = (1/n, 0, \cdots, 0)$, $\mathbf{A} \sim W_k(n, \mathbf{I}_k)$, and $\mathbf{z} \sim N_k((1, \cdots, 1)', \mathbf{I}_k)$ for $k \in \{2, 50\}$ and $n \in \{10, 30\}$

discriminant analysis under multivariate normality. Over the past ten years, there has been an increase in the number of studies on the product of a Wishart matrix and a normal vector. Although the assumption that $\mathbf{A} \sim W_k(n, \boldsymbol{\Sigma})$ and $\mathbf{z} \sim N_k(\boldsymbol{\mu}, \lambda\boldsymbol{\Sigma})$ for any positive real number λ has been employed in the literature, assuming $\mathbf{A} \sim W_k(n, \boldsymbol{\Sigma})$ and $\mathbf{z} \sim N_k(\boldsymbol{\mu}, \boldsymbol{\Omega})$ appears natural from the viewpoint of Bayesian statistics. We referred to the assumption $\mathbf{A} \sim W_k(n, \boldsymbol{\Sigma})$ and $\mathbf{z} \sim N_k(\boldsymbol{\mu}, \boldsymbol{\Omega})$, which are independent, as the uncommon covariance structure. In addition, there has been scant research on moments of the product. Thus, the purpose of this study was to present the stochastic representation, density function and exact moments under the uncommon covariance structure. We present the stochastic representation of the product, which is computationally efficient. It is applied to construct density function and to derive the higher order moments of the product. As detailed in Section 2.3, the distribution of $\mathbf{l}'\mathbf{A}\mathbf{z}$ is generally skewed and heavy-tailed. If ω_1, which is an eigenvalue of $\boldsymbol{\Sigma}\boldsymbol{\Omega}$, is large, then the distribution of $\mathbf{l}'\mathbf{A}\mathbf{z}$ is less skewed. When either dimension or degrees of freedom n of a Wishart matrix is large, the skewness and kurtosis of $\mathbf{l}'\mathbf{A}\mathbf{z}$ is close to those of a normal distribution. Section 2.4 provide the Edgeworth type expansions for the distribution of $\mathbf{l}'\mathbf{A}\mathbf{z}$, and compare the approximation with the kernel density estimators. A good performance of the approximations is documented for moderately large n.

Appendix

Proof of Lemma 2.3. We only show the proof for $r = 4$ since this proof can be applied to the proofs for $r = 1, 2, 3$. If we put $r = 4$ in (12), then $\mathrm{E}[(\mathbf{l}'\mathbf{A}\mathbf{z})^4]/(\mathbf{l}'\boldsymbol{\Sigma}\mathbf{l})^2 =$

$$\frac{2^4\Gamma\left(\frac{n}{2}+4\right)}{\Gamma\left(\frac{n}{2}\right)}\frac{1}{4!}\sum_{\nu_2=0}^{4}(-1)^{\nu_2}\binom{4}{\nu_2}\mathrm{E}\left[(h_1\mathbf{y}'\mathbf{B}\mathbf{y}+h_2\mathbf{b}'\mathbf{y})^4\right] \tag{22}$$

$$+6\frac{2^3\Gamma\left(\frac{n}{2}+3\right)}{\Gamma\left(\frac{n}{2}\right)}\sum_{\nu_1=0}^{1}\sum_{\nu_2=0}^{2}(-1)^{\nu_1+\nu_2}\binom{2}{\nu_2}\mathrm{E}\left[(h_1\mathbf{y}'\mathbf{B}\mathbf{y}+h_2\mathbf{b}'\mathbf{y})^3\right] \tag{23}$$

$$+3\frac{2^2\Gamma\left(\frac{n}{2}+2\right)}{\Gamma\left(\frac{n}{2}\right)}\frac{1}{2!}\sum_{\nu_1=0}^{2}(-1)^{\nu_1}\binom{2}{\nu_1}\mathrm{E}\left[(h_1\mathbf{y}'\mathbf{B}\mathbf{y}+h_2\mathbf{b}'\mathbf{y})^2\right], \tag{24}$$

where $h_1 = j/2 - \nu_1$ and $h_2 = 2 - j - \nu_2$.

To compute the expectation in (22), we put $r = 4$ and $j = 0$ at (17), and obtain

$$\mathrm{E}\left[(h_1\mathbf{y}'\mathbf{B}\mathbf{y}+h_2\mathbf{b}'\mathbf{y})^4\right] = 6g^{(1)}[g^{(0)}]^2 + 3[g^{(1)}]^2 + [g^{(0)}]^4,$$

with

$$g^{(1)} = \sum_{i=1}^{k} b_i^2 = (2-\nu_2)^2\frac{\mathbf{l}'\boldsymbol{\Sigma}\boldsymbol{\Omega}\boldsymbol{\Sigma}}{\mathbf{l}'\boldsymbol{\Sigma}\mathbf{l}} = (2-\nu_2)^2\frac{f}{b},$$

$$g^{(0)} = (2-\nu_2)\frac{\mathbf{l}'\boldsymbol{\Sigma}\boldsymbol{\mu}}{\sqrt{\mathbf{l}'\boldsymbol{\Sigma}\mathbf{l}}},$$

where we have used the (18). Hence (22) becomes

$$\frac{2^4\Gamma\left(\frac{n}{2}+4\right)}{\Gamma\left(\frac{n}{2}\right)}\frac{1}{4!}\sum_{\nu_2=0}^{4}(-1)^{\nu_2}\binom{4}{\nu_2}(2-\nu_2)^4\left[6\frac{fa^2}{b^2}+3\frac{f^2}{b^2}+\frac{a^4}{b^2}\right].$$

To compute the expectation in (23), we put $r=4$ and $j=1$ at (17), and obtain

$$\mathrm{E}\left[(h_1\mathbf{y}'\mathbf{B}\mathbf{y}+h_2\mathbf{b}'\mathbf{y})^3\right] = g^{(2)} + 3g^{(1)}g^{(0)} + [g^{(0)}]^3,$$

where

$$\begin{aligned}
g^{(2)} &= \sum_{i=1}^{k}(2h_1\lambda_i)^3 + 3\sum_{i=1}^{k} b_i^2(2h_1\lambda_i)\\
&= 8h_1^3\,\mathrm{tr}[\tilde{\boldsymbol{\Sigma}}(\mathbf{I}_k - Q)]^3 + 6h_1h_2^2\mathbf{a}'\mathbf{TATa} + 24h_1^2h_2\mathbf{a}'\mathbf{TATA}\boldsymbol{\theta} + 24h_1^3\boldsymbol{\theta}\mathbf{ATATA}\boldsymbol{\theta},\\
g^{(1)} &= \frac{1}{2}\sum_{i=1}^{k}(2h_1\lambda_i)^2 + \sum_{i=1}^{k} b_i^2\\
&= 2h_1^2\,\mathrm{tr}(\mathbf{T}^{1/2}\mathbf{A}\mathbf{T}^{1/2})^3 + (2h_1\mathbf{T}^{1/2}\mathbf{A}\boldsymbol{\theta} + h_2\mathbf{T}^{1/2}\mathbf{a})'(2h_1\mathbf{T}^{1/2}\mathbf{A}\boldsymbol{\theta} + h_2\mathbf{T}^{1/2}\mathbf{a})\\
&= 2h_1^2\left[t - \frac{2h}{b} + \frac{f^2}{b^2} + 2g - \frac{4ab}{b} + \frac{2fa^2}{b^2}\right] + 4h_1h_2\left[\frac{d}{\sqrt{b}} - \frac{fa}{b^{3/2}}\right] + h_2^2\frac{f}{b},\\
g^{(0)} &= h_1\left[s + c - \frac{f+a^2}{b}\right] + h_2\frac{a}{\sqrt{b}},
\end{aligned}$$

where $h_1 = 1/2 - \nu_1$ and $h_2 = 1-\nu_2$.

Although the terms h_1^3, $h_1^2h_2$, $h_1h_2^2$, h_2^3 appear in (23), only the term $h_1h_2^2$ remains. Hence, it is enough to compute the term $\mathbf{a}'\mathbf{TATa}$ in the above $g^{(2)}$ and obtain

$$\mathbf{a}'\mathbf{TATa} = \mathbf{S}_1\tilde{\boldsymbol{\Sigma}}\mathbf{S}_2'\mathbf{S}_2\tilde{\boldsymbol{\Sigma}}\mathbf{S}_1' = \frac{1}{b}\left(h - \frac{f^2}{b}\right).$$

Hence the equation (23) becomes

$$\frac{1}{6}\left[\frac{12}{b} + \frac{6f}{b}\left(s+c-\frac{f+a^2}{b}\right) + \frac{24a}{b}\left(d - \frac{fa}{b}\right) + \frac{6a^2}{b}\left(s+c-\frac{f+a^2}{b}\right)\right].$$

To compute the expectation in (24), we put $r=4$ and $j=2$ at (17), and obtain

$$\mathrm{E}\left[(h_1\mathbf{y}'\mathbf{B}\mathbf{y}+h_2\mathbf{b}'\mathbf{y})^2\right] = g^{(1)} + [g^{(0)}]^2,$$

where $h_1 = 1-\nu_1$ and $h_2 = 0$. So (24) becomes

$$3\frac{2^2\Gamma\left(\frac{n}{2}+2\right)}{\Gamma\left(\frac{n}{2}\right)}2\left(t - \frac{2h}{b} + \frac{f^2}{b^2} + 2g - \frac{4ad}{b} + \frac{2fa^2}{b^2}\right) + \left(s+c-\frac{f+a^2}{b}\right)^2.$$

If we substitute the obtained results for the equations (22), (23) and (24), $\mathrm{E}[(\mathbf{l}'\mathbf{Az})^4]$ can be obtained. □

The last part of the Appendix presents the higher order expectations of the elements of a Wishart matrix, which can be applied to derive the higher order moments of the product of a Wishart matrix and a normal vector. Higher order expectations of the elements of a Wishart matrix play an important role in multivariate statistical methods, such as discriminant analysis (see Fujikoshi, 1987; Kubokawa et al., 2013) and a growth curve model (cf. von Rosen, 1979; von Rosen, 1991b). Several different techniques have been proposed to compute the higher order expectations of the elements of a Wishart matrix (cf. Letac and Massam, 2004; von Rosen, 1979, 1991a, 1991b; Haff, 1979a; Gupta and Nagar, 2000). Here, we employed the method as detailed in Theorem 3.3.15 of Gupta and Nagar (2000), which used the definition of a Wishart matrix and the higher order moments of a matrix variate normal distribution. Let $\mathbf{A}$ be a k-dimensional Wishart matrix with n degrees of freedom and covariance matrix $\boldsymbol{\Sigma}$: that is, $\mathbf{A} \sim W_k(n, \boldsymbol{\Sigma})$. We assume that $n \geq k$, implying that the random matrix $\mathbf{A}$ is non-singular. The following third and fourth order moments of a $\mathbf{A}$ are obtained by Letac and Massam (2004):

$$\begin{aligned}
\mathrm{E}[\mathbf{A}^3] &= n\{(n^2+3n+4)\boldsymbol{\Sigma}^2 + 2(n+1)\operatorname{tr}(\boldsymbol{\Sigma})\boldsymbol{\Sigma} + (n+1)\operatorname{tr}(\boldsymbol{\Sigma}^2)\mathbf{I}_k + \operatorname{tr}(\boldsymbol{\Sigma})^2\mathbf{I}_k\}\boldsymbol{\Sigma},\\
\mathrm{E}[\mathbf{A}^4] &= n\{(n^3+6n^2+21n+20)\boldsymbol{\Sigma}^3 + (3n^2+9n+12)\operatorname{tr}(\boldsymbol{\Sigma})\boldsymbol{\Sigma}^2\\
&\quad + (2n^2+5n+5)\operatorname{tr}(\boldsymbol{\Sigma}^2)\boldsymbol{\Sigma} + 3(n+1)\operatorname{tr}(\boldsymbol{\Sigma})^2\boldsymbol{\Sigma} + (n^2+3n+4)\operatorname{tr}(\boldsymbol{\Sigma}^3)\mathbf{I}_k\\
&\quad + 3(n+1)\operatorname{tr}(\boldsymbol{\Sigma})\operatorname{tr}(\boldsymbol{\Sigma}^2)\mathbf{I}_k + \operatorname{tr}(\boldsymbol{\Sigma})^3\mathbf{I}_k\}\boldsymbol{\Sigma}.
\end{aligned}$$

In the following lemma, we derive the expression of $\mathrm{E}(\mathbf{APAQA})$ and $\mathrm{E}(\mathbf{APAQARA})$. Lemma 2.8 is not just used to derive the moments of the product of a Wishart matrix and a normal vector but also to generalize the above result of Letac and Massam (2004).

Lemma 2.8. *Let $\mathbf{A} \sim W_k(n, \boldsymbol{\Sigma})$ with $\boldsymbol{\Sigma} > 0$, and $\mathbf{P}$, $\mathbf{Q}$ and $\mathbf{R}$ be any square constant matrices. Then, the expectations of $\mathbf{APAQA}$ and $\mathbf{APAQARA}$ are given by*

$$\begin{aligned}
\mathrm{E}(\mathbf{APAQA}) &= \boldsymbol{\Sigma}\{(n\mathbf{P}+\mathbf{P}')\mathrm{E}[\mathbf{AQA}] + \mathbf{Q}\mathrm{E}[\operatorname{tr}(\mathbf{AP})\mathbf{A}] + \mathbf{Q}'\mathrm{E}[\mathbf{AP}'\mathbf{A}]\\
&\quad + \operatorname{tr}(\mathbf{P}\mathrm{E}[\mathbf{AQA}])\},\\
\mathrm{E}(\mathbf{APAQARA}) &= \boldsymbol{\Sigma}\{[(n\mathbf{P}+\mathbf{P}')\boldsymbol{\Sigma}(n\mathbf{Q}+\mathbf{Q}') + \mathbf{Q}\boldsymbol{\Sigma}(\mathbf{P}+\mathbf{P}')\\
&\quad + \mathbf{Q}'\boldsymbol{\Sigma}(n\mathbf{P}'+\mathbf{P})]\mathrm{E}[\mathbf{ARA}] + [(n\mathbf{P}+\mathbf{P}')\boldsymbol{\Sigma}\mathbf{R} + \mathbf{R}'\boldsymbol{\Sigma}\mathbf{P}']\mathrm{E}[\operatorname{tr}(\mathbf{AQ})\mathbf{A}]\\
&\quad + [(n\mathbf{P}+\mathbf{P}')\boldsymbol{\Sigma}\mathbf{R}' + \mathbf{R}\boldsymbol{\Sigma}\mathbf{P}']\mathrm{E}[\mathbf{AQ}'\mathbf{A}] + (\mathbf{R}+\mathbf{R}')\boldsymbol{\Sigma}\mathbf{P}\mathrm{E}[\mathbf{AQA}]\\
&\quad + [\mathbf{Q}\boldsymbol{\Sigma}(n\mathbf{R}+\mathbf{R}') + \mathbf{Q}'\boldsymbol{\Sigma}\mathbf{R}]\mathrm{E}[\operatorname{tr}(\mathbf{AP})\mathbf{A}] + [\mathbf{Q}'\boldsymbol{\Sigma}\mathbf{R}' + \mathbf{R}\boldsymbol{\Sigma}\mathbf{Q}]\mathrm{E}[\mathbf{APA}]\\
&\quad + [\mathbf{R}'\boldsymbol{\Sigma}(n\mathbf{Q}'+\mathbf{Q}) + \mathbf{R}\boldsymbol{\Sigma}\mathbf{Q}']\mathrm{E}[\mathbf{AP}'\mathbf{A}]\}\\
&\quad + \boldsymbol{\Sigma}\{\mathbf{Q}\operatorname{tr}(\mathbf{P}\mathrm{E}[\operatorname{tr}(\mathbf{AR})\mathbf{A}]) + \mathbf{Q}'\operatorname{tr}(\mathbf{P}\mathrm{E}[\mathbf{AR}'\mathbf{A}])\\
&\quad + (n\mathbf{R}+\mathbf{R}')\operatorname{tr}(\mathbf{P}\mathrm{E}[\mathbf{AQA}]) + \operatorname{tr}(\mathbf{P}\boldsymbol{\Sigma}(n\mathbf{Q}+\mathbf{Q}')\mathrm{E}[\mathbf{ARA}])\\
&\quad + \operatorname{tr}(\mathbf{P}\boldsymbol{\Sigma}\mathbf{R}'\mathrm{E}[\mathbf{AQ}'\mathbf{A}]) + \operatorname{tr}(\mathbf{P}\boldsymbol{\Sigma}\mathbf{R}\mathrm{E}[\operatorname{tr}(\mathbf{AQ})\mathbf{A}])\\
&\quad + \operatorname{tr}(\mathbf{Q}\mathrm{E}[\mathbf{ARA}])(n\boldsymbol{\Sigma}\mathbf{P} + \boldsymbol{\Sigma}\mathbf{P}' + \operatorname{tr}(\boldsymbol{\Sigma}\mathbf{P}))\}\boldsymbol{\Sigma}.
\end{aligned}$$

Proof of Lemma 2.8. If $\mathbf{y}_1, \cdots, \mathbf{y}_n$ are independently and identically distributed as $N_k(\mathbf{0}, \mathbf{\Sigma})$, then the Wishart matrix $\mathbf{A}$ is expressed as $\mathbf{A} = \sum_{r=1}^{n} \mathbf{y}_r \mathbf{y}_r'$. In addition, the (i, j) element of $\mathbf{A}$ can be written as $a_{ij} = \sum_{r=1}^{n} y_{ir} y_{jr}$, where $\mathbf{y}_t = (y_{1t}, \cdots, y_{kt})'$ for $t = 1, \cdots, n$. Throughout the proof, p_{ij}, q_{ij}, and r_{ij} denote the (i, j) element of $\mathbf{P}$, $\mathbf{Q}$, and $\mathbf{R}$, respectively. The expectation of an (i, j) element of $\mathbf{APAQA}$ is written by

$$\begin{aligned} \mathrm{E}(\mathbf{APAQA}) &= \mathrm{E}\left(\sum_{s=1}^{k} \sum_{l=1}^{k} \sum_{m=1}^{k} \sum_{t=1}^{k} a_{is} p_{sl} a_{lm} q_{mt} a_{tj} \right) \\ &= \sum_{s=1}^{k} \sum_{l=1}^{k} \sum_{m=1}^{k} \sum_{t=1}^{k} p_{sl} q_{mt} \mathrm{E}(a_{is} a_{lm} a_{tj}). \end{aligned} \tag{25}$$

$\mathrm{E}(a_{is} a_{lm} a_{tj})$ is expressed as

$$\begin{aligned} \mathrm{E}(a_{is} a_{lm} a_{tj}) &= \sum_{r_1=1}^{n} \sum_{r_2=1}^{n} \sum_{r_3=1}^{n} \mathrm{E}(y_{ir_1} y_{sr_1} y_{lr_2} y_{mr_2} y_{tr_3} y_{jr_3}) \\ &= \sum_{r=1}^{n} \mathrm{E}(y_{ir} y_{sr} y_{lr} y_{mr} y_{tr} y_{jr}) + \sum_{r=1}^{n} \sum_{r_3 \neq r}^{n} \mathrm{E}(y_{ir} y_{sr} y_{lr} y_{mr} y_{tr_3} y_{jr_3}) \\ &+ \sum_{r=1}^{n} \sum_{r_2 \neq r}^{n} \mathrm{E}(y_{ir} y_{sr} y_{lr_2} y_{mr_2} y_{tr} y_{jr}) + \sum_{r=1}^{n} \sum_{r_1 \neq r}^{n} \mathrm{E}(y_{ir_1} y_{sr_1} y_{lr} y_{mr} y_{tr} y_{jr}) \\ &+ \sum_{r_1=1}^{n} \sum_{r_2 \neq r_1}^{n} \sum_{r_3 \neq r_2 \neq r_1}^{n} \mathrm{E}(y_{ir_1} y_{sr_1} y_{lr_2} y_{mr_2} y_{tr_3} y_{jr_3}). \end{aligned}$$

Since $\mathbf{y}_1, \cdots, \mathbf{y}_n$ are independently and identically distributed as $N_k(\mathbf{0}, \mathbf{\Sigma})$,

$$\begin{aligned} \mathrm{E}(a_{is} a_{lm} a_{tj}) &= n\mathrm{E}(y_{ir} y_{sr} y_{lr} y_{mr} y_{tr} y_{jr}) + n(n-1)\mathrm{E}(y_{ir} y_{sr} y_{lr} y_{mr})\mathrm{E}(y_{tr_3} y_{jr_3}) \\ &+ n(n-1)\mathrm{E}(y_{ir} y_{sr} y_{tr} y_{jr})\mathrm{E}(y_{lr_2} y_{mr_2}) + n(n-1)\mathrm{E}(y_{lr} y_{mr} y_{tr} y_{jr})\mathrm{E}(y_{ir_1} y_{sr_1}) \\ &+ n(n-1)(n-2)\mathrm{E}(y_{ir_1} y_{sr_1})\mathrm{E}(y_{lr_2} y_{mr_2})\mathrm{E}(y_{tr_3} y_{jr_3}). \end{aligned}$$

Let σ_{ij} denote the (i, j) element of $\mathbf{\Sigma}$. Using Isserlis' theorem (cf. Isserlis, 1918),

$$\begin{aligned} \mathrm{E}(a_{is} a_{lm} a_{tj}) &= n(\sigma_{is}\sigma_{lm}\sigma_{tj} + \sigma_{is}\sigma_{lt}\sigma_{mj} + \sigma_{is}\sigma_{lj}\sigma_{mt} + \sigma_{il}\sigma_{sm}\sigma_{tj} + \sigma_{il}\sigma_{st}\sigma_{mj} \\ &+ \sigma_{il}\sigma_{sj}\sigma_{mt} + \sigma_{im}\sigma_{sl}\sigma_{tj} + \sigma_{im}\sigma_{st}\sigma_{lj} + \sigma_{im}\sigma_{sj}\sigma_{lt} + \sigma_{it}\sigma_{sl}\sigma_{mj} \\ &+ \sigma_{it}\sigma_{sm}\sigma_{lj} + \sigma_{it}\sigma_{sj}\sigma_{lm} + \sigma_{ij}\sigma_{sl}\sigma_{mt} + \sigma_{ij}\sigma_{sm}\sigma_{lt} + \sigma_{ij}\sigma_{st}\sigma_{lm}) \\ &+ n(n-1)(\sigma_{is}\sigma_{lm}\sigma_{tj} + \sigma_{il}\sigma_{sm}\sigma_{tj} + \sigma_{im}\sigma_{sl}\sigma_{tj} + \sigma_{is}\sigma_{lm}\sigma_{tj} \\ &+ \sigma_{it}\sigma_{sj}\sigma_{lm} + \sigma_{ij}\sigma_{st}\sigma_{lm} + \sigma_{is}\sigma_{lm}\sigma_{tj} + \sigma_{is}\sigma_{lt}\sigma_{mj} + \sigma_{is}\sigma_{lj}\sigma_{mt}) \\ &+ n(n-1)(n-2)\sigma_{is}\sigma_{lm}\sigma_{tj}. \end{aligned} \tag{26}$$

If we put (26) into (25), then

$$\mathrm{E}(\mathbf{APAQA}) = \sum_{s=1}^{k} \sum_{l=1}^{k} \sum_{m=1}^{k} \sum_{t=1}^{k} [n(\sigma_{is} p_{sl} \sigma_{lm} q_{mt} \sigma_{tj} + \sigma_{is} p_{sl} \sigma_{lt} q_{mt} \sigma_{mj} + \sigma_{mt} q_{mt} \sigma_{is} p_{sl} \sigma_{lj}$$

$$
\begin{aligned}
&+ \sigma_{il}p_{sl}\sigma_{sm}q_{mt}\sigma_{tj} + \sigma_{il}p_{sl}\sigma_{st}q_{mt}\sigma_{mj} + \sigma_{mt}q_{mt}\sigma_{il}p_{sl}\sigma_{sj} + \sigma_{sl}p_{sl}\sigma_{im}q_{mt}\sigma_{tj} \\
&+ \sigma_{im}q_{mt}\sigma_{st}p_{sl}\sigma_{lj} + \sigma_{im}q_{mt}\sigma_{lt}p_{sl}\sigma_{sj} + \sigma_{it}q_{mt}\sigma_{mj}\sigma_{sl}p_{sl} + \sigma_{it}q_{mt}\sigma_{sm}p_{sl}\sigma_{lj} \\
&+ \sigma_{it}q_{mt}\sigma_{lm}p_{sl}\sigma_{sj} + \sigma_{ij}\sigma_{sl}p_{sl}\sigma_{mt}q_{mt} + \sigma_{ij}\sigma_{sm}q_{mt}\sigma_{lt}p_{sl} + \sigma_{ij}\sigma_{st}q_{mt}\sigma_{lm}p_{sl}) \\
&+ n(n-1)(3\sigma_{is}p_{sl}\sigma_{lm}q_{mt}\sigma_{tj} + \sigma_{il}p_{sl}\sigma_{sm}q_{mt}\sigma_{tj} + \sigma_{sl}p_{sl}\sigma_{im}q_{mt}\sigma_{tj} \\
&+ \sigma_{it}q_{mt}\sigma_{lm}p_{sl}\sigma_{sj} + \sigma_{ij}p_{sl}\sigma_{st}q_{mt}\sigma_{lm} + \sigma_{is}p_{sl}\sigma_{lt}q_{mt}\sigma_{mj} + \sigma_{is}p_{sl}\sigma_{lj}\sigma_{mt}q_{mt}) \\
&+ n(n-1)(n-2)\sigma_{is}p_{sl}\sigma_{lm}q_{mt}\sigma_{tj}]
\end{aligned}
$$

$$
\begin{aligned}
&= n(\mathbf{\Sigma P \Sigma Q \Sigma} + \mathbf{\Sigma P \Sigma Q' \Sigma} + \mathbf{\Sigma Q \Sigma P \Sigma} + \mathbf{\Sigma P' \Sigma Q \Sigma} + \mathbf{\Sigma P' \Sigma Q' \Sigma} \\
&+ \operatorname{tr}(\mathbf{\Sigma Q})\mathbf{\Sigma P' \Sigma} + \operatorname{tr}(\mathbf{\Sigma P})\mathbf{\Sigma Q \Sigma} + \mathbf{\Sigma Q \Sigma P \Sigma} + \mathbf{\Sigma Q \Sigma P' \Sigma} + \operatorname{tr}(\mathbf{\Sigma P})\mathbf{\Sigma Q \Sigma} \\
&+ \mathbf{\Sigma Q' \Sigma P \Sigma} + \mathbf{\Sigma Q' \Sigma P' \Sigma} + \operatorname{tr}(\mathbf{\Sigma P})\operatorname{tr}(\mathbf{\Sigma Q})\mathbf{\Sigma} + \operatorname{tr}(\mathbf{\Sigma Q \Sigma P'})\mathbf{\Sigma} \\
&+ \operatorname{tr}(\mathbf{\Sigma Q' \Sigma P'})\mathbf{\Sigma} + n(n-1)(3\mathbf{\Sigma P \Sigma Q \Sigma} + \mathbf{\Sigma P' \Sigma Q \Sigma} + \operatorname{tr}(\mathbf{\Sigma P})\mathbf{\Sigma Q \Sigma} \\
&+ \mathbf{\Sigma Q' \Sigma P' \Sigma} + \mathbf{\Sigma P' \Sigma Q' \Sigma} + \mathbf{\Sigma P \Sigma Q' \Sigma} + \mathbf{\Sigma P \Sigma}\operatorname{tr}(\mathbf{\Sigma Q})) \\
&+ n(n-1)(n-2)\mathbf{\Sigma P \Sigma Q \Sigma}. \qquad (27)
\end{aligned}
$$

If we arrange the terms involving $\mathbf{\Sigma P}$, $\mathbf{\Sigma P'}$, $\mathbf{\Sigma Q}$, $\mathbf{\Sigma Q'}$, and the others, then

$$
\begin{aligned}
\mathrm{E}(\mathbf{APAQA}) = n\{&n\mathbf{\Sigma P}[n\mathbf{\Sigma Q} + \mathbf{\Sigma Q'} + \operatorname{tr}(\mathbf{\Sigma Q})] + \mathbf{\Sigma P'}[n\mathbf{\Sigma Q} + \mathbf{\Sigma Q'} + \operatorname{tr}(\mathbf{\Sigma Q})] \\
&+ \mathbf{\Sigma Q}[\mathbf{\Sigma P} + \mathbf{\Sigma P'} + n\operatorname{tr}(\mathbf{\Sigma P})] + \mathbf{\Sigma Q'}[\mathbf{\Sigma P} + n\mathbf{\Sigma P'} + \operatorname{tr}(\mathbf{\Sigma P})] \\
&+ \operatorname{tr}(\mathbf{\Sigma P}[n\mathbf{\Sigma Q} + \mathbf{\Sigma Q'} + \operatorname{tr}(\mathbf{\Sigma Q})])\}\mathbf{\Sigma}. \qquad (28)
\end{aligned}
$$

From Theorem 3.3.15 of Gupta and Nagar (2000), the above equation becomes

$$
\begin{aligned}
\mathrm{E}(\mathbf{APAQA}) = n\mathbf{\Sigma P}\mathrm{E}[\mathbf{AQA}] + \mathbf{\Sigma P'}\mathrm{E}[\mathbf{AQA}] + \mathbf{\Sigma Q}\mathrm{E}[\operatorname{tr}(\mathbf{AP})\mathbf{A}] \\
+ \mathbf{\Sigma Q'}\mathrm{E}[\mathbf{AP'A}] + \operatorname{tr}(\mathbf{P}\mathrm{E}[\mathbf{AQA}])\mathbf{\Sigma}.
\end{aligned}
$$

The proof for $\mathrm{E}(\mathbf{APAQA})$ is completed.

