MIXTURES OF DISCRETE AND CONTINUOUS VARIABLES: CONSIDERATIONS FOR DIMENSION REDUCTION

by

John R. Pleis

MS, The George Washington University, 2001

BS, University of Maryland - University College, 1998

BS, Peru State College, 1990

Submitted to the Graduate Faculty of

the Department of Biostatistics

Graduate School of Public Health in partial fulfillment

of the requirements for the degree of

Doctor of Philosophy

University of Pittsburgh

2018

UNIVERSITY OF PITTSBURGH

GRADUATE SCHOOL OF PUBLIC HEALTH

This dissertation was presented

by

John R. Pleis

It was defended on

May 30, 2018

and approved by

Stewart J. Anderson, PhD, Professor

Department of Biostatistics

Graduate School of Public Health

University of Pittsburgh

Chung-Chou Ho Chang, PhD, Professor

Department of Biostatistics

Graduate School of Public Health

University of Pittsburgh

Sungkyu Jung, PhD, Associate Professor

Department of Statistics

Kenneth P. Dietrich School of Arts and Sciences

University of Pittsburgh

Chaeryon Kang, PhD, Assistant Professor

Department of Biostatistics

Graduate School of Public Health

University of Pittsburgh

Dissertation Director: Stewart J. Anderson, PhD, Professor

Department of Biostatistics

Graduate School of Public Health

University of Pittsburgh

MIXTURES OF DISCRETE AND CONTINUOUS VARIABLES: CONSIDERATIONS FOR DIMENSION REDUCTION

John R. Pleis, PhD

University of Pittsburgh, 2018

ABSTRACT

For this dissertation, we will examine mixtures of different types of data, the analytic challenges that such data can present, and some approaches for addressing this issue. Specifically, we will consider mixtures of continuous and discrete data. For the theoretical developments that follow, we will focus on the general location model (GLOM)-based methodology for deriving the joint probability distribution of continuous and discrete random variables as the product of conditional and marginal probability distributions. As we will show, the general specification of this joint distribution is a finite mixture of Gaussian distributions. We will consider both the univariate and multivariate cases. For the univariate case we will first determine the distribution of the sample variance, and for the multivariate case we will first determine the distribution of the sample covariance matrix. When the component distributions of the mixture have different variances (univariate) or covariance matrices (multivariate), any analysis can become more challenging. In such cases, we propose approximating the mixture density with a non-mixture density from the same parametric family (e.g., multivariate Gaussian). Finally, we will present some extensions of this work to the field of dimension reduction.

Public Health Significance: Mixtures of continuous and discrete variables are somewhat common in public health settings (e.g., genetics, health services research), but statistical

methods for the analysis of such data are not nearly as developed and robust, compared to the analysis of only one type of data (e.g., continuous). The methods developed in this dissertation could be used to expand inferential approaches to non-normal data which are commonly seen in public health settings. For example, hypothesis testing of the proportionate contribution of eigenvalues could be adapted to mixtures of different types of data, and these methods could possibly be extended to high-dimensional data (e.g., genetics) by examining mixtures of singular Wishart distributions.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

1.0 AN OVERVIEW OF MIXTURE DISTRIBUTIONS

In the most general sense, the term "mixture data" refers to situations where the variable data structure is comprised of multiple data types (e.g., continuous, discrete). The paradigm can be further expanded outward if we consider the various types of discrete data (e.g., binary, nominal, ordinal, count) as separate data constructs. While mixtures of continuous and discrete variables are somewhat common in research settings, statistical methods for the analysis of such data are not nearly as developed and robust, compared to the analysis of only one type of data (e.g., continuous). Because of the slower pace of methodological development for the analysis of mixture data, there may be a tendency to coarsen the data so that methods developed for a single type of data could be applied. For example, some of these available options include: categorizing all of the continuous data and then analyzing the combined data using categorical data analysis methods; applying a scoring method to the categorical variables and subsequently analyzing the data using approaches developed for continuous data; or analyzing each variable type separately and then combining the results together based on some weighting mechanism. However, as illustrated by Krzanowski [1], these approaches are not necessarily ideal. The categorization of all continuous variables leads to a loss of information, and the scaling of all categorical variables introduces an unknown degree of subjectivity. Further, analyzing each variable type separately (and later combining) ignores any association between the continuous and discrete variables. Because of these consequences, the preferred approach for the analysis of multiple types of data usually starts with the specification of the joint distribution of the continuous and discrete random variables. If the different types of random variables are statistically independent, then the derivation of this joint distribution is relatively straightforward (product of the marginal distributions). However, if this assumption is not tenable, then an alternative way of calculating the joint distribution must be used. Thus, herein lies one of the primary difficulties in the analysis of multiple types of data: the lack of flexible methods for the specification of the joint distribution of the continuous and discrete random variables without the assumption of statistical independence. For this purpose, there are generally three approaches that have been demonstrated in the literature: specifying the joint distribution as a product of conditional and marginal distributions; using copula models to derive the joint distribution; or incorporating latent variables into the analysis. Typically, the latent variable approach assumes that the discrete variables are the realization of an unobserved continuous random variable. For example, let Y be an observed binary random variable taking on the values of 0 and 1. Further, let Y^* be an unobserved random variable following an, as of yet, unspecified continuous distribution. Assuming that Y has an underlying continuity represented by Y^*, then $Pr(Y = 1) = Pr(Y^* > \tau)$, where τ is an unknown thresholding parameter. Similarly, $Pr(Y = 0) = Pr(Y^* \leq \tau)$. In the following sections, these approaches will be described in more detail, including any underlying assumptions as well as examples from the literature for each approach.