The expectation of the (i, j) element of $\mathbf{APAQARA}$ is written by

$$
\begin{aligned}
\mathrm{E}(\mathbf{APAQARA}) &= \mathrm{E}\left(\sum_{s=1}^{k}\sum_{l=1}^{k}\sum_{m=1}^{k}\sum_{t=1}^{k}\sum_{u=1}^{k}\sum_{v=1}^{k} a_{is}p_{sl}a_{lm}q_{mt}a_{tu}r_{uv}a_{vj}\right) \\
&= \sum_{s=1}^{k}\sum_{l=1}^{k}\sum_{m=1}^{k}\sum_{t=1}^{k}\sum_{u=1}^{k}\sum_{v=1}^{k} p_{sl}q_{mt}r_{uv}\mathrm{E}(a_{is}a_{lm}a_{tu}a_{vj}). \qquad (29)
\end{aligned}
$$

$\mathrm{E}(a_{is}a_{lm}a_{tu}a_{vj})$ is expressed as

$$
\begin{aligned}
\mathrm{E}(a_{is}a_{lm}a_{tu}a_{vj}) &= \sum_{r_1=1}^{n}\sum_{r_2=1}^{n}\sum_{r_3=1}^{n}\sum_{r_4=1}^{n} \mathrm{E}(y_{ir_1}y_{sr_1}y_{lr_2}y_{mr_2}y_{tr_3}y_{ur_3}y_{vr_4}y_{jr_4}) \\
= \sum_{r=1}^{n}\mathrm{E}(y_{ir}y_{kr}y_{lr}y_{mr}y_{tr}y_{ur}y_{vr}y_{jr}) &+ \sum_{r=1}^{n}\sum_{r_4 \neq r}\mathrm{E}(y_{ir}y_{kr}y_{lr}y_{mr}y_{tr}y_{ur}y_{vr_4}y_{jr_4})
\end{aligned}
$$

$$
\begin{aligned}
&+\sum_{r=1}^{n}\sum_{r_3\neq r}^{n}\mathrm{E}(y_{ir}y_{sr}y_{lr}y_{mr}y_{tr_3}y_{ur_3}y_{vr}y_{jr})+\sum_{r=1}^{n}\sum_{r_2\neq r}^{n}\mathrm{E}(y_{ir}y_{sr}y_{lr_2}y_{mr_2}y_{tr}y_{ur}y_{vr}y_{jr})\\
&+\sum_{r=1}^{n}\sum_{r_1\neq r}^{n}\mathrm{E}(y_{ir_1}y_{sr_1}y_{lr}y_{mr}y_{tr}y_{ur}y_{vr}y_{jr})+\sum_{r=1}^{n}\sum_{w\neq r}^{n}\mathrm{E}(y_{ir}y_{sr}y_{lr}y_{mr}y_{tw}y_{uw}y_{vw}y_{jw})\\
&+\sum_{r=1}^{n}\sum_{w\neq r}^{n}\mathrm{E}(y_{ir}y_{sr}y_{lw}y_{mw}y_{tr}y_{ur}y_{vw}y_{jw})+\sum_{r=1}^{n}\sum_{w\neq r}^{n}\mathrm{E}(y_{ir}y_{sr}y_{lw}y_{mw}y_{tw}y_{uw}y_{vr}y_{jr})\\
&+\sum_{r=1}^{n}\sum_{r_3\neq r}^{n}\sum_{r_4\neq r_3\neq r}^{n}\mathrm{E}(y_{ir}y_{sr}y_{lr}y_{mr}y_{tr_3}y_{ur_3}y_{vr_4}y_{jr_4})\\
&+\sum_{r=1}^{n}\sum_{r_2\neq r}^{n}\sum_{r_4\neq r_2\neq r}^{n}\mathrm{E}(y_{ir}y_{sr}y_{lr_2}y_{mr_2}y_{tr}y_{ur}y_{vr_4}y_{jr_4})\\
&+\sum_{r=1}^{n}\sum_{r_2\neq r}^{n}\sum_{r_3\neq r_2\neq r}^{n}\mathrm{E}(y_{ir}y_{sr}y_{lr_2}y_{mr_2}y_{tr_3}y_{ur_3}y_{vr}y_{jr})\\
&+\sum_{r=1}^{n}\sum_{r_1\neq r}^{n}\sum_{r_4\neq r_1\neq r}^{n}\mathrm{E}(y_{ir_1}y_{sr_1}y_{lr}y_{mr}y_{tr}y_{ur}y_{vr_4}y_{jr_4})\\
&+\sum_{r=1}^{n}\sum_{r_1\neq r}^{n}\sum_{r_3\neq r_1\neq r}^{n}\mathrm{E}(y_{ir_1}y_{sr_1}y_{lr}y_{mr}y_{tr_3}y_{ur_3}y_{vr}y_{jr})\\
&+\sum_{r=1}^{n}\sum_{r_1\neq r}^{n}\sum_{r_2\neq r_1\neq r}^{n}\mathrm{E}(y_{ir_1}y_{sr_1}y_{lr_2}y_{mr_2}y_{tr}y_{ur}y_{vr}y_{jr})\\
&+\sum_{r_1\neq r_2\neq r_3\neq r_4}^{n}\mathrm{E}(y_{ir_1}y_{sr_1}y_{lr_2}y_{mr_2}y_{tr_3}y_{ur_3}y_{vr_4}y_{jr_4}). \qquad (30)
\end{aligned}
$$

In (30), there are many expectations to be evaluated. Based on Isserlis' theorem, these expectations are expressed only by elements of Σ like (26). If we substitute (30) for (29), we obtain

$$
\begin{aligned}
\mathrm{E}(\mathbf{APAQARA}) &= n\{n^3\Sigma_{\mathbf{P}}\Sigma_{\mathbf{Q}}\Sigma_{\mathbf{R}}+n^2\,[\Sigma_{\mathbf{P}}\Sigma_{\mathbf{Q}}\Sigma_{\mathbf{R}'}+\mathrm{tr}(\Sigma_{\mathbf{R}})\Sigma_{\mathbf{P}}\Sigma_{\mathbf{Q}}+\Sigma_{\mathbf{P}}\Sigma_{\mathbf{Q}'}\Sigma_{\mathbf{R}}\\
&+\mathrm{tr}(\Sigma_{\mathbf{Q}})\Sigma_{\mathbf{P}}\Sigma_{\mathbf{R}}+\Sigma_{\mathbf{P}}\Sigma_{\mathbf{R}'}\Sigma_{\mathbf{Q}'}+\mathrm{tr}(\Sigma_{\mathbf{Q}}\Sigma_{\mathbf{R}})\Sigma_{\mathbf{P}}+\Sigma_{\mathbf{P}'}\Sigma_{\mathbf{Q}}\Sigma_{\mathbf{R}}+\mathrm{tr}(\Sigma_{\mathbf{P}})\Sigma_{\mathbf{Q}}\Sigma_{\mathbf{R}}\\
&++\Sigma_{\mathbf{Q}'}\Sigma_{\mathbf{P}'}\Sigma_{\mathbf{R}}+\mathrm{tr}(\Sigma_{\mathbf{P}}\Sigma_{\mathbf{Q}})\Sigma_{\mathbf{R}}+\Sigma_{\mathbf{R}'}\Sigma_{\mathbf{Q}'}\Sigma_{\mathbf{P}'}+\mathrm{tr}(\Sigma_{\mathbf{P}}\Sigma_{\mathbf{Q}}\Sigma_{\mathbf{R}})]\\
&+n[\Sigma_{\mathbf{P}}\Sigma_{\mathbf{Q}'}\Sigma_{\mathbf{R}'}+\mathrm{tr}(\Sigma_{\mathbf{R}})\Sigma_{\mathbf{P}}\Sigma_{\mathbf{Q}'}+\Sigma_{\mathbf{P}}\Sigma_{\mathbf{R}}\Sigma_{\mathbf{Q}}+\Sigma_{\mathbf{P}}\Sigma_{\mathbf{R}}\Sigma_{\mathbf{Q}'}+\mathrm{tr}(\Sigma_{\mathbf{Q}})\Sigma_{\mathbf{P}}\Sigma_{\mathbf{R}'}\\
&+\Sigma_{\mathbf{P}}\Sigma_{\mathbf{R}'}\Sigma_{\mathbf{Q}}+\mathrm{tr}(\Sigma_{\mathbf{Q}})\,\mathrm{tr}(\Sigma_{\mathbf{R}})\Sigma_{\mathbf{P}}+\mathrm{tr}(\Sigma_{\mathbf{Q}}\Sigma_{\mathbf{R}'})\Sigma_{\mathbf{P}}+\Sigma_{\mathbf{P}'}\Sigma_{\mathbf{Q}}\Sigma_{\mathbf{R}'}+\mathrm{tr}(\Sigma_{\mathbf{R}})\Sigma_{\mathbf{P}'}\Sigma_{\mathbf{Q}}\\
&+\Sigma_{\mathbf{P}'}\Sigma_{\mathbf{Q}'}\Sigma_{\mathbf{R}}+\mathrm{tr}(\Sigma_{\mathbf{Q}})\Sigma_{\mathbf{P}'}\Sigma_{\mathbf{R}}+\Sigma_{\mathbf{P}'}\Sigma_{\mathbf{R}'}\Sigma_{\mathbf{Q}'}+\mathrm{tr}(\Sigma_{\mathbf{Q}}\Sigma_{\mathbf{R}})\Sigma_{\mathbf{P}'}+\mathrm{tr}(\Sigma_{\mathbf{P}})\Sigma_{\mathbf{Q}}\Sigma_{\mathbf{R}'}\\
&+\mathrm{tr}(\Sigma_{\mathbf{P}})\,\mathrm{tr}(\Sigma_{\mathbf{R}})\Sigma_{\mathbf{Q}}+\Sigma_{\mathbf{Q}}\Sigma_{\mathbf{P}}\Sigma_{\mathbf{R}}+\Sigma_{\mathbf{Q}}\Sigma_{\mathbf{P}'}\Sigma_{\mathbf{R}}+\Sigma_{\mathbf{Q}}\Sigma_{\mathbf{R}}\Sigma_{\mathbf{P}}+\Sigma_{\mathbf{Q}}\Sigma_{\mathbf{R}}\Sigma_{\mathbf{P}'}\\
&+\mathrm{tr}(\Sigma_{\mathbf{P}})\Sigma_{\mathbf{Q}'}\Sigma_{\mathbf{R}}+\Sigma_{\mathbf{Q}'}\Sigma_{\mathbf{P}}\Sigma_{\mathbf{R}}+\mathrm{tr}(\Sigma_{\mathbf{P}}\Sigma_{\mathbf{R}'})\Sigma_{\mathbf{Q}'}+\Sigma_{\mathbf{Q}'}\Sigma_{\mathbf{R}'}\Sigma_{\mathbf{P}}+\Sigma_{\mathbf{Q}'}\Sigma_{\mathbf{P}'}\Sigma_{\mathbf{R}'}\\
&+\mathrm{tr}(\Sigma_{\mathbf{R}})\Sigma_{\mathbf{Q}'}\Sigma_{\mathbf{P}'}+\mathrm{tr}(\Sigma_{\mathbf{P}})\,\mathrm{tr}(\Sigma_{\mathbf{Q}})\Sigma_{\mathbf{R}}+\mathrm{tr}(\Sigma_{\mathbf{P}}\Sigma_{\mathbf{Q}'})\Sigma_{\mathbf{R}}+\Sigma_{\mathbf{R}}\Sigma_{\mathbf{P}'}\Sigma_{\mathbf{Q}'}+\Sigma_{\mathbf{R}}\Sigma_{\mathbf{Q}}\Sigma_{\mathbf{P}}\\
&+\Sigma_{\mathbf{R}}\Sigma_{\mathbf{P}}\Sigma_{\mathbf{Q}}+\Sigma_{\mathbf{R}}\Sigma_{\mathbf{Q}'}\Sigma_{\mathbf{P}'}+\mathrm{tr}(\Sigma_{\mathbf{P}})\Sigma_{\mathbf{R}'}\Sigma_{\mathbf{Q}'}+\Sigma_{\mathbf{R}'}\Sigma_{\mathbf{Q}'}\Sigma_{\mathbf{P}}+\mathrm{tr}(\Sigma_{\mathbf{P}}\Sigma_{\mathbf{Q}})\Sigma_{\mathbf{R}'}
\end{aligned}
$$

$$
\begin{aligned}
&+\Sigma_{\mathbf{R}'}\Sigma_{\mathbf{P}}\Sigma_{\mathbf{Q}}+\Sigma_{\mathbf{R}'}\Sigma_{\mathbf{Q}}\Sigma_{\mathbf{P}'}+\mathrm{tr}(\Sigma_{\mathbf{Q}})\Sigma_{\mathbf{R}'}\Sigma_{\mathbf{P}'}+\mathrm{tr}(\Sigma_{\mathbf{P}})\,\mathrm{tr}(\Sigma_{\mathbf{Q}}\Sigma_{\mathbf{R}})+\mathrm{tr}(\Sigma_{\mathbf{P}}\Sigma_{\mathbf{R}'}\Sigma_{\mathbf{Q}'})\\
&+\mathrm{tr}(\Sigma_{\mathbf{P}}\Sigma_{\mathbf{Q}})\,\mathrm{tr}(\Sigma_{\mathbf{R}})+\mathrm{tr}(\Sigma_{\mathbf{P}}\Sigma_{\mathbf{Q}}\Sigma_{\mathbf{R}})+\mathrm{tr}(\Sigma_{\mathbf{P}}\Sigma_{\mathbf{Q}'}\Sigma_{\mathbf{R}})+\mathrm{tr}(\Sigma_{\mathbf{P}}\Sigma_{\mathbf{R}})\,\mathrm{tr}(\Sigma_{\mathbf{Q}})]\\
&+\Sigma_{\mathbf{P}'}\Sigma_{\mathbf{Q}'}\Sigma_{\mathbf{R}'}+\mathrm{tr}(\Sigma_{\mathbf{R}})\Sigma_{\mathbf{P}'}\Sigma_{\mathbf{Q}'}+\Sigma_{\mathbf{P}'}\Sigma_{\mathbf{R}}\Sigma_{\mathbf{Q}}+\Sigma_{\mathbf{P}'}\Sigma_{\mathbf{R}}\Sigma_{\mathbf{Q}'}+\mathrm{tr}(\Sigma_{\mathbf{Q}})\Sigma_{\mathbf{P}'}\Sigma_{\mathbf{R}'}\\
&+\Sigma_{\mathbf{P}'}\Sigma_{\mathbf{R}'}\Sigma_{\mathbf{Q}}+\mathrm{tr}(\Sigma_{\mathbf{Q}})\,\mathrm{tr}(\Sigma_{\mathbf{R}})\Sigma_{\mathbf{P}'}+\mathrm{tr}(\Sigma_{\mathbf{Q}}\Sigma_{\mathbf{R}'})\Sigma_{\mathbf{P}'}+\Sigma_{\mathbf{Q}}\Sigma_{\mathbf{P}}\Sigma_{\mathbf{R}'}+\mathrm{tr}(\Sigma_{\mathbf{R}})\Sigma_{\mathbf{Q}}\Sigma_{\mathbf{P}}\\
&+\mathrm{tr}(\Sigma_{\mathbf{P}}\Sigma_{\mathbf{R}'})\Sigma_{\mathbf{Q}}+\Sigma_{\mathbf{Q}}\Sigma_{\mathbf{R}'}\Sigma_{\mathbf{P}}+\Sigma_{\mathbf{Q}}\Sigma_{\mathbf{P}'}\Sigma_{\mathbf{R}'}+\mathrm{tr}(\Sigma_{\mathbf{P}}\Sigma_{\mathbf{R}})\Sigma_{\mathbf{Q}}+\mathrm{tr}(\Sigma_{\mathbf{R}})\Sigma_{\mathbf{Q}}\Sigma_{\mathbf{P}'}\\
&+\Sigma_{\mathbf{Q}}\Sigma_{\mathbf{R}'}\Sigma_{\mathbf{P}'}+\mathrm{tr}(\Sigma_{\mathbf{P}})\Sigma_{\mathbf{Q}'}\Sigma_{\mathbf{R}'}+\mathrm{tr}(\Sigma_{\mathbf{P}})\,\mathrm{tr}(\Sigma_{\mathbf{R}})\Sigma_{\mathbf{Q}'}+\Sigma_{\mathbf{Q}'}\Sigma_{\mathbf{P}}\Sigma_{\mathbf{R}'}\\
&+\mathrm{tr}(\Sigma_{\mathbf{R}})\Sigma_{\mathbf{Q}'}\Sigma_{\mathbf{P}}+\mathrm{tr}(\Sigma_{\mathbf{R}'}\Sigma_{\mathbf{P}'})\Sigma_{\mathbf{Q}'}+\Sigma_{\mathbf{Q}'}\Sigma_{\mathbf{R}}\Sigma_{\mathbf{P}}+\Sigma_{\mathbf{Q}'}\Sigma_{\mathbf{R}'}\Sigma_{\mathbf{P}'}+\Sigma_{\mathbf{Q}'}\Sigma_{\mathbf{R}}\Sigma_{\mathbf{P}'}\\
&+\mathrm{tr}(\Sigma_{\mathbf{P}})\Sigma_{\mathbf{R}}\Sigma_{\mathbf{Q}}+\mathrm{tr}(\Sigma_{\mathbf{P}})\Sigma_{\mathbf{R}}\Sigma_{\mathbf{Q}'}+\Sigma_{\mathbf{R}}\Sigma_{\mathbf{P}'}\Sigma_{\mathbf{Q}}+\Sigma_{\mathbf{R}}\Sigma_{\mathbf{Q}'}\Sigma_{\mathbf{P}}+\Sigma_{\mathbf{R}}\Sigma_{\mathbf{P}}\Sigma_{\mathbf{Q}'}\\
&+\mathrm{tr}(\Sigma_{\mathbf{Q}})\Sigma_{\mathbf{R}}\Sigma_{\mathbf{P}}+\Sigma_{\mathbf{R}}\Sigma_{\mathbf{Q}}\Sigma_{\mathbf{P}'}+\mathrm{tr}(\Sigma_{\mathbf{Q}})\Sigma_{\mathbf{R}}\Sigma_{\mathbf{P}'}+\mathrm{tr}(\Sigma_{\mathbf{P}})\,\mathrm{tr}(\Sigma_{\mathbf{Q}})\Sigma_{\mathbf{R}'}\\
&+\mathrm{tr}(\Sigma_{\mathbf{P}})\Sigma_{\mathbf{R}'}\Sigma_{\mathbf{Q}}+\mathrm{tr}(\Sigma_{\mathbf{P}}\Sigma_{\mathbf{Q}'})\Sigma_{\mathbf{R}'}+\Sigma_{\mathbf{R}'}\Sigma_{\mathbf{P}'}\Sigma_{\mathbf{Q}}+\Sigma_{\mathbf{R}'}\Sigma_{\mathbf{P}'}\Sigma_{\mathbf{Q}'}+\Sigma_{\mathbf{R}'}\Sigma_{\mathbf{Q}}\Sigma_{\mathbf{P}}\\
&+\Sigma_{\mathbf{R}'}\Sigma_{\mathbf{P}}\Sigma_{\mathbf{Q}'}+\mathrm{tr}(\Sigma_{\mathbf{Q}})\Sigma_{\mathbf{R}'}\Sigma_{\mathbf{P}}+\mathrm{tr}(\Sigma_{\mathbf{P}})\,\mathrm{tr}(\Sigma_{\mathbf{Q}})\,\mathrm{tr}(\Sigma_{\mathbf{R}})+\mathrm{tr}(\Sigma_{\mathbf{P}})\,\mathrm{tr}(\Sigma_{\mathbf{Q}}\Sigma_{\mathbf{R}'})\\
&+\mathrm{tr}(\Sigma_{\mathbf{P}}\Sigma_{\mathbf{Q}'})\,\mathrm{tr}(\Sigma_{\mathbf{R}})+\mathrm{tr}(\Sigma_{\mathbf{P}}\Sigma_{\mathbf{R}}\Sigma_{\mathbf{Q}'})+\mathrm{tr}(\Sigma_{\mathbf{P}}\Sigma_{\mathbf{R}}\Sigma_{\mathbf{Q}})+\mathrm{tr}(\Sigma_{\mathbf{P}}\Sigma_{\mathbf{R}'}\Sigma_{\mathbf{Q}})\\
&+\mathrm{tr}(\Sigma_{\mathbf{P}}\Sigma_{\mathbf{Q}'}\Sigma_{\mathbf{R}'})+\mathrm{tr}(\Sigma_{\mathbf{P}}\Sigma_{\mathbf{R}'})\,\mathrm{tr}(\Sigma_{\mathbf{Q}})\}\Sigma,
\end{aligned}
$$

where $\Sigma_{\mathbf{B}}=\Sigma\mathbf{B}$. If we arrange the terms involving $\Sigma\mathbf{P}$, $\Sigma\mathbf{P}'$, $\Sigma\mathbf{Q}$, $\Sigma\mathbf{Q}'$, $\Sigma\mathbf{R}$, $\Sigma\mathbf{R}'$, and the others, then $\mathrm{E}(\mathbf{APAQARA})$ is expressed by sum of the following terms:

$$
\begin{aligned}
F_1=n\Sigma_{\mathbf{P}}[&n^3\Sigma_{\mathbf{Q}}\Sigma_{\mathbf{R}}+n^2\{\Sigma_{\mathbf{Q}}\Sigma_{\mathbf{R}'}+\mathrm{tr}(\Sigma_{\mathbf{R}})\Sigma_{\mathbf{Q}}+\Sigma_{\mathbf{Q}'}\Sigma_{\mathbf{R}}+\mathrm{tr}(\Sigma_{\mathbf{Q}})\Sigma_{\mathbf{R}}+\Sigma_{\mathbf{R}'}\Sigma_{\mathbf{Q}'}\\
&+\mathrm{tr}(\Sigma_{\mathbf{Q}}\Sigma_{\mathbf{R}})\}+n\{\Sigma_{\mathbf{Q}'}\Sigma_{\mathbf{R}'}+\mathrm{tr}(\Sigma_{\mathbf{R}})\Sigma_{\mathbf{Q}'}+\Sigma_{\mathbf{R}}\Sigma_{\mathbf{Q}}+\Sigma_{\mathbf{R}}\Sigma_{\mathbf{Q}'}+\mathrm{tr}(\Sigma_{\mathbf{Q}})\Sigma_{\mathbf{R}'}\\
&+\Sigma_{\mathbf{R}'}\Sigma_{\mathbf{Q}}+\mathrm{tr}(\Sigma_{\mathbf{Q}})\,\mathrm{tr}(\Sigma_{\mathbf{R}})+\mathrm{tr}(\Sigma_{\mathbf{Q}}\Sigma_{\mathbf{R}'})\}]\Sigma,
\end{aligned}
$$

$$
\begin{aligned}
F_2=n\Sigma_{\mathbf{P}'}[&n^2\Sigma_{\mathbf{Q}}\Sigma_{\mathbf{R}}+n\{\Sigma_{\mathbf{Q}}\Sigma_{\mathbf{R}'}+\mathrm{tr}(\Sigma_{\mathbf{R}})\Sigma_{\mathbf{Q}}+\Sigma_{\mathbf{Q}'}\Sigma_{\mathbf{R}}+\mathrm{tr}(\Sigma_{\mathbf{Q}})\Sigma_{\mathbf{R}}+\Sigma_{\mathbf{R}'}\Sigma_{\mathbf{Q}'}\\
&+\mathrm{tr}(\Sigma_{\mathbf{Q}}\Sigma_{\mathbf{R}})\}+\Sigma_{\mathbf{Q}'}\Sigma_{\mathbf{R}'}+\mathrm{tr}(\Sigma_{\mathbf{R}})\Sigma_{\mathbf{Q}'}+\Sigma_{\mathbf{R}}\Sigma_{\mathbf{Q}}+\Sigma_{\mathbf{R}}\Sigma_{\mathbf{Q}'}+\mathrm{tr}(\Sigma_{\mathbf{Q}})\Sigma_{\mathbf{R}'}\\
&+\Sigma_{\mathbf{R}'}\Sigma_{\mathbf{Q}}+\mathrm{tr}(\Sigma_{\mathbf{Q}})\,\mathrm{tr}(\Sigma_{\mathbf{R}})+\mathrm{tr}(\Sigma_{\mathbf{Q}}\Sigma_{\mathbf{R}'})]\Sigma,
\end{aligned}
$$

$$
\begin{aligned}
F_3=n\Sigma_{\mathbf{Q}}[&n^2\mathrm{tr}(\Sigma_{\mathbf{P}})\Sigma_{\mathbf{R}}+n\{\mathrm{tr}(\Sigma_{\mathbf{P}})\Sigma_{\mathbf{R}'}+\mathrm{tr}(\Sigma_{\mathbf{P}})\,\mathrm{tr}(\Sigma_{\mathbf{R}})+\Sigma_{\mathbf{P}}\Sigma_{\mathbf{R}}+\Sigma_{\mathbf{P}'}\Sigma_{\mathbf{R}}\\
&+\Sigma_{\mathbf{R}}\Sigma_{\mathbf{P}}+\Sigma_{\mathbf{R}}\Sigma_{\mathbf{P}'}\}+\Sigma_{\mathbf{P}}\Sigma_{\mathbf{R}'}+\mathrm{tr}(\Sigma_{\mathbf{R}})\Sigma_{\mathbf{P}}+\mathrm{tr}(\Sigma_{\mathbf{P}}\Sigma_{\mathbf{R}'})+\Sigma_{\mathbf{R}'}\Sigma_{\mathbf{P}}+\Sigma_{\mathbf{P}'}\Sigma_{\mathbf{R}'}\\
&+\mathrm{tr}(\Sigma_{\mathbf{P}}\Sigma_{\mathbf{R}})+\mathrm{tr}(\Sigma_{\mathbf{R}})\Sigma_{\mathbf{P}'}+\mathrm{tr}(\Sigma_{\mathbf{R}'}\Sigma_{\mathbf{P}'})]\Sigma,
\end{aligned}
$$

$$
\begin{aligned}
F_4=n\Sigma_{\mathbf{Q}'}[&n^2\Sigma_{\mathbf{P}'}\Sigma_{\mathbf{R}}+n\{\mathrm{tr}(\Sigma_{\mathbf{P}})\Sigma_{\mathbf{R}}+\Sigma_{\mathbf{P}}\Sigma_{\mathbf{R}}+\mathrm{tr}(\Sigma_{\mathbf{P}}\Sigma_{\mathbf{R}'})+\Sigma_{\mathbf{R}'}\Sigma_{\mathbf{P}}\\
&+\Sigma_{\mathbf{P}'}\Sigma_{\mathbf{R}'}+\mathrm{tr}(\Sigma_{\mathbf{R}})\Sigma_{\mathbf{P}'}\}+\mathrm{tr}(\Sigma_{\mathbf{P}})\Sigma_{\mathbf{R}'}+\mathrm{tr}(\Sigma_{\mathbf{P}})\,\mathrm{tr}(\Sigma_{\mathbf{R}})+\Sigma_{\mathbf{P}}\Sigma_{\mathbf{R}'}+\mathrm{tr}(\Sigma_{\mathbf{R}})\Sigma_{\mathbf{P}}\\
&+\mathrm{tr}(\Sigma_{\mathbf{R}'}\Sigma_{\mathbf{P}'})+\Sigma_{\mathbf{R}}\Sigma_{\mathbf{P}}+\Sigma_{\mathbf{R}'}\Sigma_{\mathbf{P}'}+\Sigma_{\mathbf{R}}\Sigma_{\mathbf{P}'}]\Sigma,
\end{aligned}
$$

$$
F_5=n\Sigma_{\mathbf{R}}[n^2\mathrm{tr}(\Sigma_{\mathbf{P}}\Sigma_{\mathbf{Q}})+n\{\mathrm{tr}(\Sigma_{\mathbf{P}})\,\mathrm{tr}(\Sigma_{\mathbf{Q}})+\mathrm{tr}(\Sigma_{\mathbf{P}}\Sigma_{\mathbf{Q}'})+\Sigma_{\mathbf{P}'}\Sigma_{\mathbf{Q}'}+\Sigma_{\mathbf{Q}}\Sigma_{\mathbf{P}}
$$

$$+ \mathbf{\Sigma_P \Sigma_Q} + \mathbf{\Sigma_{Q'} \Sigma_{P'}}\} + \mathrm{tr}(\mathbf{\Sigma_P})\mathbf{\Sigma_Q} + \mathrm{tr}(\mathbf{\Sigma_P})\mathbf{\Sigma_{Q'}} + \mathbf{\Sigma_{P'}\Sigma_Q} + \mathbf{\Sigma_{Q'}\Sigma_P} + \mathbf{\Sigma_P\Sigma_{Q'}}$$
$$+ \mathrm{tr}(\mathbf{\Sigma_Q})\mathbf{\Sigma_P} + \mathbf{\Sigma_Q\Sigma_{P'}} + \mathrm{tr}(\mathbf{\Sigma_Q})\mathbf{\Sigma_{P'}}]\mathbf{\Sigma},$$

$$F_6 = n\mathbf{\Sigma_{R'}}[n^2\mathbf{\Sigma_{Q'}\Sigma_{P'}} + n\{\mathrm{tr}(\mathbf{\Sigma_P})\mathbf{\Sigma_{Q'}} + \mathbf{\Sigma_{Q'}\Sigma_P} + \mathrm{tr}(\mathbf{\Sigma_P\Sigma_Q}) + \mathbf{\Sigma_P\Sigma_Q} + \mathbf{\Sigma_Q\Sigma_{P'}}$$
$$+ \mathrm{tr}(\mathbf{\Sigma_Q})\mathbf{\Sigma_{P'}}\} + \mathrm{tr}(\mathbf{\Sigma_P})\,\mathrm{tr}(\mathbf{\Sigma_Q}) + \mathrm{tr}(\mathbf{\Sigma_P})\mathbf{\Sigma_Q} + \mathrm{tr}(\mathbf{\Sigma_P\Sigma_{Q'}}) + \mathbf{\Sigma_{P'}\Sigma_Q}$$
$$+ + \mathbf{\Sigma_Q\Sigma_P} + \mathbf{\Sigma_P\Sigma_{Q'}} + \mathrm{tr}(\mathbf{\Sigma_Q})\mathbf{\Sigma_P}]\mathbf{\Sigma},$$

$$F_7 = n\,\mathrm{tr}(\mathbf{\Sigma_P}[n^2\mathbf{\Sigma_Q\Sigma_R} + n\{\mathbf{\Sigma_Q\Sigma_{R'}} + \mathrm{tr}(\mathbf{\Sigma_R})\mathbf{\Sigma_Q} + \mathbf{\Sigma_{Q'}\Sigma_R} + \mathrm{tr}(\mathbf{\Sigma_Q})\mathbf{\Sigma_R} + \mathbf{\Sigma_{R'}\Sigma_{Q'}}$$
$$+ \mathrm{tr}(\mathbf{\Sigma_Q\Sigma_R})\} + \mathbf{\Sigma_{Q'}\Sigma_{R'}} + \mathrm{tr}(\mathbf{\Sigma_R})\mathbf{\Sigma_{Q'}} + \mathbf{\Sigma_R\Sigma_Q} + \mathbf{\Sigma_R\Sigma_{Q'}} + \mathrm{tr}(\mathbf{\Sigma_Q})\mathbf{\Sigma_{R'}}$$
$$+ \mathbf{\Sigma_{R'}\Sigma_Q} + \mathrm{tr}(\mathbf{\Sigma_Q})\,\mathrm{tr}(\mathbf{\Sigma_R}) + \mathrm{tr}(\mathbf{\Sigma_Q\Sigma_{R'}})])\mathbf{\Sigma},$$