1.1 FINDING THE JOINT PROBABILITY DISTRIBUTION OF CONTINUOUS AND DISCRETE DATA

1.1.1 Product of Conditional and Marginal Distributions

First, let us assume that we have two random variables X and Y. In addition, let $f_{X,Y}(x, y)$ be defined as their joint probability distribution. From introductory coursework in probability theory, we know that the joint distribution of any two random variables can be written as:

$$f_{X,Y}(x,y) = f_{X|Y}(x|y)f_Y(y) = f_{Y|X}(y|x)f_X(x) \tag{1.1}$$

For the mixture data setting, if we let the probability distribution for X be continuous and the probability distribution for Y be discrete, then (1.1) becomes:

$$f_{X,Y}(x,y) = Pr(x|Y=y)Pr(Y=y) = Pr(Y=y|x)f_X(x) \tag{1.2}$$

As can be seen from (1.1) and (1.2), these equations require knowledge of the form of the respective conditional probability distribution. If the form of the conditional probability distribution is known, the joint distribution between the two random variables can be expressed in a straightforward manner. Depending on the nature of the problem, one particular form of the conditional distributions specified in (1.2) may be easier to deal with than the other.

1.1.1.1 Using $Pr(x|Y=y)Pr(Y=y)$.

Some of the earliest statistical methodology for the analysis of multiple types of data was done by Tate [2]–[3] for the correlation coefficient between one continuous random variable and one binary random variable. Tate, who utilized (1.2) to express the joint distribution between X and Y, assumed that the conditional probability distribution of X given Y was Gaussian:

$$Pr(x|Y=y) = \frac{1}{\sigma\sqrt{2\pi}}e^{-[(x-\mu_y)^2/2\sigma^2]}, \qquad x, \mu_y \in \Re;\;\; y=0,1;\;\; \sigma > 0, \tag{1.3}$$

while the marginal probability distribution of Y was assumed to follow a Bernoulli distribution:

$$Pr(Y=y) = p^y q^{1-y}, \qquad y=0,1;\;\; 0 \le p \le 1;\;\; q = 1-p;\;\; p+q=1. \tag{1.4}$$

As can be seen from (1.3), X follows a Gaussian distribution for each value of Y, with different means but a common variance. That is, the shape of the conditional distributions are assumed to be the same, but with a location shift. Based on (1.3) and (1.4), the joint probability distribution of X and Y, $f_{X,Y}(x,y)$, can be expressed as:

$$f_{X,Y}(x,y) = \frac{p}{\sigma\sqrt{2\pi}} e^{-[(x-\mu_1)^2/2\sigma^2]} + \frac{q}{\sigma\sqrt{2\pi}} e^{-[(x-\mu_0)^2/2\sigma^2]}, \tag{1.5}$$

which can be recognized as a two-component mixture of Gaussian distributions.

Olkin and Tate [4] later extended Tate's previous work to the multivariate setting. Let there be C continuous variables defined by $\mathbf{X} = (X_1, X_2, \ldots, X_C)^T$, and D discrete variables defined by $\mathbf{Y} = (Y_1, Y_2, \ldots, Y_D)^T$. In addition, suppose the dth discrete variable Y_D has s_d categories. Thus, there will be a total of $S = \prod_{d=1}^{D} s_d$ possible patterns of the discrete responses for $\mathbf{Y}$ (states). Utilizing the conditional Gaussian distribution and (1.5), the joint probability density of $\mathbf{X}$ and $\mathbf{Y}$ can be written as:

$$f_{\mathbf{X},\mathbf{Y}}(\mathbf{x},\mathbf{y}) = \sum_{s=1}^{S} p_s (2\pi)^{-C/2} |\boldsymbol{\Sigma}_s|^{-1/2} exp\left(-\frac{1}{2}(\mathbf{x}-\boldsymbol{\mu}_s)^T \boldsymbol{\Sigma}_s^{-1} (\mathbf{x}-\boldsymbol{\mu}_s)\right), \tag{1.6}$$

where given that $\mathbf{Y}$ falls in the sth state, then $\mathbf{X}$ is distributed according to the multivariate normal distribution, $\mathcal{N}_C(\boldsymbol{\mu}_s, \boldsymbol{\Sigma}_s)$, and the marginal probability that $\mathbf{Y}$ falls into state s is p_s with $\sum_{s=1}^{S} p_s = 1$. One simplifying assumption for this model would be to have a common covariance matrix for each discrete state. This would lead to the following expression for the joint probability distribution of $\mathbf{X}$ and $\mathbf{Y}$:

$$f_{\mathbf{X},\mathbf{Y}}(\mathbf{x},\mathbf{y}) = \sum_{s=1}^{S} p_s (2\pi)^{-C/2} |\boldsymbol{\Sigma}|^{-1/2} exp\left(-\frac{1}{2}(\mathbf{x}-\boldsymbol{\mu}_s)^T \boldsymbol{\Sigma}^{-1} (\mathbf{x}-\boldsymbol{\mu}_s)\right), \tag{1.7}$$

which is the model utilized by Olkin and Tate [4] when examining multivariate correlation models for continuous and discrete random variables. This particular model is often referred to as the general location model (GLOM) [4]–[7]. GLOM methodology assumes a homogeneous covariance matrix across all discrete states while allowing the means to vary. After its introduction by Olkin and Tate [4], GLOM-based methods have been used in several types of analysis of continuous and discrete data. For example, Afifi and Elashoff [8] used this model for hypothesis testing in the two-sample case, and Krzanowski [1],[9]–[13] authored several articles using the GLOM approach for discrimination and classification analysis applied to mixtures of continuous and discrete data. Additional articles using the GLOM methodology for classification of mixed data have been written by Chang and Afifi [14], and de Leon et al. [15] where the latter generalized classification for the GLOM model to a general data model (e.g., continuous, binary, ordinal). Bar-Hen and Daudin [16] generalized Mahalanobis distance to the mixture data case using GLOM-based approaches, while Morales et al. [17] generalized informational distance to the multiple data types using the GLOM approach. Lauritzen and Wermuth [18] used GLOM methodology for developing graphical models for the association between quantitative and qualitative variables. GLOM-based methods have also been applied to the development of additional likelihood ratio tests for multiple types of data [19]–[20].