If we further arrange the equations $F_1, \cdots, F_7$, then

$$F_1 = n\mathbf{\Sigma_P\Sigma_Q}[n^3\mathbf{\Sigma_R} + n^2\mathbf{\Sigma_{R'}} + n^2\mathrm{tr}(\mathbf{\Sigma_R})]\mathbf{\Sigma} + n\mathbf{\Sigma_P\Sigma_{Q'}}[n^2\mathbf{\Sigma_R} + n\mathbf{\Sigma_{R'}} + n\mathrm{tr}(\mathbf{\Sigma_R})]\mathbf{\Sigma}$$
$$+ n\mathbf{\Sigma_P\Sigma_R}[n^2\,\mathrm{tr}(\mathbf{\Sigma_Q}) + n\mathbf{\Sigma_Q} + n\mathbf{\Sigma_{Q'}}]\mathbf{\Sigma} + n\mathbf{\Sigma_P\Sigma_{R'}}[n^2\mathbf{\Sigma_{Q'}} + n\,\mathrm{tr}(\mathbf{\Sigma_Q}) + n\mathbf{\Sigma_Q}]\mathbf{\Sigma}$$
$$+ n\mathbf{\Sigma_P}\,\mathrm{tr}(\mathbf{\Sigma_Q}[n^2\mathbf{\Sigma_R} + n\,\mathrm{tr}(\mathbf{\Sigma_Q}) + n\mathbf{\Sigma_{R'}}])\mathbf{\Sigma},$$
$$= n^2\mathbf{\Sigma_P\Sigma_Q}\mathrm{E}[\mathbf{ARA}] + n\mathbf{\Sigma_P\Sigma_{Q'}}\mathrm{E}[\mathbf{ARA}] + n\mathbf{\Sigma_P\Sigma_R}\mathrm{E}[\mathrm{tr}(\mathbf{QA})\mathbf{A}]$$
$$+ n\mathbf{\Sigma_P\Sigma_{R'}}\mathrm{E}[\mathbf{AQ'A}] + n\mathbf{\Sigma_P}\,\mathrm{tr}(\mathbf{Q}\mathrm{E}[\mathbf{ARA}])\mathbf{\Sigma},$$

$$F_2 = n\mathbf{\Sigma_{P'}\Sigma_Q}[n^2\mathbf{\Sigma_R} + n\mathbf{\Sigma_{R'}} + n\mathrm{tr}(\mathbf{\Sigma_R})]\mathbf{\Sigma} + n\mathbf{\Sigma_{P'}\Sigma_{Q'}}[n\mathbf{\Sigma_R} + \mathbf{\Sigma_{R'}} + \mathrm{tr}(\mathbf{\Sigma_R})]\mathbf{\Sigma}$$
$$+ n\mathbf{\Sigma_{P'}\Sigma_R}[n\,\mathrm{tr}(\mathbf{\Sigma_Q}) + \mathbf{\Sigma_Q} + \mathbf{\Sigma_{Q'}}]\mathbf{\Sigma} + n\mathbf{\Sigma_{P'}\Sigma_{R'}}[n\mathbf{\Sigma_{Q'}} + \mathrm{tr}(\mathbf{\Sigma_Q}) + \mathbf{\Sigma_Q}]\mathbf{\Sigma}$$
$$+ n\mathbf{\Sigma_{P'}}\,\mathrm{tr}(\mathbf{\Sigma_Q}[n\mathbf{\Sigma_R} + \mathrm{tr}(\mathbf{\Sigma_Q}) + \mathbf{\Sigma_{R'}}])\mathbf{\Sigma},$$
$$= n\mathbf{\Sigma_{P'}\Sigma_Q}\mathrm{E}[\mathbf{ARA}] + \mathbf{\Sigma_{P'}\Sigma_{Q'}}\mathrm{E}[\mathbf{ARA}] + \mathbf{\Sigma_{P'}\Sigma_R}\mathrm{E}[\mathrm{tr}(\mathbf{QA})\mathbf{A}]$$
$$+ \mathbf{\Sigma_{P'}\Sigma_{R'}}\mathrm{E}[\mathbf{AQ'A}] + \mathbf{\Sigma_{P'}}\,\mathrm{tr}(\mathbf{Q}\mathrm{E}[\mathbf{ARA}])\mathbf{\Sigma},$$

$$F_3 = n\mathbf{\Sigma_Q\Sigma_R}[n\mathbf{\Sigma_P} + n\mathbf{\Sigma_{P'}} + n^2\mathrm{tr}(\mathbf{\Sigma_P})]\mathbf{\Sigma} + n\mathbf{\Sigma_Q\Sigma_{R'}}[\mathbf{\Sigma_P} + \mathbf{\Sigma_{P'}} + n\mathrm{tr}(\mathbf{\Sigma_P})]\mathbf{\Sigma}$$
$$+ n\mathbf{\Sigma_Q\Sigma_P}[\mathrm{tr}(\mathbf{\Sigma_R}) + n\mathbf{\Sigma_R} + \mathbf{\Sigma_{R'}}]\mathbf{\Sigma} + n\mathbf{\Sigma_Q\Sigma_{P'}}[n\mathbf{\Sigma_R} + \mathrm{tr}(\mathbf{\Sigma_R}) + \mathbf{\Sigma_{R'}}]\mathbf{\Sigma}$$
$$+ n\mathbf{\Sigma_Q}\,\mathrm{tr}(\mathbf{\Sigma_P}[\mathbf{\Sigma_R} + n\,\mathrm{tr}(\mathbf{\Sigma_R}) + \mathbf{\Sigma_{R'}}])\mathbf{\Sigma},$$
$$= n\mathbf{\Sigma_Q\Sigma_R}\mathrm{E}[\mathbf{APA}] + \mathbf{\Sigma_Q\Sigma_{R'}}\mathrm{E}[\mathrm{tr}(\mathbf{PA})\mathbf{A}] + \mathbf{\Sigma_Q\Sigma_P}\mathrm{E}[\mathbf{ARA}] + \mathbf{\Sigma_Q\Sigma_{P'}}\mathrm{E}[\mathbf{ARA}]$$
$$+ \mathbf{\Sigma_Q}\,\mathrm{tr}(\mathbf{P}\mathrm{E}[\mathrm{tr}(\mathbf{RA})\mathbf{A}])\mathbf{\Sigma},$$

$$F_4 = n\mathbf{\Sigma_{Q'}\Sigma_{P'}}[n^2\mathbf{\Sigma_R} + n\mathbf{\Sigma_{R'}} + n\mathrm{tr}(\mathbf{\Sigma_R})]\mathbf{\Sigma} + n\mathbf{\Sigma_{Q'}\Sigma_P}[n\mathbf{\Sigma_R} + \mathbf{\Sigma_{R'}} + \mathrm{tr}(\mathbf{\Sigma_R})]\mathbf{\Sigma}$$
$$+ n\mathbf{\Sigma_{Q'}\Sigma_{R'}}[\mathrm{tr}(\mathbf{\Sigma_P}) + n\mathbf{\Sigma_P} + \mathbf{\Sigma_{P'}}]\mathbf{\Sigma} + n\mathbf{\Sigma_{Q'}\Sigma_R}[\mathbf{\Sigma_P} + n\,\mathrm{tr}(\mathbf{\Sigma_P}) + \mathbf{\Sigma_{P'}}]\mathbf{\Sigma}$$
$$+ n\mathbf{\Sigma_{Q'}}\,\mathrm{tr}(\mathbf{\Sigma_P}[n\mathbf{\Sigma_{R'}} + \mathrm{tr}(\mathbf{\Sigma_R}) + \mathbf{\Sigma_R}])\mathbf{\Sigma},$$

$$= n\boldsymbol{\Sigma}_{\mathbf{Q}'}\boldsymbol{\Sigma}_{\mathbf{P}'}\mathrm{E}[\mathbf{ARA}] + \boldsymbol{\Sigma}_{\mathbf{Q}'}\boldsymbol{\Sigma}_{\mathbf{P}}\mathrm{E}[\mathbf{ARA}] + \boldsymbol{\Sigma}_{\mathbf{Q}'}\boldsymbol{\Sigma}_{\mathbf{R}'}\mathrm{E}[\mathbf{APA}]$$
$$+ \boldsymbol{\Sigma}_{\mathbf{Q}'}\boldsymbol{\Sigma}_{\mathbf{R}}\mathrm{E}[\mathrm{tr}(\mathbf{PA})\mathbf{A}] + \boldsymbol{\Sigma}_{\mathbf{Q}'}\,\mathrm{tr}(\boldsymbol{\Sigma}_{\mathbf{P}}\mathrm{E}[\mathbf{AR}'\mathbf{A}])\boldsymbol{\Sigma},$$

$$F_5 = n\boldsymbol{\Sigma}_{\mathbf{R}}\boldsymbol{\Sigma}_{\mathbf{P}}[n\boldsymbol{\Sigma}_{\mathbf{Q}} + \boldsymbol{\Sigma}_{\mathbf{Q}'} + \mathrm{tr}(\boldsymbol{\Sigma}_{\mathbf{Q}})]\boldsymbol{\Sigma} + n\boldsymbol{\Sigma}_{\mathbf{R}}\boldsymbol{\Sigma}_{\mathbf{P}'}[n\boldsymbol{\Sigma}_{\mathbf{Q}'} + \boldsymbol{\Sigma}_{\mathbf{Q}} + \mathrm{tr}(\boldsymbol{\Sigma}_{\mathbf{Q}})]\boldsymbol{\Sigma}$$
$$+ n\boldsymbol{\Sigma}_{\mathbf{R}}\boldsymbol{\Sigma}_{\mathbf{Q}}[\mathrm{tr}(\boldsymbol{\Sigma}_{\mathbf{P}}) + n\boldsymbol{\Sigma}_{\mathbf{P}} + \boldsymbol{\Sigma}_{\mathbf{P}'}]\boldsymbol{\Sigma} + n\boldsymbol{\Sigma}_{\mathbf{R}}\boldsymbol{\Sigma}_{\mathbf{Q}'}[\boldsymbol{\Sigma}_{\mathbf{P}} + \mathrm{tr}(\boldsymbol{\Sigma}_{\mathbf{P}}) + n\boldsymbol{\Sigma}_{\mathbf{P}'}]\boldsymbol{\Sigma}$$
$$+ n\boldsymbol{\Sigma}_{\mathbf{R}}\,\mathrm{tr}(\boldsymbol{\Sigma}_{\mathbf{P}}[n^2\boldsymbol{\Sigma}_{\mathbf{Q}} + n\,\mathrm{tr}(\boldsymbol{\Sigma}_{\mathbf{Q}}) + \boldsymbol{\Sigma}_{\mathbf{Q}'}])\boldsymbol{\Sigma},$$
$$= \boldsymbol{\Sigma}_{\mathbf{R}}\boldsymbol{\Sigma}_{\mathbf{P}}\mathrm{E}[\mathbf{AQA}] + \boldsymbol{\Sigma}_{\mathbf{R}}\boldsymbol{\Sigma}_{\mathbf{P}'}\mathrm{E}[\mathbf{AQ}'\mathbf{A}] + \boldsymbol{\Sigma}_{\mathbf{R}}\boldsymbol{\Sigma}_{\mathbf{Q}}\mathrm{E}[\mathbf{APA}] + \boldsymbol{\Sigma}_{\mathbf{R}}\boldsymbol{\Sigma}_{\mathbf{Q}'}\mathrm{E}[\mathbf{AP}'\mathbf{A}]$$
$$+ n\boldsymbol{\Sigma}_{\mathbf{R}}\,\mathrm{tr}(\mathbf{P}\mathrm{E}[\mathrm{tr}(\mathbf{QA})\mathbf{A}])\boldsymbol{\Sigma},$$

$$F_6 = n\boldsymbol{\Sigma}_{\mathbf{R}'}\boldsymbol{\Sigma}_{\mathbf{Q}'}[n^2\boldsymbol{\Sigma}_{\mathbf{P}'} + n\boldsymbol{\Sigma}_{\mathbf{P}} + n\mathrm{tr}(\boldsymbol{\Sigma}_{\mathbf{P}})]\boldsymbol{\Sigma} + n\boldsymbol{\Sigma}_{\mathbf{R}'}\boldsymbol{\Sigma}_{\mathbf{P}}[n\boldsymbol{\Sigma}_{\mathbf{Q}} + \boldsymbol{\Sigma}_{\mathbf{Q}'} + \mathrm{tr}(\boldsymbol{\Sigma}_{\mathbf{Q}})]\boldsymbol{\Sigma}$$
$$+ n\boldsymbol{\Sigma}_{\mathbf{R}'}\boldsymbol{\Sigma}_{\mathbf{P}'}[n\,\mathrm{tr}(\boldsymbol{\Sigma}_{\mathbf{Q}}) + \boldsymbol{\Sigma}_{\mathbf{Q}} + \boldsymbol{\Sigma}_{\mathbf{Q}'}]\boldsymbol{\Sigma} + n\boldsymbol{\Sigma}_{\mathbf{R}'}\boldsymbol{\Sigma}_{\mathbf{Q}}[\boldsymbol{\Sigma}_{\mathbf{P}} + \mathrm{tr}(\boldsymbol{\Sigma}_{\mathbf{P}}) + n\boldsymbol{\Sigma}_{\mathbf{P}'}]\boldsymbol{\Sigma}$$
$$+ n\boldsymbol{\Sigma}_{\mathbf{R}'}\,\mathrm{tr}(\boldsymbol{\Sigma}_{\mathbf{P}}[\boldsymbol{\Sigma}_{\mathbf{Q}'} + \mathrm{tr}(\boldsymbol{\Sigma}_{\mathbf{Q}}) + n\boldsymbol{\Sigma}_{\mathbf{Q}}])\boldsymbol{\Sigma},$$
$$= n\boldsymbol{\Sigma}_{\mathbf{R}'}\boldsymbol{\Sigma}_{\mathbf{Q}'}\mathrm{E}[\mathbf{AP}'\mathbf{A}] + \boldsymbol{\Sigma}_{\mathbf{R}'}\boldsymbol{\Sigma}_{\mathbf{P}}\mathrm{E}[\mathbf{AQA}] + \boldsymbol{\Sigma}_{\mathbf{R}'}\boldsymbol{\Sigma}_{\mathbf{P}'}\mathrm{E}[\mathrm{tr}(\mathbf{QA})\mathbf{A}]$$
$$+ \boldsymbol{\Sigma}_{\mathbf{R}'}\boldsymbol{\Sigma}_{\mathbf{Q}}\mathrm{E}[\mathbf{AP}'\mathbf{A}] + n\boldsymbol{\Sigma}_{\mathbf{R}'}\,\mathrm{tr}(\mathbf{P}\mathrm{E}[\mathbf{AQA}])\boldsymbol{\Sigma},$$

$$F_7 = n\,\mathrm{tr}(\boldsymbol{\Sigma}_{\mathbf{P}}\boldsymbol{\Sigma}_{\mathbf{Q}}[n^2\boldsymbol{\Sigma}_{\mathbf{R}} + n\boldsymbol{\Sigma}_{\mathbf{R}'} + n\mathrm{tr}(\boldsymbol{\Sigma}_{\mathbf{R}})] + n\boldsymbol{\Sigma}_{\mathbf{P}}\boldsymbol{\Sigma}_{\mathbf{Q}'}[n\boldsymbol{\Sigma}_{\mathbf{R}} + \boldsymbol{\Sigma}_{\mathbf{R}'} + \mathrm{tr}(\boldsymbol{\Sigma}_{\mathbf{R}})]$$
$$+ n\boldsymbol{\Sigma}_{\mathbf{P}}\boldsymbol{\Sigma}_{\mathbf{R}}[n\,\mathrm{tr}(\boldsymbol{\Sigma}_{\mathbf{Q}}) + \boldsymbol{\Sigma}_{\mathbf{Q}} + \boldsymbol{\Sigma}_{\mathbf{Q}'}] + n\boldsymbol{\Sigma}_{\mathbf{P}}\boldsymbol{\Sigma}_{\mathbf{R}'}[n\boldsymbol{\Sigma}_{\mathbf{Q}'} + \mathrm{tr}(\boldsymbol{\Sigma}_{\mathbf{Q}}) + \boldsymbol{\Sigma}_{\mathbf{Q}}]$$
$$+ n\boldsymbol{\Sigma}_{\mathbf{P}}\,\mathrm{tr}(\boldsymbol{\Sigma}_{\mathbf{Q}}[n\boldsymbol{\Sigma}_{\mathbf{R}} + \mathrm{tr}(\boldsymbol{\Sigma}_{\mathbf{R}}) + \boldsymbol{\Sigma}_{\mathbf{R}'}]))\boldsymbol{\Sigma}$$
$$= n\,\mathrm{tr}(\boldsymbol{\Sigma}_{\mathbf{P}}\boldsymbol{\Sigma}_{\mathbf{Q}}\mathrm{E}[\mathbf{ARA}] + \boldsymbol{\Sigma}_{\mathbf{P}}\boldsymbol{\Sigma}_{\mathbf{Q}'}\mathrm{E}[\mathbf{ARA}] + \boldsymbol{\Sigma}_{\mathbf{P}}\boldsymbol{\Sigma}_{\mathbf{R}}\mathrm{E}[\mathrm{tr}(\mathbf{QA})\mathbf{A}]$$
$$+ \boldsymbol{\Sigma}_{\mathbf{P}}\boldsymbol{\Sigma}_{\mathbf{R}'}\mathrm{E}[\mathbf{AQ}'\mathbf{A}] + \boldsymbol{\Sigma}_{\mathbf{P}}\,\mathrm{tr}(\mathbf{Q}\mathrm{E}[\mathbf{ARA}]))\boldsymbol{\Sigma},$$

where we have used Theorem 3.3.15 of Gupta and Nagar (2000).

If we take the summation of F_1 to F_7, after some cumbersome calculations, we can obtain the desired result. □

As mentioned in Section 2.2, since the conditional distribution of $\mathbf{l}'\mathbf{Az}$ given $\mathbf{A}$ is $\mathbf{l}'\mathbf{Az}|\mathbf{A} \sim N(\mathbf{l}'\mathbf{A}\boldsymbol{\mu}, \mathbf{l}'\mathbf{A}\boldsymbol{\Omega}\mathbf{Al})$, the first four moments of $v = \mathbf{l}'\mathbf{Az}$ can be computed as

$$\mathrm{E}[v] = \mathbf{l}'\mathrm{E}(\mathbf{A})\boldsymbol{\mu}, \tag{31}$$
$$\mathrm{E}[v^2] = \mathrm{E}[(\mathbf{l}'\mathbf{A}\boldsymbol{\mu})^2 + \mathbf{l}'\mathbf{A}\boldsymbol{\Omega}\mathbf{Al}] = \mathbf{l}'\mathrm{E}[\mathbf{A}\boldsymbol{\mu}\boldsymbol{\mu}'\mathbf{A} + \mathbf{A}\boldsymbol{\Omega}\mathbf{A}]\mathbf{l}, \tag{32}$$
$$\mathrm{E}[v^3] = \mathrm{E}[(\mathbf{l}'\mathbf{A}\boldsymbol{\mu})^3 + 3\mathbf{l}'\mathbf{A}\boldsymbol{\mu}\mathbf{l}'\mathbf{A}\boldsymbol{\Omega}\mathbf{Al}] = \mathbf{l}'\mathrm{E}[\mathbf{A}\boldsymbol{\mu}\mathbf{l}'\mathbf{A}\boldsymbol{\mu}\boldsymbol{\mu}'\mathbf{A} + 3\mathbf{A}\boldsymbol{\mu}\mathbf{l}'\mathbf{A}\boldsymbol{\Omega}\mathbf{A}]\mathbf{l}, \tag{33}$$
$$\mathrm{E}[v^4] = \mathrm{E}[(\mathbf{l}'\mathbf{A}\boldsymbol{\mu})^4 + 6(\mathbf{l}'\mathbf{A}\boldsymbol{\mu})^2\mathbf{l}'\mathbf{A}\boldsymbol{\Omega}\mathbf{Al} + 3(\mathbf{l}'\mathbf{A}\boldsymbol{\Omega}\mathbf{Al})^2]$$
$$= \mathbf{l}'\mathrm{E}[\mathbf{A}\boldsymbol{\mu}\mathbf{l}'\mathbf{A}\boldsymbol{\mu}\mathbf{l}'\mathbf{A}\boldsymbol{\mu}\boldsymbol{\mu}'\mathbf{A} + 6\mathbf{A}\boldsymbol{\mu}\mathbf{l}'\mathbf{A}\boldsymbol{\mu}\mathbf{l}'\mathbf{A}\boldsymbol{\Omega}\mathbf{A} + 3\mathbf{A}\boldsymbol{\Omega}\mathbf{All}'\mathbf{A}\boldsymbol{\Omega}\mathbf{A}]\mathbf{l}. \tag{34}$$

Substituting $\mathrm{E}(\mathbf{A}) = n\boldsymbol{\Sigma}$ for (31), we have $\mathrm{E}(v) = n\mathbf{l}'\boldsymbol{\Sigma}\boldsymbol{\mu}$. Applying Theorem 3.3.15 (ii) of Gupta and Nagar (2000) to (32), we get

$$\begin{aligned}\mathrm{E}[v^2] = n(&\mathbf{l}'\boldsymbol{\Sigma\Omega\Sigma}\mathbf{l} + \mathrm{tr}(\boldsymbol{\Sigma\Omega})\mathbf{l}'\boldsymbol{\Sigma}\mathbf{l} + n\mathbf{l}'\boldsymbol{\Sigma\Omega\Sigma}\mathbf{l} \\ &+ \mathbf{l}'\boldsymbol{\Sigma\mu\mu}'\boldsymbol{\Sigma}\mathbf{l} + \mathrm{tr}(\boldsymbol{\Sigma\mu\mu}')\mathbf{l}'\boldsymbol{\Sigma}\mathbf{l} + n\mathbf{l}'\boldsymbol{\Sigma\mu\mu}'\boldsymbol{\Sigma}\mathbf{l}).\end{aligned}$$

We can obtain the variance of v by computing $\mathrm{E}[v^2] - [\mathrm{E}(v)]^2$. Applying Lemma 2.8 to (33) and (34), after some algebraic calculations, we can obtain the same expressions of skewness and kurtosis as Corollary 2.4.

3 On the product of a Wishart matrix and a normal vector with a dependence structure

3.1 Introduction

Over the past decade, there has been an increase in the number of studies on functions of a Wishart matrix and a normal vector. These studies have been motivated the fact that these two objects appear often together in the expressions of different statistics, e.g., Hotelling T^2-statistic, principal components, discriminant function and weights of portfolio under multivariate normality.

Let us summarize the previous studies on functions of a Wishart matrix and a normal vector. Bodnar and Okhrin (2011) derived the density functions of the product of an inverse Wishart matrix and a normal vector, in addition, discussed applications to portfolio theory and discriminant analysis. Bodnar et al. (2013) presented the stochastic representations and the density functions of the product of a Wishart matrix and a normal vector, furthermore, found an approximation for the density functions, which is based on a Taylor series expansion. Bodnar et al. (2015) investigated the distributional properties of the product of a singular Wishart matrix and a normal vector. Kotsuiba and Mazur (2016) extended the results of Bodnar and Okhrin (2011) by deriving the asymptotic and approximate density functions of the product of an inverse Wishart matrix and a normal vector. Bodnar et al. (2019) derived the stochastic representations of the product of a singular Wishart matrix and a singular normal vector, in addition, proved the asymptotic normality of the product under the high-dimensional asymptotic regime. Javed et al. (2021) derived the higher order moments of the product of an inverse Wishart matrix and a normal vector.

Although previous researches on functions have been assumed that a Wishart matrix and a normal vector are independent, we consider the situation in which a Wishart matrix and a normal vector are dependent. This situation can be considered in a Bayesian framework. In portfolio theory, for example, the weights of tangent portfolio and of multi-period optimal portfolio are expressed as the product of a Wishart matrix and a normal vector, which are dependently distributed. Bauder and his co-authors (2018, 2020) derived the stochastic representations for the weights of tangent portfolio and of multi-period optimal portfolio, respectively. In addition, they provided the first two moments and established asymptotic normality of the product by using the stochastic representations. Although the derived stochastic representations are computationally efficient, those representations could not be mathematically tractable as Bauder et al. (2020) pointed out themselves. This can cause difficulties in understanding some distributional properties of the product. The density function, for example, has not been provided and estimated by using kernel method.

Therefore the objectives of this research are: 1. to obtain more simple stochastic representation, 2. investigate some distributional properties including the explicit expression of the density function.

The rest of this paper is structured as follows. In section 3.2, we derive the stochastic representations of the product of a Wishart matrix and a normal vector, which are dependently distributed. It is applied to derive the density function and first four moments in Corollary 3.2 and 3.3. It is found that the distribution of the product is closed under conditioning, marginalization, and affine transformations. In addition, the approximation for the distribution of the product is discussed. The results of numerical studies are given in Section 3.3, while Section 3.4 summarizes the paper. In the appendix, we give the new stochastic representations of some functions which appear in the efficient frontier. These results improve the computational efficiency of the existing stochastic representations given by Bauder et al. (2019).

3.2 Main results

In previous studies attempts have been made to establish stochastic representations of the product of a Wishart matrix and a normal vector. In most cases, the obtained stochastic representations for the product are expressed in terms of the well-known distributions, such as χ^2 and normal distributions. These representations allow us to access the density function, moments and limiting distribution. In addition, the stochastic representation often makes the simulation procedure more computationally efficient (cf. Bodnar and Okhrin, 2011; Bodnar and his co-authors, 2013, 2015, 2019). In this section, therefore, we study the stochastic representation for $\mathbf{Az}$ where $\mathbf{A} \sim W_k(n, \mathbf{\Sigma})$ and $\mathbf{z}|\mathbf{A} \sim N_k(\boldsymbol{\mu}, \mathbf{A}^{-1}/\kappa)$.

Let $\mathbf{L}$ be an $p \times k$ constant matrix with $\operatorname{rank} \mathbf{L} = p < k$. Bauder and his co-authors (2018, 2020) presented the stochastic representation for $\mathbf{LAz}$ as

$$\mathbf{LAz} \stackrel{d}{=} \eta \mathbf{L}\boldsymbol{\zeta} + \sqrt{\eta}(\epsilon \mathbf{L}\boldsymbol{\Upsilon}\mathbf{L}' - \mathbf{L}\boldsymbol{\zeta}\boldsymbol{\zeta}'\mathbf{L}')^{1/2}\mathbf{z}_0, \tag{35}$$

$$\epsilon = \epsilon(Q, \mathbf{U}) = \boldsymbol{\mu}'\mathbf{\Sigma}\boldsymbol{\mu} + \frac{2\sqrt{n-k+1}}{\sqrt{\kappa}} \frac{\sqrt{kQ}}{n+1+k(Q-1)} \boldsymbol{\mu}'\mathbf{\Sigma}^{1/2}\mathbf{U}$$
$$+ \frac{1}{\kappa} \frac{kQ}{n+1+k(Q-1)} - \frac{kQ}{n+1+k(Q-1)} (\boldsymbol{\mu}'\mathbf{\Sigma}^{1/2}\mathbf{U})^2,$$

$$\boldsymbol{\zeta} = \boldsymbol{\zeta}(Q, \mathbf{U}) = \mathbf{\Sigma}\boldsymbol{\mu} + \frac{\sqrt{n-k+1}}{\sqrt{\kappa}} \frac{\sqrt{kQ}}{n+1+k(Q-1)} \mathbf{\Sigma}^{1/2}\mathbf{U}$$
$$- \frac{kQ}{n+1+k(Q-1)} \mathbf{\Sigma}^{1/2}\mathbf{U}\mathbf{U}'\mathbf{\Sigma}^{1/2}\boldsymbol{\mu}$$

$$\boldsymbol{\Upsilon} = \boldsymbol{\Upsilon}(Q, \mathbf{U}) = \mathbf{\Sigma} - \frac{kQ}{n+1+k(Q-1)} \mathbf{\Sigma}^{1/2}\mathbf{U}\mathbf{U}'\mathbf{\Sigma}^{1/2},$$

where $\eta \sim \chi^2_{n+1}$, $\mathbf{z}_0 \sim N_p(\mathbf{0}, \mathbf{I}_p)$, $Q \sim F(k, n-k+1)$ and $\mathbf{U}$ is uniformly distributed on the unit sphere in R^k; moreover, they are mutually independent.

The stochastic representation (35) simplifies the simulation procedures since random sampling is required only for the $p+k+2$ random variables. Without the representation (35), we would have to simulate $k + k(k+1)/2$ random variables. Although the uniform distribution on a unit sphere in R^k is not a standard distribution in many statistical packages, the realizations of $\mathbf{U}$ can easily be obtained from the k-dimensional standard normal vector $\mathbf{Z}$ by using $\mathbf{U} = \mathbf{Z}/\sqrt{\mathbf{Z}'\mathbf{Z}}$.

In the proof of the stochastic representation (35), the normal vector $\mathbf{z}$ was fixed, which results in $\mathbf{A}|\mathbf{z} \sim W_k(n+1, [\kappa(\mathbf{z}-\boldsymbol{\mu})(\mathbf{z}-\boldsymbol{\mu})' + \boldsymbol{\Sigma}^{-1}]^{-1})$ and $\mathbf{z} \sim t_k(n-k+1, \boldsymbol{\mu}, \boldsymbol{\Sigma})$, where $t_k(m, \boldsymbol{\theta}, \mathbf{T})$ denote the m-dimensional t-distribution with m degrees of freedom, location vector $\boldsymbol{\theta}$, and dispersion matrix $\mathbf{T}$. They applied the approach which is detailed in Bodnar and his co-authors (2013, 2015, 2019) to this setting. While, we fix the Wishart matrix $\mathbf{A}$ to obtain the novel stochastic representation of $\mathbf{Az}$ as follows:

Theorem 3.1. *Suppose* $\mathbf{A} \sim W_k(n, \boldsymbol{\Sigma})$ *with* $\boldsymbol{\Sigma} > 0$ *and* $\mathbf{z}|\mathbf{A} \sim N_k(\boldsymbol{\mu}, \mathbf{A}^{-1}/\kappa)$*,* $\kappa > 0$*. Let* $\boldsymbol{\delta} = \boldsymbol{\Sigma}\boldsymbol{\mu}$*,* $\boldsymbol{\Omega} = (\kappa^{-1} + \boldsymbol{\mu}'\boldsymbol{\Sigma}\boldsymbol{\mu})\boldsymbol{\Sigma} - \boldsymbol{\Sigma}\boldsymbol{\mu}\boldsymbol{\mu}'\boldsymbol{\Sigma}$*. Then, the stochastic representation of* $\mathbf{Az}$ *are given by*

$$\mathbf{Az} \overset{\mathrm{d}}{=} W\boldsymbol{\delta} + \sqrt{W}\boldsymbol{\Omega}^{1/2}\mathbf{Z} \tag{36}$$

where $W \sim \chi^2_n$ *and* $\mathbf{Z} \sim N_k(\mathbf{0}, \mathbf{I}_k)$*; they are independent.*

Proof. The conditional distribution of $\mathbf{l}'\mathbf{Az}$ given $\mathbf{A}$ is $N(\mathbf{l}'\mathbf{A}\boldsymbol{\mu}, \mathbf{l}'\mathbf{Al}/\kappa)$. This implies the following stochastic representation:

$$\mathbf{l}'\mathbf{Az} \overset{\mathrm{d}}{=} \sqrt{\mathbf{l}'\mathbf{Al}}\mathbf{z}_0 + \mathbf{l}'\mathbf{A}\boldsymbol{\mu} \text{ with } \mathbf{z}_0 \sim N(0, 1/\kappa). \tag{37}$$

Let $\tilde{\mathbf{L}} = (\boldsymbol{\mu}, \mathbf{l})'$. If we assume $\operatorname{rank}(\tilde{\mathbf{L}}) = 2$, then $\tilde{\mathbf{L}}\mathbf{A}\tilde{\mathbf{L}}' \sim W_2(n, \mathbf{H})$, where

$$\tilde{\mathbf{L}}\mathbf{A}\tilde{\mathbf{L}}' = \begin{pmatrix} \boldsymbol{\mu}'\mathbf{A}\boldsymbol{\mu} & \boldsymbol{\mu}'\mathbf{Al} \\ \mathbf{l}'\mathbf{A}\boldsymbol{\mu} & \mathbf{l}'\mathbf{Al} \end{pmatrix}, \mathbf{H} = \begin{pmatrix} \boldsymbol{\mu}'\boldsymbol{\Sigma}\boldsymbol{\mu} & \boldsymbol{\mu}'\boldsymbol{\Sigma}\mathbf{l} \\ \mathbf{l}'\boldsymbol{\Sigma}\boldsymbol{\mu} & \mathbf{l}'\boldsymbol{\Sigma}\mathbf{l} \end{pmatrix}.$$

Using Theorem 3.2.10 of Muirhead (1982),

$$\mathbf{l}'\mathbf{A}\boldsymbol{\mu}|\mathbf{l}'\mathbf{Al} \sim N(\mathbf{l}'\mathbf{All}'\boldsymbol{\Sigma}\boldsymbol{\mu}/\mathbf{l}'\boldsymbol{\Sigma}\mathbf{l}, h_{11\cdot 2}\mathbf{l}'\mathbf{Al}), \tag{38}$$

where $h_{11\cdot 2} = \boldsymbol{\mu}'\boldsymbol{\Sigma}\boldsymbol{\mu} - (\mathbf{l}'\boldsymbol{\Sigma}\boldsymbol{\mu})^2/\mathbf{l}'\boldsymbol{\Sigma}\mathbf{l}$. From (37) and (38), the conditional distribution of the right side of (37) given $\mathbf{l}'\mathbf{Al}$ is

$$N(\mathbf{l}'\mathbf{All}'\boldsymbol{\Sigma}\boldsymbol{\mu}/\mathbf{l}'\boldsymbol{\Sigma}\mathbf{l}, (\kappa^{-1} + h_{11\cdot 2})\mathbf{l}'\mathbf{Al}).$$

Since $W = \mathbf{l}'\mathbf{A}\mathbf{l}/\mathbf{l}'\mathbf{\Sigma}\mathbf{l}$ is χ^2_n, the stochastic representation of $\mathbf{l}'\mathbf{A}\mathbf{z}$ is

$$\mathbf{l}'\mathbf{A}\boldsymbol{z} \stackrel{\mathrm{d}}{=} cW\mathbf{l}'\mathbf{\Sigma}\boldsymbol{\mu}/c + \sqrt{cW}u_0$$

where $c = (\kappa^{-1} + \boldsymbol{\mu}'\mathbf{\Sigma}\boldsymbol{\mu})\mathbf{l}'\mathbf{\Sigma}\mathbf{l} - (\mathbf{l}'\mathbf{\Sigma}\boldsymbol{\mu})^2$ and $u_0 \sim N(0, 1)$.

From the Definition 2.1 of Barndorff-Nielsen et al. (1982)[1] , the distribution of $\mathbf{l}'\mathbf{A}\mathbf{z}$ is normal variance-mean mixture with position 0, drift $\mathbf{l}'\mathbf{\Sigma}\boldsymbol{\mu}/c$, structure matrix 1 and mixing distribution F which is the distribution function of cW. On the basis of (2.2) of Barndorff-Nielsen et al. (1982), the characteristic function of $\mathbf{l}'\mathbf{A}\mathbf{z}$ is given by

$$\begin{aligned} \mathrm{E}(\exp(i\theta\mathbf{l}'\mathbf{A}\mathbf{z})) &= \left[1 - 2c\left(i\theta\frac{\mathbf{l}'\mathbf{\Sigma}\boldsymbol{\mu}}{c} + \frac{\theta^2}{2}\right)\right]^{-n/2} \\ &= (1 - 2i\theta\mathbf{l}'\mathbf{\Sigma}\boldsymbol{\mu} + c\theta^2)^{-n/2}. \end{aligned} \tag{39}$$

If we put $\theta = 1$ and $\mathbf{\Omega} = (\kappa^{-1} + \boldsymbol{\mu}'\mathbf{\Sigma}\boldsymbol{\mu})\mathbf{\Sigma} - \mathbf{\Sigma}\boldsymbol{\mu}\boldsymbol{\mu}'\mathbf{\Sigma}$ in (39), then the characteristic function of $\mathbf{A}\mathbf{z}$ is written as

$$\begin{aligned} \mathrm{E}(\exp(i\mathbf{l}'\mathbf{A}\mathbf{z})) &= [1 - 2(i\theta\mathbf{l}'\mathbf{\Sigma}\boldsymbol{\mu} + \mathbf{l}'\mathbf{\Omega}\mathbf{l}/2)]^{-n/2} \\ &= \left[1 - 2|\mathbf{\Omega}|^{1/k}\left(i\mathbf{l}'\frac{\mathbf{\Sigma}\boldsymbol{\mu}}{|\mathbf{\Omega}|^{1/k}} - \mathbf{l}'\frac{\mathbf{\Omega}}{|\mathbf{\Omega}|^{1/k}}\mathbf{l}\right)\right]^{-n/2}. \end{aligned} \tag{40}$$

(40) implies that $\mathbf{A}\mathbf{z}$ follows the k-dimensional normal variance-mean mixture with position $\mathbf{0}$, drift $\mathbf{\Sigma}\boldsymbol{\mu}/|\mathbf{\Omega}|^{1/k}$, structure matrix $\mathbf{\Omega}/|\mathbf{\Omega}|^{1/k}$ and mixing distribution G which is the distribution function of $|\mathbf{\Omega}|^{1/k}W$.

The proof is completed. □

As pointed out in the proof of Theorem 3.1, the distribution of $\mathbf{A}\mathbf{z}$ is a normal variance-mean mixture with position $\mathbf{0}$, drift $(\mathbf{\Omega}/|\mathbf{\Omega}|^{1/k})\mathbf{\Omega}^{-1}\boldsymbol{\delta}$, structure matrix $\mathbf{\Omega}/|\mathbf{\Omega}|^{1/k}$ and mixing distribution G which is the distribution function of $|\mathbf{\Omega}|^{1/k}W$. According to Barndorff-Nielsen et al. (1982), the class of r-dimensional generalized hyperbolic distributions are obtained as r-dimensional normal variance-mean mixtures with position $\boldsymbol{\mu}$, drift $\mathbf{\Delta}\boldsymbol{\beta}$, and structure matrix $\mathbf{\Delta}$ if the mixing distribution is the generalized inverse Gaussian distribution $G(\lambda, \gamma^2, \psi^2)$, where $\psi^2 = \alpha^2 - \boldsymbol{\beta}'\mathbf{\Delta}\boldsymbol{\beta}$ with probability density function

$$f(u) = \frac{(\psi/\gamma)^\lambda}{2K_\lambda(\psi\gamma)}u^{\lambda-1}\exp\left\{-\frac{1}{2}(\gamma^2u^{-1} + \psi^2u)\right\},$$

where $u > 0$, K_λ is the modified Bessel function of the third kind, and

$$\gamma \geq 0, \alpha^2 > \boldsymbol{\beta}'\mathbf{\Delta}\boldsymbol{\beta} \text{ for } \lambda > 0,$$

[1]In Definition 2.1 of Barndorff-Nielsen et al. (1982), the row vector was used. In this paper, the row vector in their original definition is deemed to be replaced the column vector.

$$\gamma > 0, \alpha^2 > \boldsymbol{\beta}'\boldsymbol{\Delta}\boldsymbol{\beta} \text{ for } \lambda = 0,$$
$$\gamma > 0, \alpha^2 \geq \boldsymbol{\beta}'\boldsymbol{\Delta}\boldsymbol{\beta} \text{ for } \lambda < 0.$$

Since χ^2 distribution is a special case of the generalized inverse Gaussian distribution, the distribution of $\mathbf{Az}$ belongs to the the class of k-dimensional generalized hyperbolic distributions. The properties of the generalized hyperbolic distributions are well studied by Blæsild (1981), Blæsild and Jensen (1981), Barndorff-Nielsen et al. (1982), among others. In particular, the generalized hyperbolic distribution has the property of closure under affine transformation, marginalization and conditioning, these properties allow us to obtain

$$\mathbf{LAz} \overset{\mathrm{d}}{=} W\mathbf{L}\boldsymbol{\delta} + \sqrt{W}\mathbf{L}\boldsymbol{\Omega}\mathbf{L}'\mathbf{Z} \tag{41}$$

where $\boldsymbol{\delta} = \boldsymbol{\Sigma}\boldsymbol{\mu}$, $\boldsymbol{\Omega} = (\kappa^{-1} + \boldsymbol{\mu}'\boldsymbol{\Sigma}\boldsymbol{\mu})\boldsymbol{\Sigma} - \boldsymbol{\Sigma}\boldsymbol{\mu}\boldsymbol{\mu}'\boldsymbol{\Sigma}$, $\mathbf{L}$ is a constant matrix with $\operatorname{rank}\mathbf{L} = p \leq k$, $W \sim \chi^2_n$ and $\mathbf{Z} \sim N_p(\mathbf{0}, \mathbf{I}_k)$; they are independent. It is noted that the rank of L is allowed to be equal to the dimension. this admit that we can investigate the joint distribution, although most of existing literature on the product have studied marginal distribution. In contrast to the stochastic representation (35), the computational efficiency of the stochastic representation (41) is considerably high because we need to simulate $p+1$ random variables ; χ^2_n and p-dimensional standard normal vector which are independent. In particular, if we put $p = 1$ in (41), then the stochastic representation becomes,

$$\mathbf{l}'\mathbf{Az} \overset{\mathrm{d}}{=} W\mathbf{l}'\boldsymbol{\Sigma}\boldsymbol{\mu} + \sqrt{cW}u_0 \tag{42}$$

where $c = (\kappa^{-1} + \boldsymbol{\mu}'\boldsymbol{\Sigma}\boldsymbol{\mu})\mathbf{l}'\boldsymbol{\Sigma}\mathbf{l} - (\mathbf{l}'\boldsymbol{\Sigma}\boldsymbol{\mu})^2$, $W \sim \chi^2_n$ and $u_0 \sim N(0, 1)$. In addition, W and u_0 are independent. The stochastic representation (42) indicates that it is enough to simulate only two univariate random variables χ^2_n and standard normal variable for any dimension k. This result speeds up the simulation significantly, especially for larger values of k.

As pointed out in Section 3.1, the distribution of $\mathbf{Az}$ discussed here arises in the distribution of the optimal portfolio weights from Bayesian perspective. Bauder and his co-authors (2018, 2020) estimated the posterior density of the optimal portfolio weights by using Kernel method, while Javed et al. (2021) derived the higher order moments of the sampling distribution of the weights. As mentioned above, the distribution of $\mathbf{Az}$ belongs to the class of generalized hyperbolic distributions. This enables us to use the analytical formulae of the density function and the first four moments for generalized hyperbolic (GH) distributions, which are derived in Blæsild (1981), Blæsild and Jensen (1981), and Barndorff-Nielsen et al. (1982).

Corollary 3.2. *Assume the same conditions as Theorem* 3.1. *The density function of* $\mathbf{Az}$ *is given by*

$$f_{\mathbf{Az}}(\mathbf{x}) = \frac{1}{(2\pi)^{\frac{k}{2}}\Gamma\left(\frac{n}{2}\right)2^{\frac{n}{2}-1}}\left(\frac{\mathbf{x}'\boldsymbol{\Omega}^{-1}\mathbf{x}}{1+\boldsymbol{\delta}'\boldsymbol{\Omega}^{-1}\boldsymbol{\delta}}\right)^{\frac{n-k}{4}}\exp(\boldsymbol{\delta}'\boldsymbol{\Omega}^{-1}\mathbf{x})$$
$$\times K_{\frac{n-k}{2}}\left(\left[(1+\boldsymbol{\delta}'\boldsymbol{\Omega}^{-1}\boldsymbol{\delta})\mathbf{x}'\boldsymbol{\Omega}^{-1}\mathbf{x}\right]^{1/2}\right) \tag{43}$$

The density function generally involves the modified Bessel function of the third kind, which is one of the solutions of the following differential equation;

$$z^2\frac{d^2w}{dz^2} + z\frac{dw}{dz} - (z^2+v^2)w = 0,$$

where z is an arbitrary complex number and v is an arbitrary real number (Abramowitz and Stegun, 1965). In particular, if the index of the modified Bessel function is half-integer, then the following finite representations;

$$K_{i+1/2}(x) = \sqrt{\frac{\pi}{2x}}\exp(-x)\sum_{k=0}^{i}\frac{(k+i)!}{k!(i-k)!}(2x)^{-k},$$
$$K_{\pm 1/2}(x) = \sqrt{\frac{\pi}{2x}}\exp(-x)$$

are available. If $n-k$ is odd in (43), then we can get the finite representation of the density function.

In the next corollary, we provide the expression of the mean, variance, skewness and kurtosis of $\mathbf{l}'\mathbf{Az}$.

Corollary 3.3. *Assume the same conditions as Theorem* 3.1. *Let* $\mathbf{l}'$ *be a constant* $1\times k$ *vector. Then, the mean, variance, skewness and kurtosis of* $\mathbf{l}'\mathbf{Az}$ *are given by*

$$\mathrm{E}[\mathbf{l}'\mathbf{Az}] = nb,$$
$$\mathrm{V}[\mathbf{l}'\mathbf{Az}] = n\left[b^2 + a\left(d+\kappa^{-1}\right)\right],$$
$$\mathrm{Skewness}[\mathbf{l}'\mathbf{Az}] = \frac{2b\left[b^2+3a\left(d+\kappa^{-1}\right)\right]}{\sqrt{n}\left[b^2+a\left(d+\kappa^{-1}\right)\right]^{3/2}},$$
$$\mathrm{Kurtosis}[\mathbf{l}'\mathbf{Az}] = 3\left(1+\frac{2}{n}\right) + \frac{24ab^2\left(d+\kappa^{-1}\right)}{n\left[b^2+a\left(d+\kappa^{-1}\right)\right]^2},$$

where $a = \mathbf{l}'\boldsymbol{\Sigma}\mathbf{l}$, $b = \mathbf{l}'\boldsymbol{\Sigma}\boldsymbol{\mu}$ *and* $d = \boldsymbol{\mu}'\boldsymbol{\Sigma}\boldsymbol{\mu}$.

In general, the distribution of $\mathbf{l}'\mathbf{Az}$ is skewed and heavy tailed. However, if we put $\boldsymbol{\mu} = \mathbf{0}$, then the skewness is 0 and the kurtosis depends only on the degrees of freedom n. Since the second term of the kurtosis is always positive, the kurtosis with $\boldsymbol{\mu} = \mathbf{0}$ is less than the kurtosis with $\boldsymbol{\mu} \neq \mathbf{0}$.

Although Bauder and his co-authors (2018, 2020) have established the asymptotic normality of the product, the asymptotic normality for moderate sample size should not be expected to provide accurate approximations for the distribution of $\mathbf{l}'\mathbf{A}\mathbf{z}$. In addition, it may be difficult to evaluate the cumulative distribution function of $\mathbf{l}'\mathbf{A}\mathbf{z}$ because the density function generally involves the special function. To evaluate and approximate the cumulative distribution function of $\mathbf{l}'\mathbf{A}\mathbf{z}$, we apply Lemma 1 of Yonenaga and Suzukawa (2021) to (42) and we can get

$$P(S_n \le x) = \Phi(x) + n^{-1/2} p_1(x)\phi(x) + n^{-1} p_2(x)\phi(x) + o(n^{-1}), \tag{44}$$

where

$$\begin{aligned}
S_n &= \sqrt{n}\frac{\mathbf{l}'\mathbf{A}\mathbf{z}/n - \mathbf{l}'\boldsymbol{\Sigma}\boldsymbol{\mu}}{\sqrt{2(\mathbf{l}'\boldsymbol{\Sigma}\boldsymbol{\mu})^2 + c}}, \\
p_1(x) &= -\frac{1}{6}\tilde{\kappa}_3(x^2 - 1),\ p_2(x) = -x\left\{\frac{1}{24}\tilde{\kappa}_4(x^2 - 3) + \frac{1}{72}\tilde{\kappa}_3^2(x^4 - 10x^2 + 15)\right\} \\
\tilde{\kappa}_3 &= \frac{2\mathbf{l}'\boldsymbol{\Sigma}\boldsymbol{\mu}[3c + 4(\mathbf{l}'\boldsymbol{\Sigma}\boldsymbol{\mu})^2]}{[2(\mathbf{l}'\boldsymbol{\Sigma}\boldsymbol{\mu})^2 + c]^{3/2}},\ \tilde{\kappa}_4 = \frac{6[c^2 + 8c(\mathbf{l}'\boldsymbol{\Sigma}\boldsymbol{\mu})^2 + 8(\mathbf{l}'\boldsymbol{\Sigma}\boldsymbol{\mu})^4]}{[2(\mathbf{l}'\boldsymbol{\Sigma}\boldsymbol{\mu})^2 + c]^2}
\end{aligned}$$

We can observe from (44) that if $\boldsymbol{\mu} = \mathbf{0}$, the term of order $n^{-s/2}$ in the expansion is vanished, where s is odd. This implies that if we have $\boldsymbol{\mu} = \mathbf{0}$, the convergent to standard normal distribution become fast.

Remark 3.4. According to Fung and Seneta (2011), the distribution of the product belongs to the class of multivariate skewed variance-gamma (VG) distributions, which is obtained when the mixture distribution of the GH distributions is a gamma distributed.

Remark 3.5. As mentioned above, the obtained stochastic representation implies that the distribution of $\mathbf{A}\mathbf{z}$ belongs to the class of GH distributions. This fact gives us an insight of the behavior of the tail of the bivariate distribution of the product. Fung and Seneta (2011) showed that the bivariate GH distribution with the correlation $\rho \in (-1, 1)$ is asymptotically independent in the lower tail. Furthermore, let $\boldsymbol{\beta}$ be a two dimensional constant vector, and assume that its components are equal to θ and $L(u)$ be the slowing varying function which satisfies

$$\lim_{u \to 0+} \frac{L(\lambda u)}{L(u)} = 1,$$

where λ is an arbitrary positive real number (Seneta, 1976).

If $\mathbf{X}$ is GH distribution, i.e.

$$X \stackrel{\mathrm{d}}{=} W\boldsymbol{\beta} + \sqrt{W} Z,$$

where $W \sim GIG(\lambda, \delta, \kappa)$ with $\kappa > 0$, $Z \sim N_2(0, R)$, and $\mathbf{Z}$ is independent of W, and $\mathbf{R}$ is a correlation matrix with the component ρ, then Fung and Seneta (2021) derived

$$C(u,u) \sim u^{\tau} L(u),$$
$$\tau = \frac{(1+\delta^2)^{1/2}[((1+\delta^2)\theta^2+\kappa^2)^{1/2}+(1+\delta^2)^{1/2}\theta]}{(\theta^2+\kappa^2)^{1/2}+\theta},$$
$$L(u) \sim K(\rho,\theta,\lambda,\delta,\kappa)(|\log u|)^{(\lambda-1)(1-\tau)-1/2}, u \to 0+,$$

where $K(\rho, \theta, \lambda, \delta, \kappa)$ is a specific function of all the parameters, and $C(u, v)$ is the copula of $\mathbf{X}$. As for the copula, readers may refer to Nelsen (2006).

3.3 Simulation

In the foregoing section, we have derived the stochastic representation of $\mathbf{Az}$, which is different from the existing stochastic representation given by (35). In this section, we investigate whether the stochastic representation (41) yields the same distribution as the stochastic representation (35). The statistical software R is used to create all graphs and the seed value is `set.seed(1)`. The simulation is made for $p = 1$. We use two data sets in the simulation. The first one is obtained by using the stochastic representation (35), whereas the elements from the second data are obtained by using the stochastic representation (41). Each of the simulated data consists of $B = 3000$ independent realizations. The first data set is abbreviated by data 1 and it is obtained in the following way:

(a) generate independently $\eta \sim \chi^2_{n+1}$, $z_0 \sim N(0,1)$, $Q \sim F(k, n-k+1)$ and $\mathbf{Z} \sim N_2(\mathbf{0}, \mathbf{I}_2)$ with $\mathbf{U} = \mathbf{Z}/\sqrt{\mathbf{Z}'\mathbf{Z}}$;

(b) compute ϵ, ζ and Υ, respectively;

(c) put the values obtained in step (b) into the stochastic representation

$$\mathbf{l}'\mathbf{Az} \overset{\mathrm{d}}{=} \eta \mathbf{l}'\boldsymbol{\zeta} + \sqrt{\eta}(\epsilon \mathbf{l}'\boldsymbol{\Upsilon}\mathbf{l} - \mathbf{l}'\boldsymbol{\zeta}\boldsymbol{\zeta}'\mathbf{l})^{1/2} z_0;$$

(d) repeat step (a)-(c) B times.

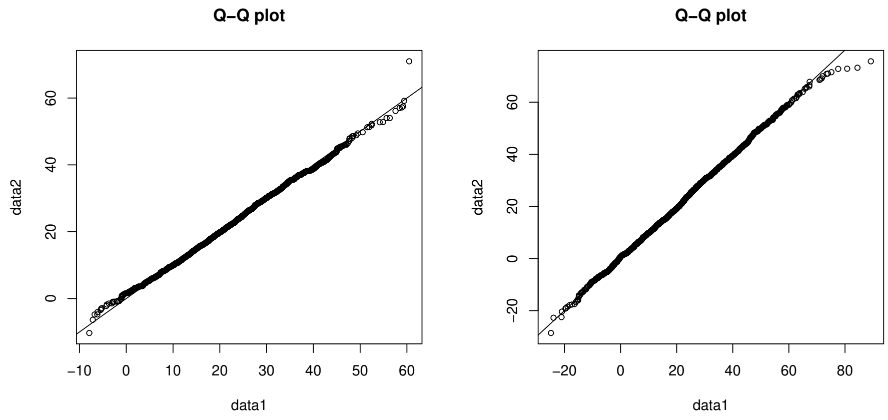
The second data set is denoted by data 2 and the corresponding algorithm is given next:

(a) generate independently $W \sim \chi^2_n$ and $u_0 \sim N(0,1)$;

(b) compute the value

$$\mathbf{l}'\mathbf{A}\boldsymbol{z} \overset{\mathrm{d}}{=} W\mathbf{l}'\boldsymbol{\Sigma}\boldsymbol{\mu} + \sqrt{(\kappa^{-1} + \boldsymbol{\mu}'\boldsymbol{\Sigma}\boldsymbol{\mu})\mathbf{l}'\boldsymbol{\Sigma}\mathbf{l} - (\mathbf{l}'\boldsymbol{\Sigma}\boldsymbol{\mu})^2}\sqrt{W} u_0;$$

(c) repeat step (a)-(b) B times.



(a) (b)

Figure 5: The both figures show that Q-Q plot of data 1 and data 2 where these data sets are drawn from the stochastic representation (35) and (41), respectively. The left panel deals with $\mathbf{A} \sim W_2(20, \mathbf{I}_2)$, $\mathbf{z} \sim N_2((1,0)', \mathbf{A}^{-1})$ and $\mathbf{l}' = (1, -1)$, whereas the right panel $\mathbf{A} \sim W_5(20, \mathbf{I}_5)$, $\mathbf{z} \sim N_5((1,0,0,0,0)', \mathbf{A}^{-1})$ and $\mathbf{l}' = (1, -1, 1, -1, 1)$. The slope of the real line in both figures is 1.

The Q-Q plot for $\mathbf{A} \sim W_2(20, \mathbf{I}_2)$, $\mathbf{z} \sim N_2((1,0)', \mathbf{A}^{-1})$ and $\mathbf{l}' = (1, -1)$ is presented in Figure 5a, and for $\mathbf{A} \sim W_5(20, \mathbf{I}_5)$, $\mathbf{z} \sim N_5((1,0,0,0,0)', \mathbf{A}^{-1})$ and $\mathbf{l}' = (1, -1, 1, -1, 1)$ in Figure 5b. We can observe from Figure 5 that quantiles of data 1 almost coincide with those of data 2, which means two data sets are drawn from the same distribution.

3.4 Summary

Previous researches on functions of a Wishart matrix and a normal vector have assumed the independence of a Wishart matrix and a normal vector. However, assuming $\mathbf{A} \sim W_k(n, \mathbf{\Sigma})$ with $\mathbf{\Sigma} > 0$ and $\mathbf{z}|\mathbf{A} \sim N_k(\boldsymbol{\mu}, \mathbf{A}^{-1}/\kappa)$ appears natural from the viewpoint of Bayesian statistics. Bauder and his co-authors (2018, 2020) derived the stochastic representations of the product, which is computationally efficient. In addition, they provided the first two moments and established asymptotic normality of the product by using the stochastic representations. Since their stochastic representations could not be mathematically tractable, we derive the novel stochastic representation of the product in Theorem 3.1. The derived stochastic representation is very simple and highly computationally efficient compared with the existing stochastic representation. In addition, it turns out that the distribution of **Az** is closed under affine transformation, marginalization and conditioning. In Corollary 3.2 and 3.3, the explicit expression of the density function and first four moments of the product are provided. We also discuss the asymptotic expansion for $\mathbf{l}'\mathbf{Az}$. The results of the present study will contribute to develop the distributional properties of the product of a Wishart matrix and a normal vector which are dependent, and improve on existing stochastic representation in computational efficiency of the stochastic representation of the product.

Appendix

As mentioned in section4.2, we comment on the results obtained by Bauder et al. (2019). By taking the approach adopted in this section, we can improve the computational efficiency of stochastic representations given in Bauder et al. (2019). This is outside the scope of this paper, but we include the results and its proofs due to its importance.

The three parameters of the efficient frontier can be expressed as follows:

$$R_{GMV} = \frac{\mathbf{1}'\mathbf{Az}}{\mathbf{1}'\mathbf{A1}},$$
$$V_{GMV} = \frac{1}{\mathbf{1}'\mathbf{A1}},$$
$$s = \mathbf{z}'\mathbf{Az} - \frac{(\mathbf{1}'\mathbf{Az})^2}{\mathbf{1}'\mathbf{A1}},$$

where $A \sim W_k(n, \mathbf{\Sigma})$ and $z \sim N_k(\boldsymbol{\mu}, \mathbf{A}^{-1}/\kappa)$.

Theorem 3.6. *Let $A \sim W_k(n, \mathbf{\Sigma})$ with $\mathbf{\Sigma} > 0$ and $z \sim N_k(\boldsymbol{\mu}, \mathbf{A}^{-1}/\kappa)$. The stochastic representation of R_{GMV}, V_{GMV} and s are given by*

$$V_{GMV} \overset{\mathrm{d}}{=} \frac{1}{\mathbf{1}'\mathbf{\Sigma 1}\chi_n^2},$$
$$R_{GMV} \overset{\mathrm{d}}{=} \frac{\mathbf{1}'\mathbf{\Sigma}\boldsymbol{\mu}}{\mathbf{1}'\mathbf{\Sigma 1}} + \sqrt{\frac{\kappa^{-1} + \sigma_{11\cdot 2}}{n\mathbf{1}'\mathbf{\Sigma 1}}}t_n,$$
$$s \overset{\mathrm{d}}{=} \frac{\chi_{k-1}^2(\kappa\sigma_{11\cdot 2}\chi_{n-1}^2)}{\kappa},$$

where $\sigma_{11\cdot 2} = \boldsymbol{\mu}'\mathbf{\Sigma}\boldsymbol{\mu} - (\mathbf{1}'\mathbf{\Sigma}\boldsymbol{\mu})^2/\mathbf{1}'\mathbf{\Sigma 1}$. In addition, χ_n^2, $\chi_n^2(\delta^2)$, and t_n denote the random variables which are distributed as χ^2 distribution with degrees of freedom n, noncentral χ^2 distribution with degrees of freedom n and non centrality parameter δ^2, and t distribution with n degrees of freedom, respectively.