1.1.1.2 Using $Pr(Y = y|x)f_X(x)$.

While the majority of the work in using conditional distributions to generate the joint probability density function has focused on GLOM-based methods (i.e., conditional Gaussian), some research has utilized the less commonly used representation, $Pr(Y = y|x)f_X(x)$. This method was first mentioned by Cox [21] for the multivariate case for both $\mathbf{X}$ and $\mathbf{Y}$. In this framework, the conditional distribution of $\mathbf{Y}|\mathbf{X}$ was assumed to follow a multivariate logistic distribution while the marginal distribution of $\mathbf{X}$ was assumed to be multivariate normal. This idea was developed further by Cox and Wermuth [22] by noting the connection between the conditional logistic method by Cox [21] and probit as well as latent variable models. However, these methods have not been pursued nearly as frequently as the GLOM-based methods. This may be due to the difficulty in working with multivariate binary data, or that probit or latent-type methods may not be as useful for nominal data if the nominal variables are not assumed to be the representation of underlying unobserved continuous random variables.

1.1.2 Copulas

Another way to derive the joint probability density function for random variables is by the use of copulas. Before proceeding further, it is important to properly define what a copula is. As stated by Nelsen [23], copulas are functions which "join or "couple" multivariate distribution functions to their one-dimensional marginal distribution functions". Stated another way, a copula itself is a multivariate distribution function whose inputs are the respective marginal cumulative probability distribution functions for the random variables of interest. Based on that description, one can see why copulas would be appealing for the derivation of the joint probability distribution for a set of random variables. A copula function, which includes a dependency parameter, only requires the specification of the respective marginal cumulative probability distribution functions. However, when not all of the random variables are continuous, there are special considerations that need to be addressed.

For a given random variable $V \in \Re$, its cumulative probability distribution function (CDF), $F_V(v)$, is defined as $Pr(V \leq v)$. The CDF of V has the following properties: 1. $F_V(v)$ is a non-decreasing fuction of v, 2. $lim_{v \to -\infty} F_V(v) = 0$, 3. $lim_{v \to \infty} F_V(v) = 1$, and 4. $F_V(v)$ is right-continuous: for every number v_0, $lim_{v \downarrow v_0} F_V(v) = F_V(v_0)$ [23]–[24]. In addition, from introductory probability theory, we know that if V is a continuous random variable, according to the probability integral transform, $F_V(v) \sim Uniform(0,1)$ [24]. However, the same is not true if V is a discrete random variable. This is due to the fact that the CDF of a discrete random variable is a step function. The essential theorem that allows the recovery of the joint probability distribution function via the copula is Sklar's Theorem which is stated below:

Sklar's Theorem. *Let H be a joint distribution function with margins F and G. Then there exists a copula C such that for all $x \in \Re$,*

$$H(x, y) = C(F(x), G(y)) \tag{1.8}$$

If F and G are continuous, then C is unique; otherwise, C is uniquely determined on Range $F \times$ Range G. Conversely, if C is a copula and F and G are distribution functions, then the function H defined by (1.8) is a joint distribution function with margins F and G [25].

As can be seen by Sklar's Theorem, when F and G are both continuous CDFs, then a given copula will uniquely determine the joint probability distribution of X and Y. However, when F and G are not both continuous CDFs, the copula, C, will only uniquely determine the joint probability distribution function of X and Y over Range $F \times$ Range G. Genest and Nešlehovà [26] provide a comprehensive treatment of the use of copulas for joint probability distribution function generation when both marginal CDFs are based on discrete count random variables. The authors state that while Sklar's Theorem can still be used when both variables are discrete, the joint distribution is not guaranteed to be unique. Song et al. [27] utilized a copula framework when determining the joint probability distribution for continuous, ordinal, and binary data by using a continuous latent variable for the ordinal variables and having a separate copula for each level of the single binary random variable. However, we are not aware of any previous research which has examined the properties of copulas and their use for generating joint probability distribution functions when the marginal CDFs are of different data types.

In spite of these issues, some authors have utilized copula methodology to determine the joint distribution of multiple types of data by utilizing a latent variable approach [27]–[30]. Specifically, discrete variables are assumed to be discretized versions of unobserved underlying continuous variables [31]. In this construct, as introduced earlier, unknown thresholding parameter(s) are used to define the observed discrete random variable(s) in terms of unobserved continuous measure(s). Let W^* be an unobserved continuous random variable. Assuming that W has an underlying continuity represented by W^*, then $Pr(W = 1) = Pr(W^* > \tau)$, where τ is an unknown thresholding parameter. Similarly, $Pr(W = 0) = Pr(W^* \leq \tau)$. While this may seem like a reasonable approach when the discrete variable is ordinal, it is less clear if this approach is reasonable when the discrete variable is not subject to an underlying continuous latent random variable (i.e.,nominal). This aspect has been noted by Wu et al. [30] who observed that "...while suitable for ordinal outcomes, the notion of continuous latent variables underlying nominal outcomes may not be appropriate".

1.1.3 Latent Variables

While the previous section demonstrated how latent variables have been used to generate the joint probability distribution function for random variables of different data types, a similar latent variable construct has been used in other ways to generate the joint distribution of different data types not via the copula method. Some of the first demonstrated work for this methodology is attributed to Cox [32]. A latent variable construct was utilized to estimate the correlation coefficient between a continuous random variable and a discrete random variable by assuming the discrete variable was the realization of an unobserved continuous random variable. By applying this mechanism, the joint distribution for the observed continuous variable and the continuous latent variable was assumed to be bivariate normal. Cox and Wermuth [22] also adopted this approach when examining response models for binary and continuous random variables. Bedrick et al. [33] and de Leon et al. [34] extended this approach to multiple latent variables for the estimation of the Mahalanobis distance for mixed continuous and discrete data. However, it is worth noting that, similar to copula models for different types of data, the latent variable approach may not be appropriate for nominal variables where the underlying continuous variable assumption may not be reasonable.