Proof. Firstly, we consider the distribution of V_{GMV}, which is the variance of the GMV portfolio. This immediately follows from Theorem 3.2.8 of Muirhead (1982) that

$$V_{GMV} \overset{\mathrm{d}}{=} \frac{1}{\mathbf{1}'\mathbf{\Sigma 1}\xi},$$

where $\xi \sim \chi_n^2$.

Next, we consider the distribution of R_{GMV}, which is the expected return of the GMV portfolio. Dividing both sides of (37) with $\mathbf{l} = \mathbf{1}$ by $\mathbf{1}'\mathbf{A}\mathbf{1}$, we obtain

$$\frac{\mathbf{1}'\mathbf{A}\mathbf{z}}{\mathbf{1}'\mathbf{A}\mathbf{1}} \overset{\mathrm{d}}{=} \frac{\mathbf{z}_0}{\sqrt{\mathbf{1}'\mathbf{A}\mathbf{1}}} + \frac{\mathbf{1}'\mathbf{A}\boldsymbol{\mu}}{\mathbf{1}'\mathbf{A}\mathbf{1}},$$

where $\mathbf{z}_0 \sim N(0, 1)$. From (38), we get

$$\frac{\mathbf{1}'\mathbf{A}\boldsymbol{\mu}}{\mathbf{1}'\mathbf{A}\mathbf{1}} | \mathbf{1}'\mathbf{A}\mathbf{1} \sim N\left(\frac{\mathbf{1}'\boldsymbol{\Sigma}\boldsymbol{\mu}}{\mathbf{1}'\boldsymbol{\Sigma}\mathbf{1}}, \frac{\sigma_{11\cdot 2}}{\mathbf{1}'\mathbf{A}\mathbf{1}}\right), \tag{45}$$

where $\sigma_{11\cdot 2} = \boldsymbol{\mu}'\boldsymbol{\Sigma}\boldsymbol{\mu} - (\mathbf{1}'\boldsymbol{\Sigma}\boldsymbol{\mu})^2/\mathbf{1}'\boldsymbol{\Sigma}\mathbf{1}$.

Therefore, the conditional distribution of $\mathbf{1}'\mathbf{A}\mathbf{z}/\mathbf{1}'\mathbf{A}\mathbf{1}$ given $\mathbf{1}'\mathbf{A}\mathbf{1}$ is

$$N\left(\frac{\mathbf{1}'\boldsymbol{\Sigma}\boldsymbol{\mu}}{\mathbf{1}'\boldsymbol{\Sigma}\mathbf{1}}, \frac{\kappa^{-1} + \sigma_{11\cdot 2}}{\mathbf{1}'\mathbf{A}\mathbf{1}}\right).$$

Since $W = \mathbf{1}'\mathbf{A}\mathbf{1}/\mathbf{1}'\boldsymbol{\Sigma}\mathbf{1}$ is χ^2_n, the stochastic representation of $\mathbf{1}'\mathbf{A}\mathbf{z}/\mathbf{1}'\mathbf{A}\mathbf{1}$ is

$$\begin{aligned}\frac{\mathbf{1}'\mathbf{A}\mathbf{z}}{\mathbf{1}'\mathbf{A}\mathbf{1}} &\overset{\mathrm{d}}{=} \frac{\mathbf{1}'\boldsymbol{\Sigma}\boldsymbol{\mu}}{\mathbf{1}'\boldsymbol{\Sigma}\mathbf{1}} + \sqrt{\frac{\kappa^{-1} + \sigma_{11\cdot 2}}{n\mathbf{1}'\boldsymbol{\Sigma}\mathbf{1}}} \frac{u_0}{\sqrt{W/n}} \\ &\overset{\mathrm{d}}{=} \frac{\mathbf{1}'\boldsymbol{\Sigma}\boldsymbol{\mu}}{\mathbf{1}'\boldsymbol{\Sigma}\mathbf{1}} + \sqrt{\frac{\kappa^{-1} + \sigma_{11\cdot 2}}{n\mathbf{1}'\boldsymbol{\Sigma}\mathbf{1}}} t_n,\end{aligned}$$

where t_n is a t-distribution with n degrees of freedom.

Finally, we consider the distribution of s which is a slope parameter of the efficient frontier. If we fix $\mathbf{A}$,and put $\sqrt{\kappa}\mathbf{A}^{1/2}\mathbf{z} = \mathbf{t}$ in the equation of s, then

$$s = \frac{1}{\kappa}\mathbf{t}'\left(\mathbf{I}_k - \frac{\mathbf{A}^{1/2}\mathbf{1}\mathbf{1}'\mathbf{A}^{1/2}}{\mathbf{1}'\mathbf{A}\mathbf{1}}\right)\mathbf{t} = \frac{\mathbf{t}'\mathbf{B}\mathbf{t}}{\kappa},$$

where $\mathbf{t}|\mathbf{A} \sim N_k(\sqrt{\kappa}\mathbf{A}^{1/2}\boldsymbol{\mu}, \mathbf{I}_k)$.

Since we can confirm the following identity holds:

(i) $\mathrm{tr}(\mathbf{B}) = k - 1,\ \kappa\boldsymbol{\mu}'\mathbf{A}^{1/2}\mathbf{B}\mathbf{A}^{1/2}\boldsymbol{\mu} = \kappa\left(\boldsymbol{\mu}'\mathbf{A}\boldsymbol{\mu} - \frac{\boldsymbol{\mu}'\mathbf{A}\mathbf{1}\mathbf{1}'\mathbf{A}\boldsymbol{\mu}}{\mathbf{1}'\mathbf{A}\mathbf{1}}\right)$,

(ii) $\mathbf{B}^2 = \mathbf{B}$,

(iii) $\kappa\boldsymbol{\mu}'\mathbf{A}^{1/2}\mathbf{B}^2\mathbf{A}^{1/2}\boldsymbol{\mu} = \kappa\boldsymbol{\mu}'\mathbf{A}^{1/2}\mathbf{B}\mathbf{A}^{1/2}\boldsymbol{\mu}$,

(iv) $\boldsymbol{\mu}'\mathbf{A}^{1/2}\mathbf{B}^2 = \boldsymbol{\mu}'\mathbf{A}^{1/2}\mathbf{B}$,

it follows from Theorem 5.1.3 of Mathai and Provost (1992) that the conditional distribution of s given $\mathbf{A}$ is $\chi^2_{k-1}(\delta^2)/\kappa$, where

$$\delta^2 = \kappa\left(\boldsymbol{\mu}'\mathbf{A}\boldsymbol{\mu} - \frac{\boldsymbol{\mu}'\mathbf{A}\mathbf{1}\mathbf{1}'\mathbf{A}\boldsymbol{\mu}}{\mathbf{1}'\mathbf{A}\mathbf{1}}\right).$$

From Theorem 3.2.10 of Muirhead, we can obtain $\delta^2 \sim \kappa\sigma_{11\cdot 2}\chi^2_{n-1}$. □

From the stochastic representations we have derived, we can see that the variance and expected return of the GMV portfolio are given by one random variable, and the slope parameter of the efficient frontier is given by two random variables. This is extremely computationally efficient compared to Theorem 1 given in Bauder et al. (2019).

4 Bayesian estimation for misclassification rate in linear discriminant analysis

4.1 Introduction

In this section, we consider discriminant analysis in the case of two multivariate normal populations with different means and common covariance matrices, namely, π_1 : $N_p(\boldsymbol{\mu}_1, \boldsymbol{\Sigma})$ and π_2 : $N_p(\boldsymbol{\mu}_2, \boldsymbol{\Sigma})$ where $\boldsymbol{\mu}_1$ is the vector of means of the i-th population, $i = 1, 2$, and $\boldsymbol{\Sigma} > 0$ is the matrix of variances and covariances of each population.

Suppose an individual $\mathbf{y}$ is an observation from either π_1 or π_2. Our classifications rule is to classify $\mathbf{y}$ to π_1 if

$$U = (\boldsymbol{\mu}_1 - \boldsymbol{\mu}_2)'\boldsymbol{\Sigma}^{-1}\left\{\mathbf{y} - \frac{1}{2}(\boldsymbol{\mu}_1 + \boldsymbol{\mu}_2)\right\} \geq x$$

is used as the best criterion of discrimination when we wish to classify an observation $\mathbf{y}$ as coming from π_1 to π_2. If q_1 and q_2 are a priori probabilities of drawing an observation from π_1 and π_2, respectively, and if the cost of misclassifying an observation from π_1 as from π_2 is $C(2|1)$ and an observation from π_2 as from π_1 is $C(1|2)$, then the x is given by $\log(q_2 C(1|2)/q_1 C(2|1))$. If $q_1 = q_2$ and $C(1|2) = C(2|1)$, then the cut-off point x reduces to 0.

In discriminant analysis, it is very important to evaluate misclassification rates. There are two misclassification rates: the optimal error rate and the actual error rate. If $\boldsymbol{\mu}_1$, $\boldsymbol{\mu}_2$ and $\boldsymbol{\Sigma}$ are fixed and $\mathbf{y} \in \pi_1$, the optimal error rate is defined as

$$\epsilon_1 = \epsilon_1(\boldsymbol{\mu}_1, \boldsymbol{\mu}_2, \boldsymbol{\Sigma}) = Pr[U < x|\boldsymbol{\mu}_1, \boldsymbol{\mu}_2, \boldsymbol{\Sigma}, \mathbf{y} \in \pi_1] = \Phi\left(\frac{x - \Delta^2/2}{\Delta}\right), \tag{46}$$

where $\Phi(\cdot)$ denotes the cumulative distribution function of the standard normal distribution, and Δ^2 denotes the square Mahalanobis distance between π_1 and π_2, which is defined by $\Delta^2 = (\boldsymbol{\mu}_1 - \boldsymbol{\mu}_2)'\boldsymbol{\Sigma}^{-1}(\boldsymbol{\mu}_1 - \boldsymbol{\mu}_2)$.

In most applications the population discriminant U is not known, and it is necessary to estimate them. Suppose we collect two mutually independent random samples $\mathbf{x}_{11}, \cdots, \mathbf{x}_{1N_1}$ and $\mathbf{x}_{21}, \cdots, \mathbf{x}_{2N_2}$ drawn from π_1 and π_2, respectively. Our estimate of $\boldsymbol{\mu}_1$ is $\bar{\mathbf{x}}_1 = \sum_{\alpha=1}^{N_1} \mathbf{x}_{1\alpha}/N_1$, of $\boldsymbol{\mu}_2$ is $\bar{\mathbf{x}}_2 = \sum_{\alpha=1}^{N_2} \mathbf{x}_{2\alpha}/N_2$, and of $\boldsymbol{\Sigma}$ is $\mathbf{S}$ defined by $n\mathbf{S} = \sum_{\alpha=1}^{N_1}(\mathbf{x}_{1\alpha} - \bar{\mathbf{x}}_1)(\mathbf{x}_{1\alpha} - \bar{\mathbf{x}}_1)' + \sum_{\alpha=1}^{N_2}(\mathbf{x}_{2\alpha} - \bar{\mathbf{x}}_2)(\mathbf{x}_{2\alpha} - \bar{\mathbf{x}}_2)'$ with $n = N_1 + N_2 - 2$. We substitute these estimates for the parameters in the population discriminant U to obtain

$$W = (\bar{\mathbf{x}}_1 - \bar{\mathbf{x}}_2)'\mathbf{S}^{-1}\left\{\mathbf{y} - \frac{1}{2}(\bar{\mathbf{x}}_1 + \bar{\mathbf{x}}_2)\right\},$$

Although we assume $N_i \geq p$ to take the inverse of $\mathbf{S}$, the opposite case, i.e. $N_i < p$, should deal with the inverse of the singular Wishart matrix. The distributional properties of (inverse) singular Wishart distribution are well studied by Díaz-García et al. (1997), Srivastava (2003), Bodnar and Okhrin (2008), Bodnar and his co-authors (2015, 2016, 2019), among others. In particular, Bodnar and Okhrin (2008), and Bodnar et al. (2016) discussed the distribution of the generalized inverse of the singular Wishart matrix in detail.

We refer to the misclassification rate associated with W as the actual error rate. Since it is very important to evaluate the misclassification rate computed based on the distribution of W, we need the exact distribution of W. However, according to Anderson (2003), the exact distribution of W is extremely complicated and very difficult to calculate. As a result, a great deal of attention has been paid to deriving an asymptotic expansion for the distribution function of W. For example, Okamoto (1963) obtained the asymptotic expansion of the distribution function of W to terms of order n^{-2}, and Siotani and Wang (1977) to terms of n^{-3}, where $n = N_1+N_2-2$. From the viewpoint of computational statistics, Bowker (1961) derived the stochastic representation of W in terms of elements of a p-dimensional normal vector and those of a 2-dimensional Wishart matrix. In recent years, Bodnar et al. (2020) established the stochastic representation of W with some univariate random variables and presented the computation routine of the actual error rate. In addition, they proved asymptotic normality under the double regime asymptotics.

In contrast to frameworks of sampling distribution theory, it is often the case that an explicit expression of the Bayes estimator of the actual error rate is available in Bayesian linear discriminant analysis. The conditional actual error rate is given by

$$\beta_1 = \beta_1(\boldsymbol{\mu}_1, \boldsymbol{\mu}_2, \boldsymbol{\Sigma}) = Pr[W < x|\boldsymbol{\mu}_1, \boldsymbol{\mu}_2, \boldsymbol{\Sigma}^{-1}, \mathbf{y} \in \pi_1] = \Phi(\theta_1),$$

where

$$\theta_1 = \frac{[(\bar{\mathbf{x}}_1 - \bar{\mathbf{x}}_2)'\mathbf{S}^{-1}\{(\bar{\mathbf{x}}_1 + \bar{\mathbf{x}}_2)/2 - \boldsymbol{\mu}_1\} + x]}{\sqrt{(\bar{\mathbf{x}}_1 - \bar{\mathbf{x}}_2)'\mathbf{S}^{-1}\boldsymbol{\Sigma}\mathbf{S}^{-1}(\bar{\mathbf{x}}_1 - \bar{\mathbf{x}}_2)}}.$$

It is noted that β_1 is the function of the random variables $\boldsymbol{\mu}_1$, $\boldsymbol{\mu}_2$ and $\boldsymbol{\Sigma}$. Under the Jeffreys prior for $\boldsymbol{\mu}_1$, $\boldsymbol{\mu}_2$ and $\boldsymbol{\Sigma}^{-1}$ and normal sampling model, Geisser (1967) showed that the Bayes estimator of β_1, i.e., the posterior mean of β_1, is given by

$$Pr\left[t_{n+1-p} \leq \frac{x - Q/2}{\sqrt{n(N_1+1)Q/(n+1-p)N_1}}\right],$$

where $Q = (\bar{\mathbf{x}}_1 - \bar{\mathbf{x}}_2)'\mathbf{S}^{-1}(\bar{\mathbf{x}}_1 - \bar{\mathbf{x}}_2)$, $n = N_1 + N_2 - 2$ and t_{n+1-p} is the t-distributed random variable with degrees of freedom $n+1-p$. Dalton and Dougherty (2011a, 2011b) proposed the minimum mean square error (MMSE) estimation theory to estimate the true error associated with the sample discriminant function conditioned on training data sets. In addition, they applied the MMSE estimation theory to estimate the misclassification rate associated with a linear discriminant $\mathbf{a}'\mathbf{y} + b$, where $\mathbf{a}$ is an arbitrary constant vector, while b is an arbitrary constant. It is noted that the proposed estimator of the actual error has the same functional form as the classical Bayes estimator (cf. Lehmann and Casella, 1998).

Whereas many studies have considered the estimation of the actual error rate, few have explored the estimation of the optimal error rate ϵ_1 defined by (46). According to Geisser (1982), the optimal error rate is useful as a criterion for the selection of variables used for discrimination. If the estimate of the optimal error rate were larger than necessary, one would search either for additional variables or another set with which to diminish the rate. Therefore, the purpose of this study is to consider the Bayesian estimation of ϵ_1. This remainder of this paper is structured as follows. In section 4.2, we present the posterior distributions for $\boldsymbol{\mu}_1$, $\boldsymbol{\mu}_2$ and $\boldsymbol{\Sigma}^{-1}$ under conjugate and diffuse priors. In section 4.3, we give the expression of the posterior predictive density and the first four moments of the linear discriminant U under both the priors. Using the expression of the posterior predictive density of U, we derive the Bayes estimator of ϵ_1. Although the expression is somewhat complicated, there are some cases in which the expression of the estimator is simply expressed. Moreover, we suggest approximations for the Bayes estimator of ϵ_1. In section 4.4, based on simulation studies, we investigate the accuracy of the approximations for the Bayes estimator of ϵ_1. Through the simulation results, we document the good performance of the suggested approximations. In section 4.5, we apply the results to estimate the optimal error rate associated with Fisher's Iris dataset. Section 4.6 contains our conclusions. All proofs are presented in the Appendix.

4.2 Preliminaries

The main purpose of this work is the Bayesian estimation of ϵ_1. In Bayesian estimation, we must choose prior distributions for $\boldsymbol{\mu}_1$, $\boldsymbol{\mu}_2$ and $\boldsymbol{\Sigma}^{-1}$. As prior distributions for $\boldsymbol{\mu}_1$, $\boldsymbol{\mu}_2$ and $\boldsymbol{\Sigma}^{-1}$, the Jeffreys prior and normal-inverse-Wishart prior (NIW prior) have been employed in the literature of Bayesian discrimination (Geisser, 1967, 1982; Dalton and Dougherty, 2011). The Jeffreys prior is given by

$$\pi_d(\boldsymbol{\mu}_1, \boldsymbol{\mu}_2, \boldsymbol{\Sigma}^{-1}) \propto |\boldsymbol{\Sigma}|^{(p+1)/2}, \tag{47}$$

and the NIW prior assumes a normal distribution for $\boldsymbol{\mu}_1$ and $\boldsymbol{\mu}_2$ and a Wishart distribution for $\boldsymbol{\Sigma}^{-1}$;

$$\boldsymbol{\mu}_1|\Sigma \sim N_p\left(\boldsymbol{\xi}_1, \Sigma/k_1\right), \ \boldsymbol{\mu}_2|\Sigma \sim N_p\left(\boldsymbol{\xi}_2, \Sigma/k_2\right), \ \boldsymbol{\Sigma}^{-1} \sim W_p(m_0, \boldsymbol{\Lambda}_0^{-1}), \tag{48}$$

where $\boldsymbol{\xi}_i$ is a prior mean, k_i is a positive real number, m_0 is a positive integer with $m_0 \geq p$, and Λ_0 is a known prior matrix of $\boldsymbol{\Sigma}$. Moreover, $\boldsymbol{\mu}_1$ and $\boldsymbol{\mu}_2$ (conditional on $\boldsymbol{\Sigma}$) are conditionally independent. Let $\mathbf{x}_{11}, \cdots, \mathbf{x}_{1N_1}$ and $\mathbf{x}_{21}, \cdots, \mathbf{x}_{2N_2}$ be two independent samples from the multivariate normal distribution that consist of independent and identically distributed random vectors with $\mathbf{x}_{1i} \sim N_p(\boldsymbol{\mu}_1, \boldsymbol{\Sigma})$ for $i = 1, \cdots, N_1$ and $\mathbf{x}_{2i} \sim N_p(\boldsymbol{\mu}_2, \boldsymbol{\Sigma})$ for $i = 1, \cdots, N_2$. Let $\mathbf{X}_1 = (\mathbf{x}_{11}, \cdots, \mathbf{x}_{1N_1})$ and $\mathbf{X}_2 = (\mathbf{x}_{21}, \cdots, \mathbf{x}_{2N_2})$ be observation matrices.

Under the Jeffreys prior, the posterior distributions for $\boldsymbol{\mu}_1$, $\boldsymbol{\mu}_2$ and $\boldsymbol{\Sigma}^{-1}$ are given by

$$\begin{aligned}
&\boldsymbol{\mu}_1|\boldsymbol{\Sigma}, \mathbf{X}_1 \sim N_p\left(\bar{\mathbf{x}}_1, \boldsymbol{\Sigma}/N_1\right), \ \boldsymbol{\mu}_2|\boldsymbol{\Sigma}, \mathbf{X}_2 \sim N_p\left(\bar{\mathbf{x}}_2, \boldsymbol{\Sigma}/N_2\right), \\
&\boldsymbol{\Sigma}^{-1}|\mathbf{X}_1, \mathbf{X}_2 \sim W_p(n, (n\mathbf{S})^{-1}),
\end{aligned}$$

where $\mathbf{1}_n$ denotes the vector of ones,

$$\begin{aligned}
\bar{\mathbf{x}}_1 &= \frac{\mathbf{1}'_{N_1}\mathbf{X}_1}{N_1}, \ \bar{\mathbf{x}}_2 = \frac{\mathbf{1}'_{N_2}\mathbf{X}_2}{N_2}, \ n = N_1 + N_2 - 2, \\
(N_1 - 1)\mathbf{S}_1 &= \sum_{j=1}^{N_1}(\mathbf{x}_{1j} - \bar{\mathbf{x}}_1)(\mathbf{x}_{1j} - \bar{\mathbf{x}}_1)', \ (N_2 - 1)\mathbf{S}_2 = \sum_{j=1}^{N_2}(\mathbf{x}_{2j} - \bar{\mathbf{x}}_2)(\mathbf{x}_{2j} - \bar{\mathbf{x}}_2)', \\
n\mathbf{S} &= (N_1 - 1)\mathbf{S}_1 + (N_2 - 1)\mathbf{S}_2.
\end{aligned}$$

Under the NIW prior,

$$\begin{aligned}
&\boldsymbol{\mu}_1|\boldsymbol{\Sigma}, \mathbf{X}_1 \sim N_p\left(\boldsymbol{\omega}_1, \frac{\boldsymbol{\Sigma}}{k_1 + N_1}\right), \ \boldsymbol{\mu}_2|\boldsymbol{\Sigma}, \mathbf{X}_2 \sim N_p\left(\boldsymbol{\omega}_2, \frac{\boldsymbol{\Sigma}}{k_2 + N_2}\right), \\
&\boldsymbol{\Sigma}^{-1}|\mathbf{X}_1, \mathbf{X}_2 \sim W_p(m, \boldsymbol{\Lambda}^{-1}),
\end{aligned}$$

where

$$\begin{aligned}
\boldsymbol{\omega}_1 &= \frac{k_1}{k_1 + N_1}\boldsymbol{\xi}_1 + \frac{N_1}{k_1 + N_1}\bar{\mathbf{x}}_1, \ \boldsymbol{\omega}_2 = \frac{k_2}{k_2 + N_2}\boldsymbol{\xi}_2 + \frac{N_2}{k_2 + N_2}\bar{\mathbf{x}}_2, \ m = m_0 + N_1 + N_2, \\
\boldsymbol{\Lambda} &= \boldsymbol{\Lambda}_0 + n\mathbf{S} + \frac{k_1 N_1}{k_1 + N_1}(\bar{\mathbf{x}}_1 - \boldsymbol{\xi}_1)(\bar{\mathbf{x}}_1 - \boldsymbol{\xi}_1)' + \frac{k_2 N_2}{k_2 + N_2}(\bar{\mathbf{x}}_2 - \boldsymbol{\xi}_2)(\bar{\mathbf{x}}_2 - \boldsymbol{\xi}_2)'.
\end{aligned}$$

4.3 Main results

4.3.1 Exact expressions of the Bayes estimator of the optimal error rate

We consider the Bayesian estimation of ϵ_1. Let $P(\boldsymbol{\mu}_1, \boldsymbol{\mu}_2, \boldsymbol{\Sigma}^{-1}|\boldsymbol{X}_1, \boldsymbol{X}_2)$ be the joint posterior distribution for $\boldsymbol{\mu}_1$, $\boldsymbol{\mu}_2$ and $\boldsymbol{\Sigma}^{-1}$, $f(u|\boldsymbol{\mu}_1, \boldsymbol{\mu}_2, \boldsymbol{\Sigma}^{-1}, \mathbf{y} \in \pi_1)$ be the conditional normal density of U with the mean $\Delta^2/2$ and variance Δ^2, and $f(u|\mathbf{y} \in \pi_{\mathbf{1}})$ be the posterior predictive density of U. The Bayes estimator of ϵ_1 is then computed as

$$\begin{aligned}
\mathrm{E}(\epsilon_1) &= \int_{\boldsymbol{\mu}_1, \boldsymbol{\mu}_2, \boldsymbol{\Sigma}^{-1}} \Phi\left(\frac{x - \Delta^2/2}{\Delta}\right) P(\boldsymbol{\mu}_1, \boldsymbol{\mu}_2, \boldsymbol{\Sigma}^{-1}|\mathbf{X}_1, \mathbf{X}_2) d\boldsymbol{\mu}_1 d\boldsymbol{\mu}_2 d\boldsymbol{\Sigma}^{-1} \\
&= \int_{-\infty}^{x} \int_{\boldsymbol{\mu}_1, \boldsymbol{\mu}_2, \boldsymbol{\Sigma}^{-1}} f(u|\boldsymbol{\mu}_1, \boldsymbol{\mu}_2, \boldsymbol{\Sigma}^{-1}, \mathbf{y} \in \pi_1) P(\boldsymbol{\mu}_1, \boldsymbol{\mu}_2, \boldsymbol{\Sigma}^{-1}|\mathbf{X}_1, \mathbf{X}_2) d\boldsymbol{\mu}_1 d\boldsymbol{\mu}_2 d\boldsymbol{\Sigma}^{-1} du
\end{aligned}$$

$$= \int_{-\infty}^{x} f(u|\mathbf{y} \in \pi_1)du. \tag{49}$$

Although Geisser (1967) discussed the distribution of U under the Jeffreys prior given by (47), the posterior predictive density of U has not been derived explicitly. In the following theorem, we present the posterior predictive density of U, which is expressed as the one-dimensional integral involving confluent hypergeometric function defined as

$${}_1F_1\left(a; b; x\right) = \sum_{k=0}^{\infty} \frac{(a)_k}{(b)_k} \frac{x^k}{k!},$$

where $(a)_k$ and $(b)_k$ are Pochhammer symbols. Let $\phi(\cdot)$ denote the standard normal density and $g_p(v)$ stand for the χ^2-density with p degrees of freedom.

Theorem 4.1. *Assume a new observation* $\mathbf{y}$ *is from* π_1.
(a) *Under the Jeffreys prior given by* (47)*, the posterior predictive density of U is given by*

$$f(u|\mathbf{y} \in \pi_1) = \left(\frac{nc}{nc+Q}\right)^{n/2} \int_0^{\infty} \frac{1}{\sqrt{cv}} \phi\left(\frac{u - cv/2}{\sqrt{cv}}\right) g_p(v) {}_1F_1\left(\frac{n}{2}; \frac{p}{2}; \frac{Q}{2(nc+Q)}v\right) dv,$$

where $n = N_1 + N_2 - 2,\ c = N_1^{-1} + N_2^{-1}$ *and* $Q = (\bar{\mathbf{x}}_1 - \bar{\mathbf{x}}_2)'\mathbf{S}^{-1}(\bar{\mathbf{x}}_1 - \bar{\mathbf{x}}_2)$.
(b) *Under the NIW prior given by* (48)*, the posterior predictive density of U can be obtained by replacing* n *with* $m = N_1 + N_2 + m_0$*,* Q *with* md*, and* c *with* $\tilde{c}$ *in Theorem* 4.1 (a)*, where* $\tilde{c} = (k_1 + N_1)^{-1} + (k_2 + N_2)^{-1}$ *and* $d = (\boldsymbol{\omega}_1 - \boldsymbol{\omega}_2)'\boldsymbol{\Lambda}^{-1}(\boldsymbol{\omega}_1 - \boldsymbol{\omega}_2)$.

To understand the characteristics of the distribution of U, it is important to compute the moments of U. In particular, the skewness and kurtosis play an important role in measuring the deviation from a normal distribution. Geisser (1967) derived the mean and variance of U under the Jeffreys prior. In Theorem 4.2, we provide not only the mean and variance but also the skewness and kurtosis of U under both the priors. As will be seen later, the mean and variance of U are needed to evaluate the asymptotic expansion for the distribution function of U which is given in Theorem 4.7. In addition, we can assess the adequacy of a normal approximation for the distribution of U by computing skewness and kurtosis of U.

Theorem 4.2. *Assume the same conditions as Theorem* 4.1.
(a) *Under the Jeffreys prior, posterior predictive mean, variance, skewness and kurtosis of U are given by*

$$\mathrm{E}[U] = \frac{Q + cp}{2},$$
$$\mathrm{V}[U] = \frac{1}{2}\left[\frac{Q^2}{n} + 2(c+1)Q + c(c+2)p\right],$$

$$\text{Skewness}[U] = \frac{1}{\text{V}(U)^{3/2}} \left[\frac{Q^3}{n^2} + 3(c+1)\frac{Q^2}{n} + 3c(c+2)Q + c^2(c+3)p \right],$$
$$\text{Kurtosis}[U] = \frac{3}{\text{V}(U)^2} \left[\frac{Q^4}{n^3} + 4(c+1)\frac{Q^3}{n^2} + 2(3c^2+6c+1)\frac{Q^2}{n} + 4c(c^2+3c+1)Q + c^2(c^2+4c+2)p \right] + 3,$$

where $n = N_1 + N_2 - 2$, $c = N_1^{-1} + N_2^{-1}$, *and* $Q = (\bar{\mathbf{x}}_1 - \bar{\mathbf{x}}_2)'\mathbf{S}^{-1}(\bar{\mathbf{x}}_1 - \bar{\mathbf{x}}_2)$.
(b) *Under the NIW prior, posterior predictive mean, variance, skewness and kurtosis of* U *can be obtained by replacing* n *with* $m = N_1 + N_2 + m_0$, Q *with* md, *and* c *with* $\tilde{c}$ *in Theorem* 4.2 (a), *where* $\tilde{c} = (k_1+N_1)^{-1} + (k_2+N_2)^{-1}$ *and* $d = (\boldsymbol{\omega}_1 - \boldsymbol{\omega}_2)'\boldsymbol{\Lambda}^{-1}(\boldsymbol{\omega}_1 - \boldsymbol{\omega}_2)$.

Let $K_s(x)$ be the modified Bessel function of the second kind, which is one of the solutions of the following differential equation;

$$z^2\frac{d^2w}{dz^2} + z\frac{dw}{dz} - (z^2 + v^2)w = 0,$$

where z is an arbitrary complex number and v is an arbitrary real number (Abramowitz & Stegun, 1965). In addition, $I(x; a, b)$ is the incomplete beta function defined as

$$I(x; a, b) = \frac{\Gamma(a+b)}{\Gamma(a)\Gamma(b)} \int_0^x t^{a-1}(1-t)^{b-1}dt,$$

where $a > 0$ and $b > 0$. In particular, Dalton and Dougherty (2011) derived the following finite representation: $I(1; 1/2; Z/2) = 1$ for any positive integer Z and

$$I\left(x; \frac{1}{2}; \frac{Z}{2}\right) = \begin{cases} (2/\pi)\sin^{-1}(\sqrt{x}) & Z = 1, \\ (2/\pi)\sin^{-1}(\sqrt{x}) & \\ +(2/\pi)\sqrt{x}\sum_{k=1}^{(Z-1)/2} \frac{(2k-2)!!}{(2k-1)!!}(1-x)^{k-1/2} & Z > 1 \text{ odd}, \\ \sqrt{x}\sum_{k=0}^{(Z-2)/2} \frac{(2k-1)!!}{(2k)!!}(1-x)^k & Z > 1 \text{ even} \end{cases}$$

for any real number $0 \le x < 1$. In Theorem 4.3, we present the exact cumulative distribution function of U (exact CDF of U) under both the priors. As in (49), the exact CDF of U is identical to the Bayes estimator of ϵ_1.

Theorem 4.3. *Assume the same conditions as Theorem* 4.1.
(a) *Under the Jeffreys prior, the exact CDF of* U *is given by*

$$F(x) = Pr(U \le x) = \frac{1}{2}\left(1 + \frac{Q}{nc}\right)\sum_{k=0}^{\infty} Pr(K = k)A(k, x),$$

where K *is a negative binomial random variable so that*

$$q = \frac{nc}{nc + Q}, \; Pr(K = k) = \frac{\Gamma\left(\frac{n-p}{2} + 1\right)}{k!\Gamma\left(\frac{n-p}{2} - k + 1\right)}(1-q)^k q^{\frac{n-p}{2} - k + 1}.$$

If $x = 0$*,*

$$A(k,0) = 1 - I\left(\frac{nc+Q}{n(c+4)+Q}, \frac{1}{2}, \frac{p}{2}+k\right)$$

and if $x \neq 0$ *and* p *is even,*

$$A(k,x) = \left(\frac{x}{|x|}+1\right)\exp\left(\frac{x-|x|}{2}\right) - \frac{e^{\frac{x}{2}}}{\sqrt{\pi}}\sum_{i=0}^{\frac{p}{2}+k-1}\frac{1}{i!}\left(\frac{n|x|}{\sqrt{nc+Q}\sqrt{n(c+4)+Q}}\right)^{i-\frac{1}{2}}$$

$$\times\left[\frac{\sqrt{n}x}{\sqrt{nc+Q}}K_{i-\frac{1}{2}}\left(\frac{|x|}{2}\sqrt{1+\frac{4n}{nc+Q}}\right) + \frac{\sqrt{n}|x|}{\sqrt{n(c+4)+Q}}K_{i+\frac{1}{2}}\left(\frac{|x|}{2}\sqrt{1+\frac{4n}{nc+Q}}\right)\right],$$

meanwhile, if $x \neq 0$ *and* p *is odd,*

$$A(k,x) = \frac{e^{\frac{x}{2}}}{\sqrt{\pi}}\left[\sum_{i=\frac{p-1}{2}+k}^{\infty}\frac{1}{\Gamma(i+\frac{3}{2})}\left(\frac{n|x|}{\sqrt{nc+Q}\sqrt{n(c+4)+Q}}\right)^{i}\right.$$

$$\left.\times\left\{\frac{\sqrt{n}x}{\sqrt{nc+Q}}K_{i}\left(\frac{|x|}{2}\sqrt{1+\frac{4n}{nc+Q}}\right) + \frac{\sqrt{n}|x|}{\sqrt{n(c+4)+Q}}K_{i+1}\left(\frac{|x|}{2}\sqrt{1+\frac{4n}{nc+Q}}\right)\right\}\right].$$

(b) *Under the NIW prior, the exact CDF of* U *is given by replacing* n *with* $m = N_1 + N_2 + m_0$*,* Q *with* md*, and* c *with* $\tilde{c}$ *in Theorem* 4.3 (a)*, where* $\tilde{c} = (k_1+N_1)^{-1} + (k_2+N_2)^{-1}$ *and* $d = (\boldsymbol{\omega}_1 - \boldsymbol{\omega}_2)'\boldsymbol{\Lambda}^{-1}(\boldsymbol{\omega}_1 - \boldsymbol{\omega}_2)$*.*

The exact CDF of U, denoted by $F(x)$, generally involves some special functions and the infinite sums. However, there are some cases where $F(x)$ is represented only by elementary functions and the finite sums. Tables 1 and 2 summarize when and how $F(x)$ is expressed. We observe from Table 1 that if both n and p are even or odd, the expression of $F(0)$ consists of elementary functions only, and has the finite sums. Table 2 shows that if both n and p are even, then the expression of $F(x)$ consists of elementary functions only and has the finite sums for all x. These facts imply that we can quite easily and exactly evaluate the estimate of the optimal error rate for the above cases. We state these cases as Corollary 4.4 and 4.5.

Corollary 4.4. *Assume the same conditions as Theorem* 4.1*. Under the Jeffreys prior, if both* n *and* p *are even or odd, then*

$$F(0) = \frac{1}{2}\left(1+\frac{Q}{nc}\right)\sum_{k=0}^{(n-p)/2} Pr(K=k)\left[1 - I\left(\frac{nc+Q}{n(c+4)+Q}, \frac{1}{2}, \frac{p}{2}+k\right)\right], \quad (50)$$

where $I(1;1/2;Z/2) = 1$ *for any positive integer* Z*, and*

$$I\left(x;\frac{1}{2};\frac{Z}{2}\right) = \begin{cases} (2/\pi)\sin^{-1}(\sqrt{x}) & Z = 1, \\ (2/\pi)\sin^{-1}(\sqrt{x}) & \\ +(2/\pi)\sqrt{x}\sum_{k=1}^{(Z-1)/2}\frac{(2k-2)!!}{(2k-1)!!}(1-x)^{k-1/2} & Z > 1 \text{ odd}, \\ \sqrt{x}\sum_{k=0}^{(Z-2)/2}\frac{(2k-1)!!}{(2k)!!}(1-x)^{k} & Z > 1 \text{ even} \end{cases}$$

Table 1: Classification of expressions for the exact CDF of U at $x = 0$.

p	n	
	even	odd
even	finite sums and elementary functions	infinite sums and elementary functions
odd	infinite sums and elementary functions	finite sums and elementary functions

Table 2: Classification of expressions for the exact CDF of U at $x \neq 0$

p	n	
	even	odd
even	finite sums and elementary functions	infinite sums and elementary functions
odd	double infinite sums and special functions	infinite sums and special functions

for any real number $0 \leq x < 1$.

Corollary 4.5. *Assume the same conditions as Theorem* 4.1. *Under the Jeffreys prior, if both* n *and* p *are even, then*

$$F(x) = \frac{1}{2}\left(1 + \frac{Q}{nc}\right) \sum_{k=0}^{(n-p)/2} Pr(K = k) A(k, x), \tag{51}$$

where

$$A(k,x) = \left(\frac{x}{|x|} + 1\right) \exp\left(\frac{x - |x|}{2}\right) - \frac{e^{x/2}}{\sqrt{\pi}} \sum_{i=0}^{p/2+k-1} \frac{1}{i!} \left(\frac{n|x|}{\sqrt{nc+Q}\sqrt{n(c+4)+Q}}\right)^{i-1/2}$$
$$\times \left[\frac{\sqrt{n}x}{\sqrt{nc+Q}} K_{i-1/2}\left(\frac{|x|}{2}\sqrt{1 + \frac{4n}{nc+Q}}\right) + \frac{\sqrt{n}|x|}{\sqrt{n(c+4)+Q}} K_{i+1/2}\left(\frac{|x|}{2}\sqrt{1 + \frac{4n}{nc+Q}}\right)\right]$$

with

$$K_{i+1/2}(x) = \sqrt{\frac{\pi}{2x}} \exp(-x) \sum_{k=0}^{i} \frac{(k+i)!}{k!(i-k)!}(2x)^{-k},$$
$$K_{\pm 1/2}(x) = \sqrt{\frac{\pi}{2x}} \exp(-x).$$

If either n or p is not even, however, we need to evaluate the (double) infinite sums and the modified Bessel function of the second kind. In this case, the infinite sums can be truncated whenever convergence is observed and we may use mathematical software such as R, which is able to evaluate the modified Bessel function of the second kind.

4.3.2 Approximations for the Bayes estimator of the optimal error rate

In the previous section, we observed that the exact CDF of U generally involves special functions and (double) infinite sums. To avoid the computation of these complicated elements, we discuss the approximations for the Bayes estimator of ϵ_1 under both the priors.

This section suggests the approximations based on Edgeworth expansion. Concerning Edgeworth expansions, Javed et al. (2021) discussed Edgeworth expansion of random sum of independent and identically distributed random vectors.

Under the Jeffreys prior, Geisser (1986) proposed a normal approximation

$$\mathrm{E}(\epsilon_1) \approx \Phi\left(\frac{x-(pc+Q)/2}{[pc+(1+pc)Q]^{1/2}}\right), \tag{52}$$

where $Q = (\bar{\mathbf{x}}_1 - \bar{\mathbf{x}}_2)'\mathbf{S}^{-1}(\bar{\mathbf{x}}_1 - \bar{\mathbf{x}}_2)$ and $c = N_1^{-1} + N_2^{-1}$. We refer to the approximation given by (52) as the Geisser approximation. The Geisser approximation could be obtained by approximating the variance of U in Theorem 4.2 by $pc+(1+pc)Q$ and then regarding the distribution of U as the normal distribution with the mean $(pc+Q)/2$ and the approximated variance $pc + (1 + pc)Q$. However, the Geisser approximation appears somewhat rough, because it is apparent from Theorem 4.2 that the distribution of U is skewed and heavy tailed, and the approximated variance is obtained by deleting the terms pc^2 and Q^2/n in $\mathrm{V}(U)$. This section suggests a more accurate approximation for the exact CDF of U. Let $\overset{\mathrm{d}}{=}$ denote the equality in distribution. From the proof of Theorem 4.1, the stochastic representation of U is given by

$$U \overset{\mathrm{d}}{=} \frac{cv}{2} + \sqrt{cv}u_0, \tag{53}$$

where $v|z \sim \chi^2_p(Qz/(nc))$, $z \sim \chi^2_n$, and $u_0 \sim N(0,1)$; u_0 is independent of v and z. Suppose $X \overset{\mathrm{d}}{\approx} Y$ means that the distribution of X is approximated by that of Y. At first, we approximate the distribution of U based on the approximation for the distribution of cv. Geisser (1967) equated the first two moments of cv to those of a constant R times a χ^2 random variable with f degrees of freedom resulting in

$$cv \overset{\mathrm{d}}{\approx} RW \tag{54}$$

where

$$R = \frac{pc^2 + 2Qc + Q^2/n}{cp + Q},\ f = \frac{(pc+Q)^2}{pc^2 + 2Qc + Q^2/n},\ W \sim \chi^2_f.$$

Applying approximation (54) to (53), we can get

$$U \overset{\mathrm{d}}{\approx} U_1 \text{ and } U_1 \overset{\mathrm{d}}{=} W\frac{R}{2} + \sqrt{WR}u_0, \tag{55}$$

where $u_0 \sim N(0,1)$ and $W \sim \chi^2_f$, and they are independently distributed. The following lemma is directly used to derive the Edgeworth expansion for U_1.

Lemma 4.6. *Suppose the stochastic representation of a random variable X is*

$$X \stackrel{\mathrm{d}}{=} W\delta + \sqrt{W}\Omega^{1/2}u_0$$

where $u_0 \sim N(0,1)$ and $W \sim \chi^2_f$; they are independently distributed. Then the Edgeworth expansion for the distribution function of the standardized X/f is given by

$$P(S_f \leq x) = \Phi(x) + f^{-1/2}p_1(x)\phi(x) + f^{-1}p_2(x)\phi(x) + o(f^{-1}),$$

where

$$S_f = \sqrt{f}\frac{X/f - \delta}{\sqrt{2\delta^2 + \Omega}},$$
$$p_1(x) = -\frac{1}{6}\tilde{\kappa}_3(x^2-1),\ p_2(x) = -x\left\{\frac{1}{24}\tilde{\kappa}_4(x^2-3) + \frac{1}{72}\tilde{\kappa}_3^2(x^4 - 10x^2 + 15)\right\},$$
$$\tilde{\kappa}_3 = \frac{2\delta(3\Omega + 4\delta^2)}{(2\delta^2+\Omega)^{3/2}},\ \tilde{\kappa}_4 = \frac{6(\Omega^2 + 8\Omega\delta^2 + 8\delta^4)}{(2\delta^2+\Omega)^2}.$$

Using Lemma 4.6 with $\delta = R/2$ and $\Omega = R$, we can derive the Edgeworth expansion for the distribution function of U_1, whose stochastic representation is given by (55). We refer to the Edgeworth expansion for U_1 as the approximate-Edgeworth expansion for U. We summarize the results in Theorem 4.7.

Theorem 4.7. *Assume the same conditions as Theorem* 4.1.
(a) *Under the Jeffreys prior, the approximate-Edgeworth expansion for U is given by*

$$Pr\left(\frac{U - \mathrm{E}(U)}{\sqrt{\mathrm{V}(U)}} \leq x\right) \approx Pr\left(\frac{U_1 - \mathrm{E}(U)}{\sqrt{\mathrm{V}(U)}} \leq x\right)$$
$$= \Phi(x) + f^{-1/2}p_1(x)\phi(x) + f^{-1}p_2(x)\phi(x) + o(f^{-1}), \quad (56)$$

where

$$f = \frac{(cp+Q)^2}{pc^2 + 2Qc + Q/n},\ R = \frac{pc^2 + 2Qc + Q^2/n}{cp+Q},$$
$$\tilde{\kappa}_3 = \frac{\sqrt{R}(R+3)}{(R/2+1)^{3/2}},\ \tilde{\kappa}_4 = \frac{6(1 + 2R + R^2/2)}{(R/2+1)^2}.$$

(b) *Under the NIW prior, the approximate-Edgeworth expansion for U is given by replacing n with $m = N_1 + N_2 + m_0$, Q with md, and c with $\tilde{c}$ in Theorem* 4.7(a)*, where $\tilde{c} = (k_1+N_1)^{-1} + (k_2+N_2)^{-1}$ and $d = (\boldsymbol{\omega}_1 - \boldsymbol{\omega}_2)'\boldsymbol{\Lambda}^{-1}(\boldsymbol{\omega}_1 - \boldsymbol{\omega}_2)$.*

We refer to the leading term $\Phi(x)$ in (56) as the leading term approximation, and to the approximation including the second and third terms of (56) as the second order approximation, which is denoted by $F_2(x)$. We suggest these approximations as the approximations for the exact CDF of U.

4.4 Numerical studies

In the foregoing section, we discussed the approximations for the exact CDF of U under both the priors. In this section, we investigate the accuracy of the approximations for the exact CDF of U under the Jeffreys prior. The statistical software R is used to create all tables and the seed value is `set.seed(1)`. All simulations are made for $\boldsymbol{\mu}_1 = (\Delta, 0, \cdots, 0)'$, $\boldsymbol{\mu}_2 = (0, \cdots, 0)'$ and $\boldsymbol{\Sigma} = \mathbf{I}_p$, where Δ^2 is the squared Mahalanobis distance between π_1 and π_2, and $\mathbf{I}_p$ denotes the $p \times p$ identity matrix. In addition, we choose the sample sizes N_1 and N_2 and the dimension p, so that we can use the finite representation formula (50) or (51) which enables use to compute the exact CDF of U exactly and quickly. Let $F_2(x)$ be the second order approximation, $\Phi(x)$ be the leading term approximation, and $\Phi(Rx)$ with $R = \sqrt{\mathrm{V}(U)/[(c+1)Q + cp]}$ be the Geisser approximation. For various pairs of (N_1, N_2) and several cut-off point x, we compare the relative errors of $F_2(x)$, $\Phi(x)$, and the $\Phi(Rx)$ to the exact CDF of the standardized U denoted by $F(x)$. The simulation results are presented in Tables 3, 4, 5 and 6. We highlight in bold the lowest absolute value of the relative errors in these tables. We observe from Table 3 that $\Phi(x)$ shows better performance than $F_2(x)$ when both the squared Mahalanobis distance and dimension are small. However, $F_2(x)$ shows considerable improvement over $\Phi(x)$ and $\Phi(Rx)$ in most cases. Based on the relative errors of $F_2(x)$ to $F_(x)$, It seems that $F_2(x)$ tends to underestimate $F_(x)$. The performance of $\Phi(Rx)$ is similar to that of $\Phi(x)$ for small dimension, but for large dimension, the deviation of $\Phi(Rx)$ from $\Phi(x)$ is also found.

4.5 Empirical studies

In this section, we apply the results of Section 2 to estimate the optimal error rate for Fisher's iris data. The data involves the measurements in cm of the sepal length and width and petal length and width of 50 plants for each of three types of iris; Iris setosa, Iris versicolor, and Iris virginica. For the sake of illustration, we consider only two species: versicolor and virginica. We assume the cut-off point is 0. We calculate the Bayes estimate of the optimal error rate, the leading term approximation $\Phi(\eta)$, and the second order approximation

$$F_2(\eta) = \Phi(\eta) + f^{-1/2} p_1(\eta)\phi(\eta) + f^{-1} p_2(\eta)\phi(\eta),$$

where $\eta = [x - \mathrm{E}(U)] / \sqrt{\mathrm{V}(U)}$. The Geisser approximation given by (52) is omitted, since as was pointed out in the previous section, the performance of the Geisser approximation is similar to that of the normal approximation for small p. For comparison, we consider the classical estimators of the optimal error rate: $\Phi(-Q/2)$ and $\Phi(-Q_1/2)$,

Table 3: Relative errors (%) of the second order approximation $F_2(x)$, leading term approximation $\Phi(x)$, and Geisser approximation $\Phi(Rx)$ to the exact CDF $F(x)$, for $\Delta^2 = 0.5, 10$, $p = 2, 10$, and cut-off point is 0.

Δ^2	N_1	N_2	$p=2$			$p=10$		
			$F_2(x)$	$\Phi(x)$	$\Phi(Rx)$	$F_2(x)$	$\Phi(x)$	$\Phi(Rx)$
	10	10	3.31	**-0.82**	1.41	**1.77**	10.5	69.1
	20	10	**1.88**	5.82	9.88	**1.06**	9.51	67.6
	20	20	1.16	**-0.68**	0.96	**0.70**	1.63	21.6
	30	20	0.82	**0.46**	2.46	**0.53**	0.83	15.6
0.5	30	30	0.59	**-0.33**	1.00	**0.39**	0.66	14.1
	40	30	0.46	**0.07**	1.47	**0.31**	2.00	23.2
	40	40	0.36	**-0.05**	1.12	**0.24**	1.77	21.0
	50	40	**0.25**	1.41	2.91	**0.21**	1.22	16.9
	50	50	0.24	**-0.21**	0.68	**0.18**	0.43	10.7
	10	10	**-2.38**	74.2	32.8	**-46.3**	288	727
	20	10	**-29.8**	202	33.5	**-32.8**	221	682
	20	20	**-1.89**	37.7	19.8	**-3.90**	54.5	215
	30	20	**-2.88**	44.2	19.2	**-2.43**	40.3	173
10	30	30	**-1.26**	27.2	14.3	**-1.06**	24.8	116
	40	30	**-0.83**	21.4	12.5	**-1.85**	34.4	141
	40	40	**-0.78**	20.2	11.0	**-1.69**	33.4	130
	50	40	**-1.40**	29.7	11.3	**-1.22**	27.4	111
	50	50	**-0.46**	15.0	8.72	**-0.52**	16.0	75.5

Table 4: Relative errors (%) of the second order approximation $F_2(x)$, leading term approximation $\Phi(x)$, and Geisser approximation $\Phi(Rx)$ to the exact CDF $F(x)$, for $\Delta^2 = 0.5, 10$, $p = 2, 10$, and cut-off point is 0.1.

Δ^2	N_1	N_2	$p = 2$			$p = 10$		
			$F_2(x)$	$\Phi(x)$	$\Phi(Rx)$	$F_2(x)$	$\Phi(x)$	$\Phi(Rx)$
	10	10	**0.00**	-6.34	-4.74	**1.57**	6.35	53.6
	20	10	**1.26**	2.12	5.40	**0.99**	6.06	53.5
	20	20	**-0.09**	-3.51	-2.32	0.49	**-0.01**	15.9
	30	20	0.33	-1.47	**0.09**	**0.34**	-0.53	11.2
0.5	30	30	**0.07**	-2.06	-1.06	**0.25**	-0.49	10.1
	40	30	**0.16**	-1.30	-0.22	**0.24**	0.82	18.2
	40	40	**0.11**	-1.25	-0.36	**0.19**	0.73	16.5
	50	40	**0.18**	0.39	1.62	**0.16**	0.32	13.1
	50	50	**0.05**	-1.21	-0.54	**0.11**	-0.33	7.82
	10	10	**0.04**	54.4	21.1	**-25.0**	210	522
	20	10	**-19.6**	157	21.2	**-21.8**	172	520
	20	20	**-0.97**	29.9	14.6	**-2.47**	44.3	176
	30	20	**-1.89**	36.2	14.5	**-1.59**	33.1	144
10	30	30	**-0.77**	22.2	10.9	**-0.64**	20.3	97.5
	40	30	**-0.51**	17.5	9.64	**-1.33**	28.9	120
	40	40	**-0.50**	16.7	8.53	**-1.27**	28.3	111
	50	40	**-1.06**	25.2	8.80	**-0.91**	23.2	95.7
	50	50	**-0.29**	12.4	6.83	**-0.35**	13.4	65.2

Table 5: Relative errors (%) of the second order approximation $F_2(x)$, leading term approximation $\Phi(x)$, and Geisser approximation $\Phi(Rx)$ to the exact CDF $F(x)$, for $\Delta^2 = 0.5, 10$, $p = 2, 10$, and cut-off point is 0.3.

Δ^2	N_1	N_2	$p = 2$			$p = 10$		
			$F_2(x)$	$\Phi(x)$	$\Phi(Rx)$	$F_2(x)$	$\Phi(x)$	$\Phi(Rx)$
	10	10	**-2.34**	-11.3	-10.6	0.92	**0.09**	30.4
	20	10	**0.21**	-3.13	-1.05	**0.63**	0.77	31.9
	20	20	**-1.27**	-6.74	-6.22	**0.09**	-2.44	7.06
	30	20	**-0.38**	-4.11	-3.27	**0.00**	-2.51	4.13
0.5	30	30	**-0.58**	-4.33	-3.87	**-0.02**	-2.20	3.83
	40	30	**-0.27**	-3.21	-2.66	**0.09**	-1.02	10.2
	40	40	**-0.25**	-2.93	-2.49	**0.07**	-0.89	9.29
	50	40	**0.04**	-1.18	-0.39	**0.04**	-1.07	6.96
	50	50	**-0.22**	-2.61	-2.30	**-0.02**	-1.47	3.18
	10	10	**1.47**	26.9	4.82	**-5.04**	113	280
	20	10	-6.99	93.8	**3.54**	**-8.36**	105	309
	20	20	**0.02**	17.7	6.39	**-0.69**	28	118
	30	20	**-0.61**	23.2	6.92	**-0.48**	21.3	99.6
10	30	30	**-0.16**	14.0	5.39	**-0.11**	12.8	68.6
	40	30	**-0.09**	11.0	4.98	**-0.58**	19.6	86.1
	40	40	**-0.13**	10.8	4.45	**-0.62**	19.6	81.2
	50	40	**-0.52**	17.5	4.57	**-0.44**	16.1	70.5
	50	50	**-0.07**	7.80	3.66	**-0.11**	8.81	47.9

Table 6: Relative errors (%) of the second order approximation $F_2(x)$, leading term approximation $\Phi(x)$, and Geisser approximation $\Phi(Rx)$ to the exact CDF $F(x)$, for $\Delta^2 = 0.5, 10$, $p = 2, 10$, and cut-off point is 0.5.

Δ^2	N_1	N_2	$p=2$			$p=10$		
			$F_2(x)$	$\Phi(x)$	$\Phi(Rx)$	$F_2(x)$	$\Phi(x)$	$\Phi(Rx)$
	10	10	**-2.11**	-12.0	-11.9	**0.30**	-3.80	14.8
	20	10	**-0.38**	-5.98	-4.79	**0.22**	-2.67	16.9
	20	20	**-1.29**	-7.68	-7.65	**-0.16**	-3.82	0.86
	30	20	**-0.62**	-5.32	-5.03	**-0.20**	-3.61	-0.75
0.5	30	30	**-0.71**	-5.25	-5.19	**-0.17**	-3.15	-0.59
	40	30	**-0.43**	-4.13	-3.98	**-0.04**	-2.19	4.19
	40	40	**-0.37**	-3.74	-3.64	**-0.03**	-1.93	3.90
	50	40	**-0.07**	-2.16	-1.73	**-0.05**	-1.94	2.40
	50	50	**-0.31**	-3.28	-3.24	**-0.09**	-2.13	-0.16
	10	10	**1.15**	10.3	-4.71	**0.77**	59.8	156
	20	10	**-1.83**	54.7	-7.49	**-2.35**	62.3	189
	20	20	**0.32**	8.89	0.67	**0.09**	16.3	78.7
	30	20	**-0.01**	13.6	1.45	**0.04**	12.5	67.9
10	30	30	**0.10**	7.79	1.34	**0.12**	7.10	47.2
	40	30	**0.09**	6.04	1.50	**-0.15**	12.4	61.5
	40	40	**0.04**	6.20	1.38	**-0.21**	12.8	58.5
	50	40	**-0.19**	11.5	1.35	**-0.14**	10.4	51.1
	50	50	**0.03**	4.57	1.25	**0.02**	5.23	34.4

Table 7: Comparison of the exact Bayes estimate for the optimal error rate, approximations for the Bayes estimate and classical estimates for $N_1 = N_2 = 50$. These values are calculated for several sets of variables; SL, SW, PL and PW (sepal length, sepal width, petal length and petal width, respectively.

variables	Exact	Approximations		Classical estimates	
	Bayes estimates	Normal	Edgeworth	$\Phi(-Q/2)$	$\Phi(-Q_1/2)$
(SL, SW)	0.284	0.284	0.285	0.286	0.289
(PL, PW)	0.058	0.064	0.058	0.056	0.059
(SL, SW, PL, PW)	0.031	0.036	0.031	0.030	0.033

where $Q = (\bar{\mathbf{x}}_1 - \bar{\mathbf{x}}_2)'\mathbf{S}^{-1}(\bar{\mathbf{x}}_1 - \bar{\mathbf{x}}_2)$and $Q_1 = (n-p-1)Q/n$. These values are reported in Table 7. As may be seen in Table 7, the Bayes estimate is parallel to the classical estimates. If we employ the sepal length and width, and petal length and width as the variables used for discrimination, then the estimate of the optimal error rate gives the smallest value. On the other hand, the error rate associated with sepal length and width is even higher than the error rate associated with the other variables. It could be concluded that one may select all the variables as the discriminant variables in terms of the optimal error rate.

4.6 Summary

Geisser (1967) pointed out that the estimation of the optimal error rate is useful as a guide to the optimal discriminatory power of the variables used for allocation. However, previous studies on misclassification rates in linear discriminant analysis have mainly focused on the estimation of the actual error rate. In section 4, we derived the exact expressions for the Bayes estimator of the optimal error rate in Theorem 4.3. In general, the expression of the Bayes estimator of the optimal error rate involves special functions and the infinite sums. However, the estimator can be represented only by the elementary functions and the finite sums for specific combinations of dimension p and $n = N_1 + N_2 - 2$. We also suggested the leading term and second order approximations for the estimate of the optimal error rate. The suggested approximations were derived based on the approximate distribution of U, the stochastic representation of which is given by (55). As shown in section 4.4, the second order approximation showed considerable improvement over the leading term and Geisser approximations in most cases. This deviation could be caused by the deletion of the term pc^2 in the exact variance of U, because pc^2 may not be ignored when p is large.

It should be noted that if the value of f, which is defined in Theorem 4.7, is large, then the estimate of the optimal error rate would be well approximated by the leading term approximation. It is easily found that the dimension p and sample squared Mahalanobis distance Q monotonically increase the value of f, while $c = N_1^{-1} + N_2^{-1}$ monotonically decrease the value of f. These facts could indicate the good performance of the leading term approximation for large p, Δ^2 or sample sizes.

We have assumed the degrees of freedom of the posterior distribution of $\boldsymbol{\Sigma}^{-1}$ is a natural number. However, by referring to Rao (1973) and others, we can apply the discussion here not only to the natural degrees of freedom, but also to the real degrees of freedom. We will leave this as a future issue.