2.0 PROPOSAL

For this dissertation, we will examine mixtures of different types of data; specifically we will consider mixtures of continuous and discrete data. For the developments that follow, we will focus on the GLOM-based methodology for deriving the joint probability distribution of continuous and discrete random variables as the product of conditional and marginal probability distributions (as shown in section 1.1.1). As we noticed from (1.6), the general specification of the joint distribution is a finite mixture of multivariate Gaussian distributions with potentially different covariance matrices. One item to note from this specification of the PDF is that some confusion arises when this specification is erroneously interpreted as the sum of separate multivariate Gaussian distributions. Rather, using the notation of (1.6), the correct interpretation is that the random variable $\mathbf{x}$ is assumed to have been generated from one of the component "s" multivariate Gaussian distributions but it is unknown which one. That is, it is important to differentiate between a random variable with a PDF that is the sum of a group of component distributions (mixture distribution), and a random variable that is the sum of "s" random variables where the distribution can be found using established approaches for sums of independent random variables (e.g., convolution).

In a general sense, we can think of a random variable with a mixture distribution as one having been generated from at least two subpopulations. For example, in a genetic association study, we may think of a genetic variant (e.g., single-nucleotide polymorphism (SNP)) as having been derived from two or more subpopulations each having different allele frequencies. As we notice from (1.6), this form of the mixture distribution allows for different covariance matrices for each component distribution. However, under such situations, analysis if generally more complicated. Therefore, for this dissertation, we propose approximating a mixture distribution with a distribution from the same parametric family. For this work, we are also assuming that the mixture distribution has component distributions all from the same parametric family as well. For example, if we have a random variable with a distribution that is a mixture of multivariate Gaussians, the approximating distribution would be a single multivariate Gaussian distribution.

Several authors have undertaken a similar problem: the weighted sum of central chi-squared distributions (special case of the weighted sum of gamma distributions). Arguably, the most well-known approach for dealing with such a problem was introduced by Satterthwaite [35] - [36] and Welch [37]; equations are constructed by matching the corresponding mean(s) and variance(s) for the original distribution (weighted sum of central chi-squared random variables) and an approximate chi-squared distribution (with a degrees of freedom adjustment). While the method attributed to Satterthwaite as well as Welch have been been widely utilized, other authors have since suggested other approaches, with varying degrees of complexity. For example, Davis [38] evaluated the distribution of weighted sums of central chi-squared random variables using a differential equation approach. Solomon and Stephens [39] approximated the distribution of the weighted sum of central chi-squared random variables by first fitting a Pearson curve with the same first four moments as the weighted sum and also by fitting $Q_k = Aw^r$ where $w \sim \chi_p$ and where A, r, and p are determined by the first three moments of Q_k. Oman and Zacks [40] utilized negative binomial mixture distributions. Mathai [41] evaluated the distribution of the weighted sum of gamma random variables using incomplete gamma functions. In addition, Moschopoulos and Canada [42] inverted the MGF of the weighted sum of central chi-squared random variables to obtain the distribution of the weighted sum as an infinite series of incomplete gamma integrals. Lindsay et al. [43] utilized gamma mixture distributions to approximate the distribution of the weighted sum of central chi-squared random variables by matching moments. More recently, Di Salvo [44] expressed the exact distribution of the weighted sum of gamma random variables as the product between a gamma density and a confluent hypergeometric function. These examples illustrate that the distributional form of the weighted sum of gamma random variables is somewhat complex and can be subject to computational challenges.

While the preceding methods were primarily proposed for deriving (or approximating) the distribution of the sum of central chi-squared random variables, approaches for approximating the mixture distribution for a set of component distributions have received less attention. Taking a mixture of gamma distributions as an example, we know that complications can arise when the scale parameters for each component distribution are not identical. In these situations, an approximation method originally developed for estimating the distribution of a sum of central chi-squared random variables may have some utility in estimating a corresponding mixture distribution. Therefore, we propose using a method similar to that developed by Satterthwaite and Welch for approximating a mixture of gamma distributions. Simulations will be utilized to assess the adequacy of the applicable approximation.

The remainder of this dissertation is organized as follows. Chapter 3 is devoted to determining the distribution of the sample variance for a Gaussian finite mixture distribution. Section 3.1 focuses on the univariate case, and section 3.2 focuses on the multivariate case. Theoretical developments supporting these efforts are shown in Appendices A-C. In Appendix A we utilize contour integration from the field of complex analysis to obtain the PDF from the MGF in the univariate case. In Appendices B-C we perform similar developments for the multivariate case. In Appendix D we determine the marginal distribution when the joint distribution is that of a mixture of multivariate Gaussian distributions. In sections 3.2 and 3.2 we also present simulations for the univariate and multivariate cases, respectivley. Section 3.2 also contains a discussion of considerations for simulating from a Wishart distribution. In Section 4 we suggest some future directions for this work. Finally, it is worth noting that for this dissertation we will be focusing on situations where the parameters for the various component distributions from the mixture are reasonably close to each other.

3.0 DISTRIBUTION OF THE SAMPLE VARIANCE - GAUSSIAN FINITE MIXTURE DISTRIBUTION

3.1 UNIVARIATE CASE

A k-component finite mixture distribution has the following PDF:

$$f_k(x_\ell) = \sum_{j=1}^{k} w_j f_j(x_\ell|\theta_j), \tag{3.1}$$

where x_ℓ is a random variable, $f_j(x_\ell|\theta_j)$ may be a continuous or discrete distribution, θ_j represents the parameters of the j^{th} component distribution, w_j represents the weight for the j^{th} component distribution, and k is finite. We also note that a random variable distributed as in (3.1) is assumed to have been generated in a heterogeneous manner. That is, some data points were generated from each of the k component distributions, but we do not know which point was generated from which distribution. Further, the w_js satisfy:

$$\sum_{j=1}^{k} w_j = 1, w_j \geq 0 \tag{3.2}$$

Letting $f_j(x_\ell|\theta_j)$ in (3.1) be represented by a Gaussian PDF with $\theta_j = \{\mu_j, \sigma_j^2\}$ we have:

$$f_k(x_\ell) = \sum_{j=1}^{k} w_j \phi(x_\ell|\mu_j, \sigma_j^2), \tag{3.3}$$

where $\phi(x_\ell|\mu_j, \sigma_j^2)$ represents the j^{th} Gaussian distribution with its mean $= \mu_j$ and variance $= \sigma_j^2$. We note that the same constraint in (3.2) applies, and k remains finite.

Given the mixture parameters in (3.3), $\{w_j, \mu_j, \sigma_j^2\}, j = 1, \ldots, k$, the expected value of x_ℓ can be written as:

$$E\left(x_\ell | w_j, \mu_j, \sigma_j^2\right) = \sum_{j=1}^{k} w_j E\left(x_\ell | \mu_j, \sigma_j^2\right) \tag{3.4}$$

$$= \sum_{j=1}^{k} w_j \mu_j \tag{3.5}$$

$$= \mu_{mix} \tag{3.6}$$

Similarly, we can compute the variance of x_ℓ by first calculating its second moment, given $\{w_j, \mu_j, \sigma_j^2\}, j = 1, \ldots, k$:

$$E\left(x_\ell^2 | w_j, \mu_j, \sigma_j^2\right) = \sum_{j=1}^{k} w_j E\left(x_\ell^2 | \mu_j, \sigma_j^2\right) \tag{3.7}$$

$$= \sum_{j=1}^{k} w_j \left(\sigma_j^2 + \mu_j^2\right), \tag{3.8}$$

which leads to the variance of x_ℓ as:

$$Var\left(x_\ell\right) = E\left(x_\ell^2 | w_j, \mu_j, \sigma_j^2\right) - \left[E\left(x_\ell | w_j, \mu_j, \sigma_j^2\right)\right]^2 \tag{3.9}$$

$$= \sum_{j=1}^{k} w_j \left(\sigma_j^2 + \mu_j^2\right) - \mu_{mix}^2 \tag{3.10}$$

$$= \sigma_{mix}^2 \tag{3.11}$$

If we consider the class of 0-mean k-component finite Gaussian mixture distributions, (3.6) and (3.11) become:

$$\mu_{mix} = 0, \sigma_{mix}^2 = \sum_{j=1}^{k} w_j \sigma_j^2 \tag{3.12}$$

Given a random sample $\{e_1, \ldots, e_n\}$ from such a distribution, an unbiased sample variance estimator would be:

$$\hat{\sigma}_{mix}^2 = \left(\frac{1}{n-1}\right) \sum_{\ell=1}^{n} e_\ell^2 \tag{3.13}$$

Next, let us define a random variable

$$\varepsilon_n = \sum_{\ell=1}^{n} y_\ell^2, \tag{3.14}$$

where y_i is a random variable from a 0-mean k-component finite Gaussian mixture distribution as defined in (3.12). Define the scaled random variable

$$\varepsilon = \left(\frac{1}{n-1}\right) \varepsilon_n, \tag{3.15}$$

which produces the unbiased variance estimator in (3.13). To determine the density for (3.15), we can use the property that a PDF is uniquely determined by its moment generating function (MGF), when it exists. The MGF of the random variable ε_n is defined as

$$\begin{aligned} M_{\varepsilon_n}(t) &= E\left(e^{t\varepsilon_n}\right) = E\left(e^{t\sum_{\ell=1}^{n} y_\ell^2}\right) && (3.16)\\ &= \prod_{\ell=1}^{n} E\left(e^{ty_\ell^2}\right) && (3.17)\\ &= \left(E\left(e^{ty_\ell^2}\right)\right)^n && (3.18)\\ &= \left(M_{y_\ell^2}(t)\right)^n, && (3.19) \end{aligned}$$

where:

$$M_{y_\ell^2}(t) = \int_{-\infty}^{\infty} e^{ty_\ell^2} \sum_{j=1}^{k} w_j \phi\left(y_\ell | 0, \sigma_j^2\right) dy_\ell \tag{3.20}$$

Because k is finite, the order of the summation and integration can be reversed:

$$\begin{aligned} M_{y_\ell^2}(t) &= \sum_{j=1}^{k} w_j \int_{-\infty}^{\infty} e^{ty_\ell^2} \left(\frac{1}{\sqrt{2\pi}\sigma_j} e^{-\frac{y_\ell^2}{2\sigma_j^2}}\right) dy_\ell && (3.21)\\ &= \sum_{j=1}^{k} w_j \int_{-\infty}^{\infty} \left(\frac{1}{\sqrt{2\pi}\sigma_j}\right) \exp\left(ty_\ell^2 - \frac{y_\ell^2}{2\sigma_j^2}\right) dy_\ell && (3.22)\\ &= \sum_{j=1}^{k} w_j \int_{-\infty}^{\infty} \left(\frac{1}{\sqrt{2\pi}\sigma_j}\right) \exp\left(\frac{2\sigma_j^2 t y_\ell^2 - y_\ell^2}{2\sigma_j^2}\right) dy_\ell && (3.23)\\ &= \sum_{j=1}^{k} w_j \int_{-\infty}^{\infty} \left(\frac{1}{\sqrt{2\pi}\sigma_j}\right) \exp\left(-\frac{y_\ell^2}{2\sigma_j^2}\left(1 - 2\sigma_j^2 t\right)\right) dy_{i\ell} && (3.24) \end{aligned}$$