Appendix

Initially, we give the proof for the joint posterior distribution for $\boldsymbol{\mu}_1$, $\boldsymbol{\mu}_2$, $\boldsymbol{\Sigma}^{-1}$ under the NIW prior. It follows from (48) the joint density function of $\boldsymbol{\mu}_1$, $\boldsymbol{\mu}_2$, $\boldsymbol{\Sigma}^{-1}$ is given by

$$P(\boldsymbol{\mu}_1, \boldsymbol{\mu}_2, \boldsymbol{\Sigma}^{-1}) \propto |\boldsymbol{\Sigma}^{-1}|^{\frac{m_0-p+1}{2}} \operatorname{etr}\left[-\frac{1}{2}\{\boldsymbol{\Lambda}_0 + \sum_{i=1}^{2} k_i(\boldsymbol{\mu}_i - \boldsymbol{\xi}_i)(\boldsymbol{\mu}_i - \boldsymbol{\xi}_i)'\}\boldsymbol{\Sigma}^{-1}\right],$$

Since $\bar{\mathbf{x}}_i \sim N_p(\boldsymbol{\mu}_i, \Sigma/N_i)$ and $n\mathbf{S} \sim W_p(n, \Sigma)$, and $\boldsymbol{t} = (\bar{\mathbf{x}}_1, \bar{\mathbf{x}}_2, \mathbf{S})$ are sufficient for the parameters of the population, the likelihood function is given by

$$L(\mathbf{t}|\boldsymbol{\mu}_1, \boldsymbol{\mu}_2, \boldsymbol{\Sigma}^{-1}) \propto |\boldsymbol{\Sigma}^{-1}|^{\frac{N_1+N_2}{2}} \operatorname{etr}\left[-\frac{1}{2}\{n\mathbf{S} + \sum_{i=1}^{2} N_i(\bar{\mathbf{x}}_i - \boldsymbol{\mu}_i)(\bar{\mathbf{x}}_i - \boldsymbol{\mu}_i)'\}\boldsymbol{\Sigma}^{-1}\right].$$

The posterior distribution becomes

$$\begin{aligned} P(\boldsymbol{\mu}_1, \boldsymbol{\mu}_2, \boldsymbol{\Sigma}^{-1}|\mathbf{t}) &\propto |\boldsymbol{\Sigma}^{-1}|^{\frac{N_1+N_2+m_0-p+1}{2}} \operatorname{etr}\left[-\frac{1}{2}\left(\boldsymbol{\Lambda}_0 + n\mathbf{S}\right)\boldsymbol{\Sigma}^{-1}\right] \\ &\times \operatorname{etr}\left[-\frac{1}{2}\left\{\sum_{i=1}^{2} k_i(\boldsymbol{\mu}_i - \boldsymbol{\xi}_i)(\boldsymbol{\mu}_i - \boldsymbol{\xi}_i)' + N_i(\bar{\mathbf{x}}_i - \boldsymbol{\mu}_i)(\bar{\mathbf{x}}_i - \boldsymbol{\mu}_i)'\right\}\boldsymbol{\Sigma}^{-1}\right]. \end{aligned} \tag{57}$$

If we put $B_i = k_i(\boldsymbol{\mu}_i - \boldsymbol{\xi}_i)(\boldsymbol{\mu}_i - \boldsymbol{\xi}_i)' + N_i(\bar{\mathbf{x}}_i - \boldsymbol{\mu}_i)(\bar{\mathbf{x}}_i - \boldsymbol{\mu}_i)'$, then

$$\begin{aligned} B_i &= k_i(\boldsymbol{\mu}_i - \bar{\mathbf{x}}_i + \bar{\mathbf{x}}_i - \boldsymbol{\xi}_i)(\boldsymbol{\mu}_i - \bar{\mathbf{x}}_i + \bar{\mathbf{x}}_i - \boldsymbol{\xi}_i)' + N_i(\bar{\mathbf{x}}_i - \boldsymbol{\mu}_i)(\bar{\mathbf{x}}_i - \boldsymbol{\mu}_i)' \\ &= (k_i + N_i)(\boldsymbol{\mu}_i - \bar{\mathbf{x}}_i)(\boldsymbol{\mu}_i - \bar{\mathbf{x}}_i)' \\ &+ k_i[(\boldsymbol{\mu}_i - \bar{\mathbf{x}}_i)(\bar{\mathbf{x}}_i - \boldsymbol{\xi}_i)' + (\bar{\mathbf{x}}_i - \boldsymbol{\xi}_i)(\boldsymbol{\mu}_i - \bar{\mathbf{x}}_i)' + (\bar{\mathbf{x}}_i - \boldsymbol{\xi}_i)(\bar{\mathbf{x}}_i - \boldsymbol{\xi}_i)'] \\ &= (k_i + N_i)(\boldsymbol{\mu}_i - \bar{\mathbf{x}}_i)(\boldsymbol{\mu}_i - \bar{\mathbf{x}}_i)' + k_i[(\boldsymbol{\mu}_i - \bar{\mathbf{x}}_i)(\bar{\mathbf{x}}_i - \boldsymbol{\xi}_i)' + (\bar{\mathbf{x}}_i - \boldsymbol{\xi}_i)(\boldsymbol{\mu}_i - \bar{\mathbf{x}}_i)'] \\ &+ \frac{k_i^2}{k_i + N_i}(\bar{\mathbf{x}}_i - \boldsymbol{\xi}_i)(\bar{\mathbf{x}}_i - \boldsymbol{\xi}_i)'\frac{k_i + N_i}{k_i} \\ &= (k_i + N_i)(\boldsymbol{\mu}_i - \bar{\mathbf{x}}_i)(\boldsymbol{\mu}_i - \bar{\mathbf{x}}_i)' + k_i[(\boldsymbol{\mu}_i - \bar{\mathbf{x}}_i)(\bar{\mathbf{x}}_i - \boldsymbol{\xi}_i)' + (\bar{\mathbf{x}}_i - \boldsymbol{\xi}_i)(\boldsymbol{\mu}_i - \bar{\mathbf{x}}_i)'] \\ &+ \frac{k_i^2}{k_i + N_i}(\bar{\mathbf{x}}_i - \boldsymbol{\xi}_i)(\bar{\mathbf{x}}_i - \boldsymbol{\xi}_i)'\left(1 + \frac{N_i}{k_i}\right) \end{aligned}$$

$$
\begin{aligned}
&= (k_i+N_i)(\boldsymbol{\mu}_i-\bar{\mathbf{x}}_i)(\boldsymbol{\mu}_i-\bar{\mathbf{x}}_i)' + k_i[(\boldsymbol{\mu}_i-\bar{\mathbf{x}}_i)(\bar{\mathbf{x}}_i-\boldsymbol{\xi}_i)' + (\bar{\mathbf{x}}_i-\boldsymbol{\xi}_i)(\boldsymbol{\mu}_i-\bar{\mathbf{x}}_i)']\\
&+ \frac{k_i^2}{k_i+N_i}(\bar{\mathbf{x}}_i-\boldsymbol{\xi}_i)(\bar{\mathbf{x}}_i-\boldsymbol{\xi}_i)' + \frac{k_iN_i}{k_i+N_i}(\bar{\mathbf{x}}_i-\boldsymbol{\xi}_i)(\bar{\mathbf{x}}_i-\boldsymbol{\xi}_i)'\\
&= (k_i+N_i)\left[(\boldsymbol{\mu}_i-\bar{\mathbf{x}}_i)-\frac{k_i}{k_i+N_i}(\boldsymbol{\xi}_i-\bar{\mathbf{x}}_i)\right]\left[(\boldsymbol{\mu}_i-\bar{\mathbf{x}}_i)-\frac{k_i}{k_i+N_i}(\boldsymbol{\xi}_i-\bar{\mathbf{x}}_i)\right]'\\
&+ \frac{k_iN_i}{k_i+N_i}(\bar{\mathbf{x}}_i-\boldsymbol{\xi}_i)(\bar{\mathbf{x}}_i-\boldsymbol{\xi}_i)'\\
&= (k_i+N_i)\left[\boldsymbol{\mu}_i-\left\{\frac{k_i}{k_i+N_i}\boldsymbol{\xi}_i+\left(1-\frac{k_i}{k_i+N_i}\right)\bar{\mathbf{x}}_i\right\}\right]\\
&\times \left[\boldsymbol{\mu}_i-\left\{\frac{k_i}{k_i+N_i}\boldsymbol{\xi}_i+\left(1-\frac{k_i}{k_i+N_i}\right)\bar{\mathbf{x}}_i\right\}\right]' + \frac{k_iN_i}{k_i+N_i}(\bar{\mathbf{x}}_i-\boldsymbol{\xi}_i)(\bar{\mathbf{x}}_i-\boldsymbol{\xi}_i)'\\
&= (k_i+N_i)(\boldsymbol{\mu}_i-\boldsymbol{\omega}_i)(\boldsymbol{\mu}_i-\boldsymbol{\omega}_i)' + \frac{k_iN_i}{k_i+N_i}(\bar{\mathbf{x}}_i-\boldsymbol{\xi}_i)(\bar{\mathbf{x}}_i-\boldsymbol{\xi}_i)',
\end{aligned}
$$

where we put

$$\boldsymbol{\omega}_i = \frac{k_i}{k_i+N_i}\boldsymbol{\xi}_i + \left(1-\frac{k_i}{k_i+N_i}\right)\bar{\mathbf{x}}_i.$$

If we substitute B_i for (57), the desired result follows immediately.

Proof of Theorem 4.1. We only show the proof of Theorem 4.1 (a), because this proof can be applied to the proof of Theorem 4.1 (b) in an obvious way. When a new observation is from π_1, the conditional distribution of U given $\boldsymbol{\mu}_1$, $\boldsymbol{\mu}_2$ and $\boldsymbol{\Sigma}$ is $U|\boldsymbol{\mu}_1,\boldsymbol{\mu}_2,\boldsymbol{\Sigma},\mathbf{y}\in\pi_1 \sim N(\Delta^2/2,\Delta^2)$ with $\Delta^2 = (\boldsymbol{\mu}_1-\boldsymbol{\mu}_2)'\boldsymbol{\Sigma}^{-1}(\boldsymbol{\mu}_1-\boldsymbol{\mu}_2)$. Define $v=\Delta^2/c$ with $c = N_1^{-1}+N_2^{-1}$. Since

$$\frac{1}{\sqrt{c}}\boldsymbol{\Sigma}^{-1/2}(\boldsymbol{\mu}_1-\boldsymbol{\mu}_2)|\boldsymbol{\Sigma} \sim N_p\left(\frac{1}{\sqrt{c}}\boldsymbol{\Sigma}^{-1/2}(\bar{\mathbf{x}}_1-\bar{\mathbf{x}}_2), \boldsymbol{I}_p\right),$$

the conditional distribution of $v|\boldsymbol{\Sigma}$ is the non-central χ^2 distribution with non-centrality parameter $\delta = (\bar{\mathbf{x}}_1-\bar{\mathbf{x}}_2)'\boldsymbol{\Sigma}^{-1}(\bar{\mathbf{x}}_1-\bar{\mathbf{x}}_2)/c$. This is expressed as

$$v|\boldsymbol{\Sigma} \sim \chi_p^2\left(\frac{Q}{nc}z\right),$$

where

$$Q = (\bar{\mathbf{x}}_1-\bar{\mathbf{x}}_2)'\mathbf{S}^{-1}(\bar{\mathbf{x}}_1-\bar{\mathbf{x}}_2),\ z = \frac{(\bar{\mathbf{x}}_1-\bar{\mathbf{x}}_2)'\boldsymbol{\Sigma}^{-1}(\bar{\mathbf{x}}_1-\bar{\mathbf{x}}_2)}{(\bar{\mathbf{x}}_1-\bar{\mathbf{x}}_2)'(n\mathbf{S})^{-1}(\bar{\mathbf{x}}_1-\bar{\mathbf{x}}_2)}.$$

Since z is distributed as χ_m^2 from Theorem 3.2.8 of Muirhead (1982), the unconditional distribution of U is given by

$$
\begin{aligned}
f(u|\mathbf{y}\in\pi_1) = \int_0^\infty\int_0^\infty \frac{1}{\sqrt{cv}}\phi\left(\frac{u-cv/2}{\sqrt{cv}}\right) g_p(v)\exp\left(-\frac{Q}{2nc}z\right){}_0F_1\left(\frac{p}{2};\frac{Q}{4nc}vz\right)\\
\times g_m(z)dvdz.
\end{aligned}
$$

Using Lemma 1.3.3 of Muirhead (1982), the desired result follows immediately. □

Proof of Theorem 4.2. We derive the first four moments of U using the expression of density given in Theorem 4.1. If we define

$$W(k)=\left(\frac{nc}{nc+Q}\right)^{n/2}\int_0^\infty \frac{v^{p/2+k-1}e^{-v/2}}{2^{p/2}\Gamma(p/2)}{}_1F_1\left(\frac{n}{2};\frac{p}{2};\frac{Q}{2(nc+Q)}v\right)dv, \tag{58}$$

for a non-negative number k, then the first four moments of U can be represented by

$$\begin{aligned}
\mathrm{E}(U)&=\frac{c}{2}W(1), \qquad (59)\\
\mathrm{E}(U^2)&=\frac{c^2}{4}W(2)+cW(1),\\
\mathrm{E}(U^3)&=\frac{c^3}{8}W(3)+\frac{3c^2}{2}W(2),\\
\mathrm{E}(U^4)&=\frac{c^4}{16}W(4)+\frac{3c^3}{2}W(3)+3c^2W(2).
\end{aligned}$$

Hence, the central moments of U are expressed as

$$\begin{aligned}
\mathrm{V}(U)&=\frac{c^2}{4}(W(2)-W(1)^2)+cW(1), \qquad (60)\\
\mathrm{E}(U-\mathrm{E}(U))^3&=\frac{c^3}{8}W(3)+\frac{3c^2}{2}W(2)-\frac{3c^3}{8}W(2)W(1)\\
&\quad-\frac{3c^2}{2}W(1)^2+\frac{c^3}{4}W(1)^3, \qquad (61)\\
\mathrm{E}(U-\mathrm{E}(U))^4&=\frac{c^4}{16}W(4)+\frac{3c^3}{2}W(3)+3c^2W(2)\\
&\quad-\frac{c^4}{4}W(3)W(1)-3c^3W(2)W(1)+\frac{3c^4}{8}W(2)W(1)^2\\
&\quad+\frac{3c^3}{2}W(1)^3-\frac{3c^4}{16}W(1)^4. \qquad (62)
\end{aligned}$$

Applying Lemma 1.3.3 of Muirhead (1982) to (58),

$$W(k)=\left(\frac{nc}{nc+Q}\right)^{n/2}\frac{\Gamma(p/2+k)}{\Gamma(p/2)}2^k{}_2F_1\left(\frac{n}{2},\frac{p}{2}+k;\frac{p}{2};\frac{Q}{nc+Q}\right). \tag{63}$$

Using equality

$${}_2F_1(a,b;c;z)=(1-z)^{-a}{}_2F_1\left(a,c-b;c;\frac{z}{z-1}\right)$$

in (63),

$$\begin{aligned}
W(k)&=\frac{\Gamma(p/2+k)}{\Gamma(p/2)}2^k{}_2F_1\left(\frac{n}{2},-k;\frac{p}{2};-\frac{Q}{nc}\right)\\
&=\frac{\Gamma(p/2+k)}{\Gamma(p/2)}2^k\sum_{r=0}^{k}\frac{(n/2)_r(-k)_r}{(p/2)_r}\frac{1}{r!}\left(-\frac{Q}{nc}\right)^r. \qquad (64)
\end{aligned}$$

If we put $k=$1, 2, 3 and 4 into (64), then

$$W(1)=p\left(1+\frac{Q}{pc}\right),$$

$$W(2)=p(p+2)\left[1+\frac{2Q}{pc}+\frac{n(n+2)}{p(p+2)}\left(\frac{Q}{nc}\right)^2\right],$$

$$W(3)=p(p+2)(p+4)\left[1+\frac{3Q}{pc}+\frac{3n(n+2)}{p(p+2)}\left(\frac{Q}{nc}\right)^2+\frac{n(n+2)(n+4)}{p(p+2)(p+4)}\left(\frac{Q}{nc}\right)^3\right],$$

$$W(4)=p(p+2)(p+4)(p+6)\left[1+\frac{4Q}{pc}+\frac{6n(n+2)}{p(p+2)}\left(\frac{Q}{nc}\right)^2\right.$$
$$\left.+4\frac{n(n+2)(n+4)}{p(p+2)(p+4)}\left(\frac{Q}{nc}\right)^3+\frac{n(n+2)(n+4)(n+6)}{p(p+2)(p+4)(p+6)}\left(\frac{Q}{nc}\right)^4\right].$$

If we substitute these for (59), (60), (61) and (62), and then make arrangements of the equations, mean, variance, skewness and kurtosis of U can be obtained. □

Proof of Theorem 4.3. We only show the proof of Theorem 4.3 (a), because this proof can be applied to the proof of Theorem 4.3 (b) in an obvious way. If we integrate the posterior predictive density of U given in Theorem 4.1 (a) from $-\infty$ to x, then

$$F(x)=\left(\frac{nc}{nc+Q}\right)^{n/2}\int_0^\infty \Phi\left(\frac{x-cv/2}{\sqrt{cv}}\right)g_p(v)\,{}_1F_1\left(\frac{n}{2};\frac{p}{2};\frac{Q}{2(nc+Q)}v\right)dv. \quad (65)$$

If we apply an identity

$${}_1F_1(a;b;x)=e^x\,{}_1F_1\left(b-a;b;-x\right)$$

to (65), then

$$\begin{aligned}F(x)&=\left(\frac{nc}{nc+Q}\right)^{n/2}\int_0^\infty \Phi\left(\frac{x-cv/2}{\sqrt{cv}}\right)g_p(v)\exp\left(\frac{Q}{2(nc+Q)}v\right)\\&\quad\times\sum_{k=0}^{\infty}\frac{\left(\frac{p-n}{2}\right)_k}{\left(\frac{p}{2}\right)_k}\frac{v^k}{k!}\left(\frac{-Q}{2(nc+Q)}\right)^k dv\\&=\sum_{k=0}^{\infty}\frac{\left(\frac{p-n}{2}\right)_k}{k!\left(\frac{p}{2}\right)_k}\left(\frac{-Q}{2(nc+Q)}\right)^k\left(\frac{nc}{nc+Q}\right)^{n/2}\\&\quad\frac{1}{2^{\frac{p}{2}}\Gamma\left(\frac{p}{2}\right)}\int_0^\infty \Phi\left(\frac{x-cv/2}{\sqrt{cv}}\right)v^{\frac{p}{2}+k-1}\exp\left(-\frac{1}{2}\left(1-\frac{Q}{nc+Q}\right)v\right)dv.\end{aligned}$$

If we change the variable to $nc(nc+Q)^{-1}v=v^*$, then

$$F(x)=\sum_{k=0}^{\infty}\frac{\left(\frac{p-n}{2}\right)_k}{k!\left(\frac{p}{2}\right)_k}\left(\frac{-Q}{2(nc+Q)}\right)^k\left(\frac{nc}{nc+Q}\right)^{\frac{n-p}{2}-k}$$

$$\frac{1}{2^{\frac{p}{2}}\Gamma\left(\frac{p}{2}\right)}\int_0^\infty \Phi\left(\frac{x-(c+Q/n)v^*/2}{\sqrt{(c+Q/n)v^*}}\right) v^{*\frac{p}{2}+k-1}\exp\left(-\frac{v^*}{2}\right)dv. \tag{66}$$

If we apply equalities

$$(x)_n = \frac{\Gamma(x+n)}{\Gamma(x)},\ (-x)_n = (-1)^n(x-n+1)_n$$

to (66), then

$$F(x) = \left(\frac{nc}{nc+Q}\right)^{-1}\sum_{k=0}^{\infty}\frac{\Gamma\left(\frac{n-p}{2}+1\right)}{k!\Gamma\left(\frac{n-p}{2}-k+1\right)}\left(1-\frac{nc}{nc+Q}\right)^k\left(\frac{nc}{nc+Q}\right)^{\frac{n-p}{2}-k+1}M(x),$$

with

$$M(x) = \int_0^\infty \Phi\left(\frac{x-(c+Q/n)v/2}{\sqrt{(c+Q/n)v}}\right) g_{p+2k}(v)dv,$$

where $g_{p+2k}(v)$ is the χ^2 density of degrees of freedom $p+2k$. To calculate $F(x)$, we need to compute $M(x)$. First, if we put $x=0$ and $\alpha = c+Q/n$ in $M(x)$, then

$$\begin{aligned}
M(0) &= \int_0^\infty \Phi\left(-\frac{\sqrt{\alpha v}}{2}\right) g_{p+2k}(v)dv \\
&= \frac{1}{\sqrt{2\pi}}\frac{1}{2^{\frac{p}{2}+k}\Gamma(\frac{p}{2}+k)}\int_0^\infty\int_{-\infty}^{-\sqrt{\alpha v}/2} v^{\frac{p}{2}+k-1}\exp\left(-\frac{z^2}{2}-\frac{v}{2}\right)dzdv \\
&= \frac{1}{\sqrt{2\pi}}\frac{1}{2^{\frac{p}{2}+k}\Gamma(\frac{p}{2}+k)}\int_0^\infty\int_{\sqrt{\alpha v}/2}^{\infty} v^{\frac{p}{2}+k-1}\exp\left(-\frac{z^2}{2}-\frac{v}{2}\right)dzdv.
\end{aligned}$$

If we make the change of variable $y=\sqrt{v}$, then

$$M(0) = \sqrt{\frac{2}{\pi}}\frac{1}{2^{\frac{p}{2}+k}\Gamma(\frac{p}{2}+k)}\int_0^\infty\int_{\sqrt{\alpha}y/2}^{\infty} y^{p+2k-1}\exp\left(-\frac{z^2}{2}-\frac{y^2}{2}\right)dzdy. \tag{67}$$

The integral (67) can be evaluated by applying the proof of Lemma D.1 given by Dalton and Dougherty (2011). Next, we consider the case $x \neq 0$. Let $\phi(\cdot)$ be the standard normal density and $G_s(\cdot)$ be the CDF of a χ^2 variable with degrees of freedom s. If we apply the integration by parts to $M(x)$, then

$$\begin{aligned}
M(x) &= \left[\Phi\left(\frac{x-\alpha v/2}{\sqrt{\alpha v}}\right)G_{p+2k}(v)\right]_0^\infty \\
&\quad + \frac{1}{2}\int_0^\infty \phi\left(\frac{x-\alpha v/2}{\sqrt{\alpha v}}\right)\left(\frac{x}{\sqrt{\alpha}}v^{-3/2}+\frac{\sqrt{\alpha}}{2}v^{-1/2}\right)G_{p+2k}(v)dv \\
&= \frac{1}{2}\int_0^\infty \phi\left(\frac{x-\alpha v/2}{\sqrt{\alpha v}}\right)\left(\frac{x}{\sqrt{\alpha}}v^{-3/2}+\frac{\sqrt{\alpha}}{2}v^{-1/2}\right)G_{p+2k}(v)dv \\
&= \frac{e^{x/2}}{2\sqrt{2\pi}}\left[\frac{x}{\sqrt{\alpha}}\int_0^\infty v^{-3/2}\exp\left(-\frac{x^2}{2\alpha}\frac{1}{v}-\frac{\alpha}{8}v\right)G_{p+2k}(v)dv\right.
\end{aligned}$$

$$+\frac{\sqrt{\alpha}}{2}\int_0^\infty v^{-1/2}\exp\left(-\frac{x^2}{2\alpha}\frac{1}{v}-\frac{\alpha}{8}v\right)G_{p+2k}(v)dv\Bigg].$$

If we define

$$N(t)=\int_0^\infty v^{-t/2}\exp\left(-\frac{x^2}{2\alpha}\frac{1}{v}-\frac{\alpha}{8}v\right)G_{p+2k}(v)dv, \tag{68}$$

then,

$$M(x)=\frac{e^{x/2}}{2\sqrt{2\pi}}\left[\frac{x}{\sqrt{\alpha}}N(3)+\frac{\sqrt{\alpha}}{2}N(1)\right]. \tag{69}$$

Equation (68) can be evaluated using the following expressions for $G_{p+2k}(v)$ given in the Appendix of Fatti (1983):

$$G_{p+2k}(v)=\begin{cases}1-\exp\left(-\frac{v}{2}\right)\sum_{i=0}^{p/2+k-1}\frac{1}{i!}\left(\frac{v}{2}\right)^i, & p \text{ is even,}\\ 2\Phi(\sqrt{v})-1-\exp\left(-\frac{v}{2}\right)\sum_{i=0}^{(p-1)/2+k-1}\frac{1}{\Gamma(i+3/2)}\left(\frac{v}{2}\right)^{i+1/2} & p \text{ is odd.}\end{cases}$$

When p is even, (68) becomes

$$\begin{aligned}N(t)=&\,2\left(\frac{2|x|}{\alpha}\right)^{1-t/2}K_{1-t/2}\left(\frac{|x|}{2}\right)\\&-\sum_{i=0}^{p/2+k-1}\frac{2}{2^i i!}\left(\frac{2|x|}{\sqrt{\alpha}\sqrt{\alpha+4}}\right)^{i+1-t/2}K_{i+1-t/2}\left(\frac{|x|}{2}\sqrt{1+\frac{4}{\alpha}}\right),\end{aligned} \tag{70}$$

where we applied the equality

$$\int_0^\infty v^{s-1}e^{-t_1v^{-1}-t_2v}dv=2\left(\frac{t_1}{t_2}\right)^{s/2}K_s\left(2\sqrt{t_1t_2}\right), \tag{71}$$

where $t_1>0$ and $t_2>0$. This equality is given by equation 9 of 3.471 in Gradshteyn and Ryzhik (2007). If we put (70) into (69), then

$$\begin{aligned}M(x)=&\,\frac{e^{x/2}}{2\sqrt{2\pi}}\left\{\frac{x}{\sqrt{\alpha}}\left[2\left(\frac{2|x|}{\alpha}\right)^{-1/2}K_{-1/2}\left(\frac{|x|}{2}\right)\right.\right.\\&\left.-\sum_{i=0}^{p/2+k-1}\frac{2}{2^i i!}\left(\frac{2|x|}{\sqrt{\alpha}\sqrt{\alpha+4}}\right)^{i-1/2}K_{i-1/2}\left(\frac{|x|}{2}\sqrt{1+\frac{4}{\alpha}}\right)\right]\\&+\frac{\sqrt{\alpha}}{2}\left[2\left(\frac{2|x|}{\alpha}\right)^{-1/2}K_{1/2}\left(\frac{|x|}{2}\right)\right.\end{aligned}$$

$$-\sum_{i=0}^{p/2+k-1}\frac{2}{2^i i!}\left(\frac{2|x|}{\sqrt{\alpha}\sqrt{\alpha+4}}\right)^{i+1/2}K_{i+1/2}\left(\frac{|x|}{2}\sqrt{1+\frac{4}{\alpha}}\right)\Bigg]\Bigg\}. \tag{72}$$

If we use the equality

$$K_{1/2}(x)=K_{-1/2}(x)=\sqrt{\frac{\pi}{2x}}\exp(-x)$$

in (72), then the desired result for even p and $x \neq 0$ can be obtained. When p is odd, equation (68) becomes

$$\begin{aligned}
N(t) &= \int_0^\infty v^{-\frac{t}{2}}\exp\left(-\frac{x^2}{2\alpha}\frac{1}{v}-\frac{\alpha}{8}v\right) \\
&\quad\times\left[2\Phi(\sqrt{v})-1-\exp\left(-\frac{v}{2}\right)\sum_{i=0}^{\frac{p-1}{2}+k-1}\frac{1}{\Gamma(i+\frac{3}{2})}\left(\frac{v}{2}\right)^{i+\frac{1}{2}}\right]dv \\
&= \int_0^\infty v^{-\frac{t}{2}}\exp\left(-\frac{x^2}{2\alpha}\frac{1}{v}-\frac{\alpha}{8}v\right) \\
&\quad\times\left[2\int_0^{\sqrt{v}}\phi(z)dz-\exp\left(-\frac{v}{2}\right)\sum_{i=0}^{\frac{p-1}{2}+k-1}\frac{1}{\Gamma(i+\frac{3}{2})}\left(\frac{v}{2}\right)^{i+\frac{1}{2}}\right]dv.
\end{aligned} \tag{73}$$

Since the error function is defined as

$$\operatorname{erf}(z)=\frac{2}{\sqrt{\pi}}\int_0^z e^{-t^2}dt,$$

we can express equation (73) as

$$\begin{aligned}
N(t) &= \int_0^\infty v^{-\frac{t}{2}}\exp\left(-\frac{x^2}{2\alpha}\frac{1}{v}-\frac{\alpha}{8}v\right) \\
&\quad\times\left[\operatorname{erf}\left(\sqrt{\frac{v}{2}}\right)-\exp\left(-\frac{v}{2}\right)\sum_{i=0}^{\frac{p-1}{2}+k-1}\frac{1}{\Gamma(i+\frac{3}{2})}\left(\frac{v}{2}\right)^{i+\frac{1}{2}}\right]dv.
\end{aligned} \tag{74}$$

Using the equality

$$\operatorname{erf}(z)=\frac{2ze^{-z^2}}{\sqrt{\pi}}{}_1F_1\left(1;\frac{3}{2};z^2\right),$$

equation (74) becomes

$$\begin{aligned}
N(t) &= \int_0^\infty v^{-\frac{t}{2}}\exp\left(-\frac{x^2}{2\alpha}\frac{1}{v}-\frac{\alpha}{8}v\right) \\
&\quad\times\left[\sqrt{\frac{2}{\pi}}\sqrt{v}\exp\left(-\frac{v}{2}\right){}_1F_1\left(1;\frac{3}{2};\frac{v}{2}\right)-\exp\left(-\frac{v}{2}\right)\sum_{i=0}^{\frac{p-1}{2}+k-1}\frac{1}{\Gamma(i+\frac{3}{2})}\left(\frac{v}{2}\right)^{i+\frac{1}{2}}\right]dv
\end{aligned}$$

$$= \sqrt{\frac{2}{\pi}} \int_0^\infty v^{(1-t)/2} \exp\left(-\frac{x^2}{2\alpha}\frac{1}{v} - \frac{\alpha+4}{8}v\right) {}_1F_1\left(1; \frac{3}{2}; \frac{v}{2}\right) dv$$
$$- \sum_{i=0}^{\frac{p-1}{2}+k-1} \frac{1}{2^{i+1/2}} \frac{1}{\Gamma(i+3/2)} \int_0^\infty v^{(1-t)/2+i} \exp\left(-\frac{x^2}{2\alpha}\frac{1}{v} - \frac{\alpha+4}{8}v\right) dv. \tag{75}$$