Using the substitution method:

$$v = y_\ell \sqrt{1 - 2\sigma_j^2 t} \tag{3.25}$$

$$dv = \sqrt{1 - 2\sigma_j^2 t}\; dy_\ell \tag{3.26}$$

$$dy_\ell = \left(\frac{1}{\sqrt{1 - 2\sigma_j^2 t}} \right) dv \tag{3.27}$$

Substituting into (3.24):

$$\begin{aligned}
M_{y_\ell^2}(t) &= \sum_{j=1}^{k} w_j \int_{-\infty}^{\infty} \left(\frac{1}{\sqrt{2\pi}\sigma_j} \right) \exp\left(-\frac{w^2}{2\sigma_j^2} \right) \left(\frac{1}{\sqrt{1 - 2\sigma_j^2 t}} \right) dv && (3.28)\\
&= \sum_{j=1}^{k} w_j \left(1 - 2\sigma_j^2 t\right)^{-\frac{1}{2}} \underbrace{\int_{-\infty}^{\infty} \left(\frac{1}{\sqrt{2\pi}\sigma_j} \right) \exp\left(-\frac{w^2}{2\sigma_j^2} \right) dv}_{Integrates\, to\, 1} && (3.29)\\
&= \sum_{j=1}^{k} w_j \left(1 - 2\sigma_j^2 t\right)^{-\frac{1}{2}}, t < \frac{1}{2\sigma_j^2}, && (3.30)
\end{aligned}$$

which is a weighted sum of MGFs for a *Gamma* $\left(\frac{1}{2}, 2\sigma_j^2\right)$ distribution. Combining (3.30) with (3.16):

$$\begin{aligned}
M_{\varepsilon_n}(t) &= \left(\sum_{j=1}^{k} w_j \left(1 - 2\sigma_j^2 t\right)^{-\frac{1}{2}} \right)^n && (3.31)\\
&= \left(\sum_{j=1}^{k} w_j M_j(t) \right)^n && (3.32)
\end{aligned}$$

As is shown in Appendix A, $\sum_{j=1}^{k} w_j M_j(t)$ corresponds to a PDF which is a k-component mixture of *Gamma* $\left(\frac{1}{2}, 2\sigma_j^2\right)$ PDFs. We note that from (3.30) the scale parameters, $2\sigma_j^2$, are not necessarily identical. If all the scale parameters were the same,

$\sigma_1^2 = \sigma_2^2 = \cdots = \sigma_k^2 = \sigma^2$, (3.32) becomes:

$$M_{\varepsilon_n}(t) = \left(\sum_{j=1}^{k} w_j \left(1 - 2\sigma_j^2 t\right)^{-\frac{1}{2}} \right)^n \tag{3.33}$$

$$= \left(\sum_{j=1}^{k} w_j \left(1 - 2\sigma^2 t\right)^{-\frac{1}{2}} \right)^n \tag{3.34}$$

$$= \left(\left(1 - 2\sigma^2 t\right)^{-\frac{1}{2}} \sum_{j=1}^{k} w_j \right)^n \tag{3.35}$$

$$= \left(\left(1 - 2\sigma^2 t\right)^{-\frac{1}{2}} \right)^n \tag{3.36}$$

$$= \left(1 - 2\sigma^2 t\right)^{-\frac{n}{2}}, \tag{3.37}$$

which is the MGF corresponding to a $Gamma\left(\frac{n}{2}, 2\sigma^2\right)$ PDF. Thus, under this scenario, $\varepsilon_n \sim Gamma\left(\frac{n}{2}, 2\sigma^2\right)$. However, when the $\sigma_j^2 s$ from (3.30) are not all equal, the PDF corresponding to the MGF is decidedly more complicated. From Appendix A, $M_{y_\ell^2}(t) = \sum_{j=1}^{k} w_j \left(1 - 2\sigma_j^2 t\right)^{-\frac{1}{2}}$ corresponds to the k-component mixture of Gamma distributions:

$$f\left(y_i^2\right) = \sum_{j=1}^{k} w_j \; Gamma\left(\frac{1}{2}, 2\sigma_j^2\right) \tag{3.38}$$

First, we will calculate the first and second moments for y_ℓ^2 by utilizing the MGF shown in (3.30).

$$\frac{\partial M_{y_\ell^2}(t)}{\partial t} = \sum_{j=1}^{k} w_j \left(-\frac{1}{2}\right) \left(1 - 2\sigma_j^2 t\right)^{-\frac{3}{2}} \left(-2\sigma_j^2\right) \tag{3.39}$$

$$\mathbf{E}\left(y_\ell^2\right) = \left.\frac{\partial M_{y_\ell^2}(t)}{\partial t}\right|_{t=0} = \left.\sum_{j=1}^{k} w_j \sigma_j^2 \left(1 - 2\sigma_j^2 t\right)^{-\frac{3}{2}}\right|_{t=0} \tag{3.40}$$

$$= \sum_{j=1}^{k} w_j \sigma_j^2 \tag{3.41}$$

Similarly, we have for the second moment of y_ℓ^2 we have:

$$\frac{\partial^2 M_{y_\ell^2}(t)}{\partial t^2} = \sum_{j=1}^{k} w_j \sigma_j^2 \left(-\frac{3}{2}\right) \left(1 - 2\sigma_j^2 t\right)^{-\frac{5}{2}} \left(-2\sigma_j^2\right) \tag{3.42}$$

$$= 3 \sum_{j=1}^{k} w_j \sigma_j^4 \left(1 - 2\sigma_j^2 t\right)^{-\frac{5}{2}} \tag{3.43}$$

$$\mathbf{E}\left(y_\ell^2\right)^2 = \left.\frac{\partial^2 M_{y_\ell^2}(t)}{\partial t^2}\right|_{t=0} = \left.\left(3\sum_{j=1}^{k} w_j\sigma_j^4\left(1-2\sigma_j^2 t\right)^{-\frac{5}{2}}\right)\right|_{t=0} \tag{3.44}$$

$$= 3\sum_{j=1}^{k} w_j\sigma_j^4 \tag{3.45}$$

Finally, using the results from (3.41) and (3.45), we have:

$$\mathbf{Var}\left(y_\ell^2\right) = \mathbf{E}\left(y_\ell^2\right)^2 - \left[\mathbf{E}\left(y_\ell^2\right)\right]^2 \tag{3.46}$$

$$= 3\sum_{j=1}^{k} w_j\sigma_j^4 - \left[\sum_{j=1}^{k} w_j\sigma_j^2\right]^2 \tag{3.47}$$