The second term of (75) can be computed as

$$\sum_{i=0}^{\frac{p-1}{2}+k-1} \frac{2}{2^{i+1/2}} \frac{1}{\Gamma(i+3/2)} \left(\frac{2|x|}{\sqrt{\alpha}\sqrt{\alpha+4}}\right)^{\frac{1-t}{2}+1+i} K_{\frac{1-t}{2}+1+i}\left(\frac{|x|}{2}\sqrt{1+\frac{4}{\alpha}}\right). \tag{76}$$

by using the equality (71), while the first term of (75) as

$$\sqrt{\frac{2}{\pi}} \int_0^\infty v^{(1-t)/2} \exp\left(-\frac{x^2}{2\alpha}\frac{1}{v} - \frac{\alpha+4}{8}v\right) \int_0^\infty e^{-w} {}_0F_1\left(\frac{3}{2}; \frac{vw}{2}\right) dwdv$$
$$= 2\int_0^\infty v^{-t/2} \exp\left(-\frac{x^2}{2\alpha}\frac{1}{v} - \frac{\alpha}{8}v\right)$$
$$\times \int_0^\infty \frac{v^{3/2-1}e^{-v/2}}{2^{3/2}\Gamma(3/2)} e^{-2w/2} {}_0F_1\left(\frac{3}{2}; \frac{2w}{4}v\right) dwdv. \tag{77}$$

Since the integrand in (77) is the noncentral chi squared density with degrees of freedom 3 and noncentrality parameter $2w$ and Corollary 1.3.5 of Muirhead (1982), (77) becomes

$$2\sum_{i=0}^\infty \int_0^\infty v^{-t/2} \exp\left(-\frac{x^2}{2\alpha}\frac{1}{v} - \frac{\alpha}{8}v\right) g_{3+2i}(v) dv$$
$$= \sum_{i=0}^\infty \frac{1}{2^{1/2+i}\Gamma(3/2+i)} \int_0^\infty v^{(3+2i-t)/2-1} \exp\left(-\frac{x^2}{2\alpha}\frac{1}{v} - \frac{\alpha+4}{8}v\right) dv$$
$$= \sum_{i=0}^\infty \frac{2}{2^{1/2+i}\Gamma(3/2+i)} \left(\frac{2|x|}{\sqrt{\alpha}\sqrt{\alpha+4}}\right)^{\frac{1-t}{2}+1+i} K_{\frac{1-t}{2}+1+i}\left(\frac{|x|}{2}\sqrt{1+\frac{4}{\alpha}}\right). \tag{78}$$

Putting (76) and (78) into (75),

$$N(t) = \sum_{i=(p-1)/2+k}^\infty \frac{2}{2^{1/2+i}\Gamma(3/2+i)} \left(\frac{2|x|}{\sqrt{\alpha}\sqrt{\alpha+4}}\right)^{\frac{1-t}{2}+1+i}$$
$$K_{\frac{1-t}{2}+1+i}\left(\frac{|x|}{2}\sqrt{1+\frac{4}{\alpha}}\right). \tag{79}$$

If we put (79) into (69), then we can obtain the result for odd p and $x \neq 0$. □

Proof of Corollary 4.4 *and* 4.5. Since $(n-p)/2$ is an even number,

$$ {}_1F_1\left(\frac{n}{2}; \frac{p}{2}; \frac{Q}{2(nc+Q)}v\right) = \exp\left(\frac{Q}{2(nc+Q)}v\right) {}_1F_1\left(\frac{p-n}{2}; \frac{p}{2}; \frac{Q}{2(nc+Q)}v\right)$$

$$= \sum_{k=0}^{|p-n|/2} \frac{\left(\frac{p-n}{2}\right)_k}{\left(\frac{p}{2}\right)_k} \frac{v^k}{k!} \left(\frac{-Q}{2(nc+Q)}\right)^k .$$

Thus, $F(x)$ is represented by

$$F(x) = \sum_{k=0}^{|p-n|/2} \frac{\left(\frac{p-n}{2}\right)_k}{k! \left(\frac{p}{2}\right)_k} \left(\frac{-Q}{2(nc+Q)}\right)^k \left(\frac{nc}{nc+Q}\right)^{n/2}$$
$$\frac{1}{2^{\frac{p}{2}}\Gamma\left(\frac{p}{2}\right)} \int_0^\infty \Phi\left(\frac{x-cv/2}{\sqrt{cv}}\right) v^{\frac{p}{2}+k-1} \exp\left(-\frac{1}{2}\left(1-\frac{Q}{nc+Q}\right)v\right) dv.$$

The rest of the proof is the same as that of Theorem 4.3 (a). In addition, the finite representation for $K_{i\pm 1/2}$ is found in 8.468 and 8.469 of Gradshteyn and Ryzhik (2007). □

Proof of Lemma 4.6. Since mean and variance of X/f are given by $\mathrm{E}(X/f) = \delta$ and $\mathrm{V}(X/f) = \gamma/f$ with $\gamma = \Omega + 2\delta^2$, the conditional distribution of the standardized X/f is given by

$$S_f \sim N\left(-\frac{\sqrt{f}\delta}{\sqrt{\gamma}} + \frac{\delta}{g\sqrt{f\gamma}}(gW), (gW)\right),$$

where $W \sim \chi^2_f$ and $g = \Omega/(f\gamma)$. From Definition 2.1 of Barndorff-Nielsen et al. (1982), the distribution of S_f is the normal variance-mean mixture with position $-\sqrt{f}\delta/\sqrt{\gamma}$, drift $\delta/(g\sqrt{f\gamma})$, structure matrix 1, and mixing distribution $gW \overset{\mathrm{d}}{=} g\chi^2_f$. Hence, from equation (2.2) of Barndorff-Nielsen et al. (1982), the characteristic function of S_f is given by

$$\hat{g}(\theta) = \exp\left(-\frac{i\theta\sqrt{f}\delta}{\sqrt{\gamma}}\right)\left(1 - \frac{2i\theta\delta}{\sqrt{f\gamma}} + \frac{\theta^2\Omega}{f\gamma}\right)^{-f/2}.$$

The cumulant generating function of S_f becomes

$$\begin{aligned}
K(\theta) &= -\frac{i\theta\sqrt{f}\delta}{\sqrt{\gamma}} - \frac{f}{2}\log\left(1 - \frac{2i\theta\delta}{\sqrt{f\gamma}} + \frac{\theta^2\Omega}{f\gamma}\right) \\
&= -\frac{i\theta\sqrt{f}\delta}{\sqrt{\gamma}} - \frac{f}{2}\sum_{r=1}^\infty \frac{(-1)^{r+1}}{r}\left(-\frac{2i\theta\delta}{\sqrt{f\gamma}} + \frac{\theta^2\Omega}{f\gamma}\right)^r \\
&= -\frac{i\theta\sqrt{f}\delta}{\sqrt{\gamma}} - \frac{1}{2}\sum_{r=1}^\infty \frac{(-1)^{r+1}}{r}\sum_{s=0}^r \binom{r}{s}\left(-\frac{2i\delta}{\sqrt{\gamma}}\right)^{r-s}\left(\frac{\Omega}{\gamma}\right)^s\left(\frac{\theta}{\sqrt{f}}\right)^{r+s}\left(\frac{1}{\sqrt{f}}\right)^{-2}.
\end{aligned}$$

For $l \geq 2$, the l th derivative of $K(\theta)$ is given by

$$K^{(l)}(\theta) = -\frac{1}{2}\sum_{r=1}^\infty \frac{(-1)^{r+1}}{r}\sum_{s=0}^r \binom{r}{s}\left(-\frac{2i\delta}{\sqrt{\gamma}}\right)^{r-s}\left(\frac{\Omega}{\gamma}\right)^s$$
$$\times (r+s)(r+s-1)\cdots(r+s-l+1)\left(\frac{\theta}{\sqrt{f}}\right)^{r+s-l}\left(\frac{1}{\sqrt{f}}\right)^{l-2}. \qquad (80)$$

To compute the lth cumulant, we express (80) as

$$K^{(l)}(\theta) = \sum_{r+s=l, s\le r} \left[-\frac{1}{2}\frac{(-1)^{r+1}}{r}\binom{r}{s}\left(-\frac{2i\delta}{\sqrt{\gamma}}\right)^{r-s}\left(\frac{\Omega}{\gamma}\right)^{s} \right.$$
$$\left. \times (r+s)(r+s-1)\cdots(r+s-l+1)\left(\frac{\theta}{\sqrt{f}}\right)^{r+s-l}\left(\frac{1}{\sqrt{f}}\right)^{l-2} \right]$$
$$+ \sum_{r+s\neq l, s\le r} \left[-\frac{1}{2}\frac{(-1)^{r+1}}{r}\binom{r}{s}\left(-\frac{2i\delta}{\sqrt{\gamma}}\right)^{r-s}\left(\frac{\Omega}{\gamma}\right)^{s} \right.$$
$$\left. \times (r+s)(r+s-1)\cdots(r+s-l+1)\left(\frac{\theta}{\sqrt{f}}\right)^{r+s-l}\left(\frac{1}{\sqrt{f}}\right)^{l-2} \right]. \qquad (81)$$

If we put $l = 3$ in (81), then the first summation becomes

$$-\frac{1}{2}\left[\frac{(-1)^3}{2}\binom{2}{1}\left(-\frac{2i\delta}{\sqrt{\gamma}}\right)\left(\frac{\Omega}{\gamma}\right)3!\frac{1}{\sqrt{f}} + \frac{(-1)^4}{3}\binom{3}{0}\left(-\frac{2i\delta}{\sqrt{\gamma}}\right)^3 3!\frac{1}{\sqrt{f}}\right]$$
$$= -i\frac{2\delta(3\Omega+4\delta^2)}{\gamma^{3/2}}\frac{1}{\sqrt{f}}.$$

Hence, the third cumulant of S_f is given by

$$\frac{2\delta(3\Omega+4\delta^2)}{(2\delta^2+\Omega)^{3/2}}\frac{1}{\sqrt{f}}.$$

The fourth cumulant of S_f can be obtained by following the procedures to obtain the third cumulant of S_f. The characteristic function of S_f is represented by

$$\hat{g}(\theta) = \exp\left\{-\frac{1}{2}\theta^2 + \frac{1}{\sqrt{f}}\frac{1}{3!}\tilde{\kappa}_3(i\theta)^3 + \frac{1}{f}\frac{1}{4!}\tilde{\kappa}_4(i\theta)^4 + \cdots\right\}, \qquad (82)$$

where

$$\tilde{\kappa}_3 = \frac{2\delta(3\Omega+4\delta^2)}{(2\delta^2+\Omega)^{3/2}} \text{ and } \tilde{\kappa}_4 = \frac{6(\Omega^2+8\Omega\delta^2+8\delta^4)}{(2\delta^2+\Omega)^2}.$$

Since (82) is identical to (2.7) of Hall (1992), the Edgeworth expansion of S_f is followed immediately. □

5 Conclusion and future problems

There are a lot of papers concerning the properties of a Wishart matrix and a normal vector, respectively. However, these random objects does not always appear in isolation. For instance, the Hotelling's T^2 statistics are expressed as the combination of a Wishart matrix and a normal vector in the form of quadratic forms, while linear discriminant coefficients are expressed as the product of an inverse Wishart matrix and a normal vector (cf. Siotani et al. 1985). Consequently, in this dissertation, we have investigated the distribution of some functions of a Wishart matrix and a normal vector.

In section 2, we derive the stochastic representation of the product of a Wishart matrix and a normal vector with different covariance matrices. The derivation is based on the result of Bodnar et al. (2013). The obtained representation enables us to construct the density function and to express the higher order moment of the product. As for the moment, we provided the explicit expression of the first four moments of the product. We can observe from the moment formulae that the distribution of the product is mainly affected by degrees of freedom of a Wishart matrix n, and dimension k; if n or k is large, this distribution could be approximated by a normal distribution well. In addition, the large eigenvalues of the product of a different covariance matrices make the distribution less skewed. Using the first four moments of the product, we provided a Edgeworth type expansion given by Theorem 3.2.2 of Kollo and von Rosen (2005). Comparing the Edgeworth type expansion with kernel density estimator of the product, the good performance of the Edgeworth type expansions is documented for moderately large n. Even though the Edgeworth type expansion with small n performed badly, the performance of the Edgeworth type expansions is greatly improved for large dimension. Since the error bound for the Edgeworth expansion derived in section 2 has not been evaluated, the evaluation of the error bound will be an important issue in the future. In addition, we proposed a Edgeworth type expansion for linear combinations of the product, and a future work may be to provide Edgeworth type expansion for linear transformations of the product. The product discussed in section 2 corresponds to a semi-conjugate prior for the weights of the contact portfolio and the coefficients of the linear discriminant function (cf. Hoff, 2009). Therefore, the discussion in this section can be applied to Bayesian inference for the weights of tangency portfolios and coefficients of the linear discriminant function.

In section 3, we consider the distribution of the product of a Wishart matrix and a conditional normal vector given a Wishart matrix. This kind of product appears in Bayesian statistics. Our stochastic representation of the product improved the existing representations in terms of computational efficiency. We can observe from the stochastic representation of the product that the distribution of the product belongs to the family of generalized hyperbolic distributions (cf. Barndorff-Nielsen, 1982; Blæsild, 1981; Blæsild, 1982). which indicate that the distribution is closed under affine transformation, marginalization, and conditioning. Some useful formulae—density function, first four moments, Edgeworth expansion of the product—have been provided. In Appendix, we have improved the computational efficiency of some stochastic representations given in Bauder et al. (2019). As a future work, we would like to consider the product of the matrix variate generalized inverse Gaussian distribution (MGIG), which includes the Wishart matrix as a special case, and the multivariate normal distribution. For the properties of MGIG distribution, readers may refer to Butler (1998), Seshadri (2003), Massam and Wesołowski (2006), and Seshadri and Wesołowski (2008)

In section 4, we discuss the posterior predictive distribution of the linear discriminant in a Bayesian setting. We derived the first four moments and the density function of the population linear discriminant under the Jeffreys and normal inverse Wishart prior distributions. We applied these results to Bayesian estimation of the optimal error rate, which is associated with population linear discriminant. Although, the obtained estimator is generally expressed by infinite series or some special functions, we can represent the estimator only as a finite sum and an elementary function under certain conditions. In addition, we suggested the Edgeworth type approximations for the estimator of the optimal error rate, which is generally documented the better result than existing approximations for the estimator of the optimal error rate. According to Geisser (1967, 1982), "The estimation of $\epsilon = q_1\epsilon_1 + q_2\epsilon_2$, the total optimal error rate, is useful as a guide to the optimal discriminatory power of the variables used for allocation. If the estimate of ϵ indicates that ϵ is larger than the accuracy required for the allocation procedure, one would search for additional or another set of variables that would diminish the total error rate. " Therefore, the obtained results in the section 4 are useful for the selection of variables for linear discriminant analysis. Further, it seems that the obtained Bayes estimator crucially depends on the choices of hyper-parameters of the prior distributions. Thus, it is important to propose (theoretically supported) guidelines on some choices of hyper-parameters for the prior distributions. This is left for future works. In addition to the linear discriminant function in two multivariate normal populations treated in this paper, there are other important discriminant procedures. There are, for example, quadratic discriminant procedure that arise when the covariance matrices of two normal populations are different, linear discriminant procedure in k multivariate normal discrimination, and discriminant procedure obtained by the Between-Within Method (cf. Siotani et al. 1985). Also, in the situation where the correlation between variables is extremely strong, the covariance matrix will be close to be singular. In this case, we can also consider the linear discriminant procedure that arises when $\mathbf{\Sigma}^{-1}$ is replaced by the Moore-Penrose inverse $\mathbf{\Sigma}^{+}$ (cf. Harville, 1997). A future task is to estimate the probabilities of misclassification associated with these discriminant procedures under the Bayesian framework.

References

[1] Anderson, T. W. (1973). *An asymptotic expansion of the distribution of the studentized classification statistic W*, *The Annals of Statistics*, **1**(5), 964-972.

[2] Anderson, T, W. (2003). *An introduction to multivariate statistical analysis*, Third Edition, Wiley-Interscience.

[3] Abramowitz, M., & Stegun, I. A. (1965). *Handbook of mathematical functions : with formulas, graphs, and mathematical tables*. Dover Publications.

[4] Barndorff - Nielsen, O., Kent, J., & Sorensen, M. (1982). Normal variance - mean mixtures and z distributions. *International Statistical Review*, **50**(2), 145-159.

[5] Bauder, D., Bodnar, R., Bodnar, T., & Schmid, W. (2019). Bayesian estimation of the efficient frontier, *Scandinavian Journal of Statistics*, **46**, 802 - 830.

[6] Bauder, D., Bodnar, T., Mazur, S., & Okhrin, Y. (2018). Bayesian inference for the tangency portfolio. *International Journal of Theoretical and Applied Finance*, **21**, 1-27.

[7] Bauder, D., Bodnar, T., Paloya, N., & Schmid, W (2020). Bayesian inference of the multi-period optimal portfolio for an exponential utility. *Journal of Multivariate Analysis*, **175**, 1-22.

[8] Blæsild, P. (1981). The two-dimensional hyperbolic distribution and related distributions, with an application to Johannsen's data. *Biometrika*, **68**, 251-263.

[9] Blæsild, P., & Jensen, J, L. (1981). Multivariate distributions of hyperbolic type, *Statistical Distributions in Scientific Work*(eds. Taillie, C., Patil, G, P., & Baldesari, A), Vol.4, 45-66, Reidel, Dordrecht.

[10] Bodnar, T. (2004). Optimal portfolios in an elliptical model – Statistical analysis and tests for efficiency, PhD thesis, European University Viadrina, Frankfurt (Oder), Germany.

[11] Bodnar, T., Mazur, S., Muhinyuza, S. & N. Parolya (2019). On the product of a singular Wishart matrix and a singular Gaussian vector in high dimension. *Theory of Probability and Mathematical Statistics*, **99**, 39-52.

[12] Bodnar, T., Mazur, S., Ngailo, E., & Parolya, N. (2020). Discriminant analysis in small and large dimensions. *Theory of Probability and Mathematical Statistics*, **100**, 21-41. https://doi.org/10.1090/tpms/1096

[13] Bodnar, T., Mazur, S., & Okhrin, Y. (2013). On the exact and approximate distributions of the product of a Wishart matrix with a normal vector, *Journal of Multivariate Analysis*. **122**, 70 - 81.

[14] Bodnar, T., Mazur, S. & Y. Okhrin (2015). Distribution of the product of singular Wishart matrix and normal vector, *Theory of Probability and Mathematical Statistics*, **91**, 1-15.

[15] Bodnar, T., Mazur, S. & Y. Okhrin (2017). Bayesian estimation of the global minimum variance portfolio, *European Journal of Operational Research*, **256**, 292-307.

[16] Bodnar, T., Mazur, S. & K. Podgórski. (2016). Singular inverse Wishart distribution and its application to portfolio theory, *Journal of Multivariate Analysis*, **143**, 314-326.

[17] Bodnar, T., Mazur, S., & Parolya, N. (2019). Central limit theorems for functionals of large sample covariance matrix and mean vector in matrix variate location mixture of normal distributions, *Scandinavian Journal of Statistics*, **46**(2), 636 - 660.

[18] Bodnar, T., & Okhrin, Y. (2008). Properties of the singular, inverse and generalized inverse partitioned Wishart distributions. *Journal of Multivariate Analysis*, **99**, 2389-2405.

[19] Bodnar, T., & Okhrin, Y. (2011). On the product of inverse Wishart and normal distributions with applications to discriminant analysis and portfolio theory, *Scandinavian Journal of Statistics*, **38**, 311 - 331.

[20] Bodnar, T., & Okhrin, Y. (2019). Tangency portfolio weights for singular covariance matrix in small and large dimensions: Estimation and test theory, *Journal of Statistical Planning and Inference*, **201**, 40 - 57.

[21] Bodnar, T., & Schmid, W. (2006). Distributional properties of portfolio weights, *Journal of Econometrics*, **134**, 235 - 256.

[22] Bodnar, T., & Schmid, W. (2008). A test for the weights of the global minimum variance portfolio in an elliptical model, *Metrika*, **67**, 127 - 143.

[23] Bodnar, T., & Schmid, W. (2009). Econometrical analysis of the sample efficient frontier, *European Journal of Finance*, **15**(3), 317 - 335.

[24] Bowker, A. (1961). Representation of Hotelling's T2 and Anderson's classification statistics W in terms of simple statistics. In H. Solomon (Eds.), *Studies in Item analysis and prediction* (pp. 285-292). Stanford University Press.

[25] Box, G. E. P., & Tiao, G. C. (1973). *Bayesian inference in statistical analysis*, Addison-Wesley, London.

[26] Butler, R. W (1998). Generalizd inverse Gaussian distributions and their Wishart connections, *Scandinavian Journal of Statistics*, **25**, 69 - 75.

[27] Dalton, L. A., & Dougherty, E. R. (2011a). Bayesian minimum mean-square error estimation for classification error - Part I: Definition and the Bayesian MMSE error estimator for discrete classification. *IEEE Transactions on signal processing*, **59**(1), 115-129.

[28] Dalton, L. A., & Dougherty, E. R. (2011b). Bayesian minimum mean-square error estimation for classification error - Part II: The Bayesian MMSE error estimator for linear classification of Gaussian distributions. *IEEE Transactions on signal processing*, **59**(1), 130-144.

[29] Díaz-García, J.A., Gitiérrez-Jáimez, R. & Mardia, K.V. (1997). Wishart and pseudo-Wishart distributions and some applications to shape theory. *Journal of Multivariate Analysis*, **63**, 73-87.

[30] Drton, M., Massam, H., & Olkin, I. (2008). Moments of minors of Wishart matrices, *The Annals of Statistics*, **36**(5), 2261-2283.

[31] Fatti, L. P. (1982). Predictive discrimination under the random effects model. *South African Statistical Journal*, **16**, 55-77.

[32] Fatti, L. P. (1983). The random-effects model in discriminant analysis. *Journal of the American Statistical Association*, **78**(383), 679-687.

[33] Fujikoshi, Y. (1987). Error bounds for asymptotic expansion of the distribution of the MLE in a gmanova model, *Annals of Institute of Statistical Mathematics*, **39**, 153 - 161.

[34] Fung, T., & Seneta, E (2011). Tail dependence and skew distributions, *Quantitive Finance*, **11**, 327-333.

[35] Fung, T., & Seneta, E (2021). Tail asymptotics for the bivariate equi-skew generalized hyperbolic distribution and its Variance-Gamma special case, *Statistics and Probability Letters*, **178**.

[36] Geisser, S. (1964). Posterior Odds for Multivariate Normal Classifications, *Journal of the Royal Statistical Society. Series B (Methodological)*, **26**(1), 69-76.

[37] Geisser, S. (1966). Predictive discrimination, *Multivariate Analysis*, In P. R. Krishnaiah (Eds.), Academic Press, New York, (pp.149-163).

[38] Geisser, S. (1967). Estimation associated with linear discriminants. *The Annals of Mathematical Statistics*, **38**(3), 807-817.

[39] Geisser, S. (1982). Bayesian discrimination. In P. R. Krishnaiah, & L. N. Kanal (Eds.), *Handbook of Statistics: Classification, pattern recognition, and reduction of dimensionality* (Vol 2., pp. 101-120). North-Holland Publish Company.

[40] Gradshteyn, I. S & Ryzhik, I. M. (2007). *Table of integrals, series, products*. Elsevier.

[41] Gupta, A. K & Nagar, D. K. (2000). *Matrix Variate Distributions*, Chapman and Hall/CRC, Boca Raton.

[42] Gupta, A. K., Varga, T., & Bodnar, T. (2013), *Elliptically contoured models in statistics and portfolio theory*, Springer, New York.

[43] Haff, L. R. (1977). Minimax estimators for a multinormal precision matrix, *Journal of Multivariate Analysis*, **7**(3), 374-385.

[44] Haff, L. R. (1979a). An identity for the Wishart distribution with applications, *Journal of Multivariate Analysis*, **9**(4), 531 - 544.

[45] Haff, L. R. (1979b). Estimation of the Inverse Covariance Matrix: Random Mixtures of the Inverse Wishart Matrix and the Identity, *Journal of Multivariate Analysis*, **7**(6), 1264-1276.

[46] Haff, L. R. (1979c). Estimation of the Inverse Covariance Matrix: Random Mixtures of the Inverse Wishart Matrix and the Identity. *The Annals of Statistics*, **7**(6), 1264-1276.

[47] Haff, L. R. (1980). Empirical Bayes Estimation of the Multivariate Normal Covariance Matrix, *The Annals of Statistics*, **8**(3), 586-597.

[48] Hall, P. (1992). *The bootstrap and Edgeworth expansion*. Springer - Verlag.

[49] Harville, D. A. (1997). *Matrix algebra from a statistician's perspective*. Springer.

[50] Hillier, G., & Kan, R. (2021). *Moments of a Wishart matrix*, To appear in *Journal of Quantitative Economics*.

[51] Hoff, P.D. (2009). *A first course in Bayesian statistical methods*, Springer, New York.

[52] Isserlis, L. (1918). On a formula for the product-moment coefficient of any order of a normal frequency distribution in any number of variables, Biometrika. **12**, 134 - 139.

[53] Javed, F., Mazur, S., & Ngailo, E. (2021). Higher order moments of the estimated tangency portfolio weights, *Journal of Applied Statistics*, **48**, 517-535.

[54] Javed, F., Loperdo, N., & Mazur,S. (2021). Edgeworth expansions for multivariate random sums. To appear in *Econometrics and Statistics*.

[55] Jobson, J. D. (1991). Confidence region for the mean - variance efficient set: An alternative approach to estimation risk, *Review of Quantitative Finance and Accounting*, **1**, 235 - 257.

[56] Jobson, J. D., & Korkie, B. (1980). Estimation of Markowitz efficient portfolios, *Journal of the American Statistical Association*, **75**, 544 - 554.

[57] Kan, R. (2008). From moments of sum to moments of product, *Journal of Multivariate Analysis*, **99**, 542 - 554.

[58] Kan, R & Smith, D. R. (2008). The distribution of the sample minimum variance frontier, *Management Science*, **54**(7), 1364-1380.

[59] Kotsiuba, I., & Mazur, S. (2016). On the asymptotic and approximate distributions of the product of an inverse Wishart matrix and a Gaussian vector. *Theory of Probability and Mathematical Statistics*, **93**, 103-112.

[60] Kubokawa, T., Hyodo, M., & Srivastava, M. S. (2013). Asymptotic expansion and estimation of EPMC for linear classification rules in high dimension, *Journal of Multivariate Analysis*. **115**, 496 - 515.

[61] Lehmann, E. L. & Casella, G. (1998). *Theory of point estimation*, Second Edition, Springer.

[62] Letac, G & Massam, H. (2004). All invariant moments of the Wishart distribution, *Scandinavian Journal of Statistics*, **31**, 295 - 318.

[63] Markowitz, H. (1952). Portfolio selection, *The Journal of Finance*, **7**, 77 - 91.

[64] Massam, H. & Wesołowski, J. (2006). The Matsumoto-Yor property and the structure of the Wishart distribution, *Journal of Multivariate Analysis*, **97**, 103 - 123.

[65] Mathai, A. M & Provost, S. B. (1992). *Quadratic forms in random variables*, Mercel Dekker.

[66] Merton, R. C. (1972). An analytical derivation of the efficient frontier, *Journal of Financial and Quantitative Analysis*, **7**, 1851 - 1872.

[67] Muirhead, R. J. (1982). *Aspects of multivariate statistical theory*, Wiley.

[68] Nelsen, R. B. (2006). *An introduction to copulas*, Second Edition, Springer.

[69] Okamoto, M. (1963). An asymptotic expansion for the distribution of the linear discriminant function. *Annals of Mathematical Statistics*, **34**, 1286-1301.

[70] Rao, C. R. (1973). *Linear statistical inference and its applications*, Second Edition, John Wiliey & Sons.

[71] Rigby, R. A. (1982). A Credibility Interval for the Probability that a New Observation Belongs to One of Two Multivariate Normal Populations, *Journal of the Royal Statistical Society: Series B (Methodological)*, **44**(2), 212-220.

[72] Rigby, R. A. (1997). Bayesian Discrimination between Two Multivariate Normal Populations with Equal Covariance Matrices, *Journal of the American Statistical Association*, **92**(439), 1151-1154.

[73] Seneta, E. (1976). *Regularly Varying Functions*. In: Lecture Notes in Mathematics, vol 508, Springer, Berlin.

[74] Seshadri, V. (2003). Some properties of the matrix generalized inverse Gaussian distribution. In: Balakrishnan, N., Kannan, N., and Srinivasan, M. R. (eds), *Statistical methods and practice. Recent advances*, (pp. 47-56), Narosa Publishing House, New Delhi.

[75] Seshadri, V., & Wesołowski. (2008). More on connections between Wishart and matrix GIG distributions, *Metrika*, **68**, 219-232.

[76] Siotani, M., Hayakawa, T., & Fujikoshi, Y. (1985). *Modern multivariate statistical analysis : a graduate course and handbook*, American series in mathematical and management sciences ; (Vol. 9). Columbus, Ohio : American Sciences Press.

[77] Siotani, M., & Wang, R. H. (1977). Asymptotic expansions for error rates and comparison of the W-procedure and the Z-procedure in discriminant analysis. In P. R. Krishnaiah (Eds.), *Multivariate Analysis IV: proceedings of the fourth International Symposium on Multivariate Analysis* (pp.523-545). Elsevier/North Holland.

[78] Srivastava, M. S. (2003). Singular Wishart and multivariate beta distributions, *The Annals of Statistics*, **31**(5), 1537-1560.

[79] Stein, C. M. (1981). Estimation of the Mean of a Multivariate Normal Distribution, *The Annals of Statistics*, **9**(6), 1135-1151.

[80] Sultan, S. A., & Tracy, D. S. (1996). Moments of Wishart distribution, *Stochastic Analysis and Applications*, **14**(2), 237-243.

[81] von Rosen, D. (1979). Distribution and density approximations of a maximum likelihood estimator in the growth curve model, *Statistics*. **29**, 37 - 59.

[82] von Rosen, D. (1991). Moments for the inverted Wishart distribution,*Scandinavian Journal of Statistics*, **15**, 97 - 109.

[83] von Rosen, D. (1991). Moments of maximum likelihood estimators in the growth curve model, *Statistics*. **22**, 111 - 131.

[84] Watamori, Y. (1990), Moments of the traces of Wishart and inverted Wishart matrices, *South African Statistical Journal*, **24**(2), 153-176.

[85] Yonenaga, K., & Suzukawa, A. (2021). Bayesian estimation for misclassification rate in linear discriminant analysis, *Japanese Journal of Statistics and Data Science*, **4**, 861-885.