Let us assume we are going to approximate the distribution of y_ℓ^2 using a single gamma distribution (instead of a mixture of gamma densities). We shall approximate the random variable y_ℓ^2 by the random variable $\tilde{y}$ where $\tilde{y} \sim Gamma(\alpha, \beta)$. By (3.37), we can see that $\tilde{y}$ has the following MGF:

$$M_{\tilde{y}}(t) = (1-\beta t)^{-\alpha} \tag{3.48}$$

Similarly, we will calculate the first and second moments for $\tilde{y}$ using the MGF:

$$\mathbf{E}(\tilde{y}) = \left.\frac{\partial M_{\tilde{y}}(t)}{\partial t}\right|_{t=0} \tag{3.49}$$

$$= \left.\left\{(-\alpha)(1-\beta t)^{-\alpha-1}(-\beta)\right\}\right|_{t=0} \tag{3.50}$$

$$= \alpha\beta \tag{3.51}$$

$$\mathbf{E}\left(\tilde{y}^2\right) = \left.\frac{\partial^2 M_{\tilde{y}}(t)}{\partial t^2}\right|_{t=0} \tag{3.52}$$

$$= \left.\frac{\partial\left[\alpha\beta(1-\beta t)^{-\alpha-1}\right]}{\partial t}\right|_{t=0} \tag{3.53}$$

$$= \left.\left\{(\alpha\beta)(-\alpha-1)(1-\beta t)^{-\alpha-2}(-\beta)\right\}\right|_{t=0} \tag{3.54}$$

$$= (\alpha\beta)(-\alpha-1)(-\beta) \tag{3.55}$$

$$= \alpha\beta^2(\alpha+1) \tag{3.56}$$

Finally, using the results from (3.51) and (3.56) we have:

$$\begin{aligned}
\mathbf{Var}\left(\tilde{y}\right) &= \mathbf{E}\left(\tilde{y}^2\right) - \left[\mathbf{E}\left(\tilde{y}\right)\right]^2 && (3.57)\\
&= \alpha\beta^2\left(\alpha+1\right) - \left[\alpha\beta\right]^2 && (3.58)\\
&= \alpha\beta^2\left(\alpha+1\right) - \alpha^2\beta^2 && (3.59)\\
&= \alpha\beta^2[(\alpha+1) - \alpha] && (3.60)\\
&= \alpha\beta^2 && (3.61)
\end{aligned}$$

Now, similar to the approach used by Satterthwaite as well as Welch, let us equate the corresponding means and variances of y_i^2 and $\tilde{y}$:

$$\alpha\beta = \sum_{j=1}^{k} w_j\sigma_j^2 \tag{3.62}$$

$$\alpha\beta^2 = 3\sum_{j=1}^{k} w_j\sigma_j^4 - \left[\sum_{j=1}^{k} w_j\sigma_j^2\right]^2 \tag{3.63}$$

Based on equation (3.62), we note that

$$\alpha = \frac{\sum_{j=1}^{k} w_j\sigma_j^2}{\beta} \tag{3.64}$$

Substituting this into (3.63), we have:

$$\left(\frac{\sum_{j=1}^{k} w_j\sigma_j^2}{\beta}\right)\beta^2 = 3\sum_{j=1}^{k} w_j\sigma_j^4 - \left[\sum_{j=1}^{k} w_j\sigma_j^2\right]^2 \tag{3.65}$$

$$\beta = \frac{3\sum_{j=1}^{k} w_j\sigma_j^4 - [\sum_{j=1}^{k} w_j\sigma_j^2]^2}{\sum_{j=1}^{k} w_j\sigma_j^2} \tag{3.66}$$

Finally, substituting (3.66) into (3.64), we have:

$$\alpha = \frac{\left(\sum_{j=1}^{k} w_j\sigma_j^2\right)^2}{3\sum_{j=1}^{k} w_j\sigma_j^4 - [\sum_{j=1}^{k} w_j\sigma_j^2]^2} \tag{3.67}$$

Because $\mathbf{Var}\left(y_i^2\right) = 3\sum_{j=1}^{k} w_j\sigma_j^4 - [\sum_{j=1}^{k} w_j\sigma_j^2]^2$, and is greater than zero by definition, we also know that the expressions for α and β in (3.67) and (3.66), respectively, are both greater than zero (properties of the gamma distribution). Next, simulations were performed to evaluate the adequacy of the approximation method.

3.1.1 Data Simulations - Univariate Case

As previously mentioned, data simulations were performed to assess the adequacy of the approximation method shown in (3.62) - (3.67). Initial simulations were performed with a 2-component mixture of gamma distributions. For the 2-component mixture distribution parametrized as in (3.38), 3 different scenarios were evaluated:

- $w_1 = 0.2, w_2 = 0.8, \alpha = 0.5, \beta_1 = 2\sigma_1^2 = 1, \beta_2 = 2\sigma_2^2 = 1.1$
- $w_1 = 0.5, w_2 = 0.5, \alpha = 0.5, \beta_1 = 2\sigma_1^2 = 1, \beta_2 = 2\sigma_2^2 = 1.1$
- $w_1 = 0.8, w_2 = 0.2, \alpha = 0.5, \beta_1 = 2\sigma_1^2 = 1, \beta_2 = 2\sigma_2^2 = 1.1$

where w_1 and w_2 represent the mixture weights from the first and second component gamma distributions, respectively. Further, α represents the common location parameter from the component gamma distributions, and β_1 and β_2 represent the scale parameters from the first and second component gamma distributions, respectively. The generation of data from a 2-component gamma mixture distribution was a 2-step process. First, a random variate (u) was generated from a $Uniform(0, 1)$ distribution. If $u < w_1$, then a random variate was generated from a $Gamma(\alpha = 0.5, \beta_1 = 1)$ distribution. Otherwise, a random variate was generated from a $Gamma(\alpha = 0.5, \beta_2 = 1.1)$ distribution. Similarly for the approximation method, a random variate was generated from a $Gamma(\alpha, \beta)$ distribution using the expressions in (3.66) and (3.67). For each simulation scenario, 1,000 replicates were generated each with a sample size of 100. In addition to density plots comparing the mixture distribution with the approximation method, the following statistics were utilized

$$\xi_{k_f} = \sum_{g=1}^{100} \left(h_{fg} - \tilde{h}_{fg} \right) \tag{3.68}$$

$$\xi_k = \sum_{f=1}^{1000} \xi_{k_f} \tag{3.69}$$

$$s^2_{\xi_k} = \frac{\sum\limits_{f=1}^{1000} \left(\xi_{k_f} - \bar{\xi}_k \right)^2}{f - 1} \tag{3.70}$$

$$\bar{\xi}_k = \frac{\sum\limits_{f=1}^{1000} \xi_{k_f}}{f} \tag{3.71}$$

where k = the number of mixture components, f = the number of replicates, and g = the sample size per replicate. In addition, in (3.68), h_{fg} = a random deviate from a mixture of gamma distributions, and $\tilde{h}_{fg}$ = a random deviate from the approximating gamma distribution. The density plots assessing the adequacy of the approximation method are shown in Figure 1. Based on a review of Figure 1 and Table 1 we note that the best fitting approximating distributions are those applied to the situation when $w_1 = 0.5, 0.8$. In addition, the average rates of error are summarized in Table 1.

3.2 MULTIVARIATE CASE

We shall now extend the work of section 3.1 to the multivariate case. However, before proceeding to the multivariate developments, we will define some notation as well as some concepts for multivariate data.

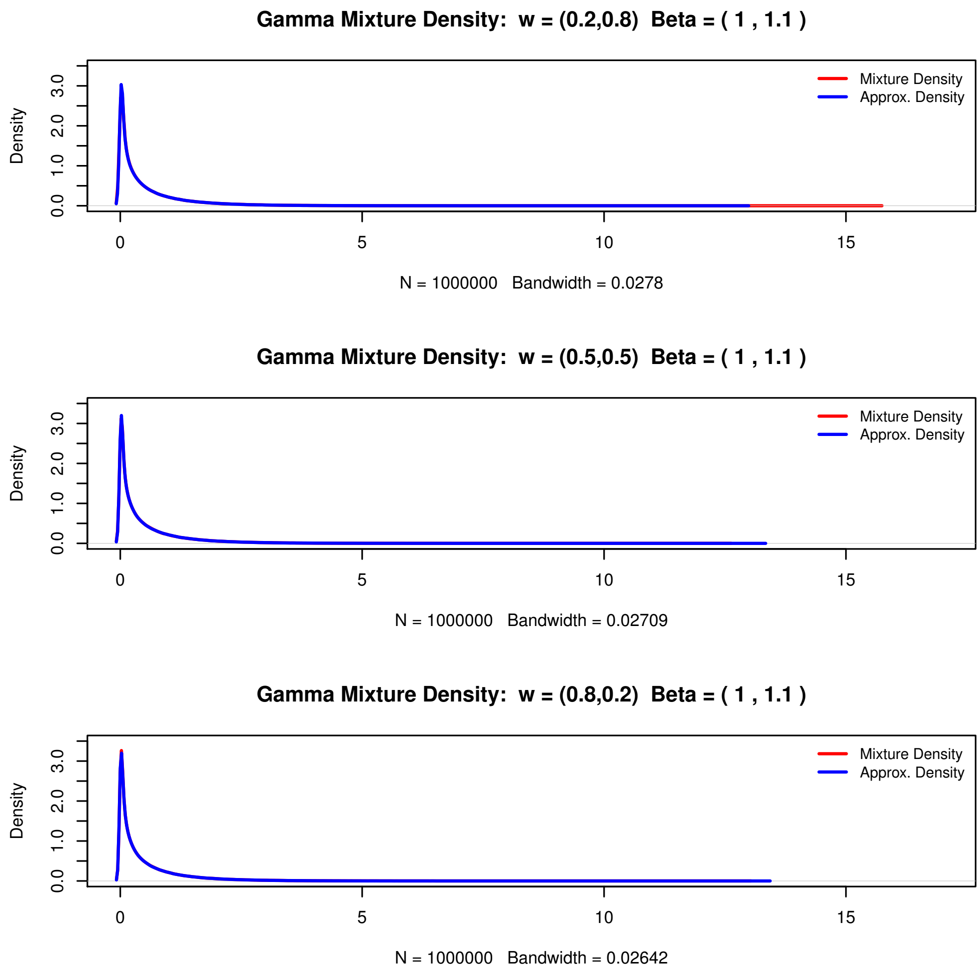

Figure 1: Simulation - 2-component Mixture of Gamma Distributions

Table 1: Comparison of Average Squared Error Between Mixture Distribution and Approximating Distribution

Scenario	Average Squared Error
$w_1 = 0.2,\ w_2 = 0.8$	119.45
$w_1 = 0.5,\ w_2 = 0.5$	107.86
$w_1 = 0.8,\ w_2 = 0.2$	101.83

First, let us define the random vector $\mathbf{x}_\ell$ and the random matrix $\mathbf{X}_{n\times p}$ as follows:

$$\mathbf{x}_\ell = \begin{pmatrix} x_{\ell 1} \\ x_{\ell 2} \\ \vdots \\ x_{\ell p} \end{pmatrix}_{p\times 1} , \quad \ell = 1, \ldots, n \tag{3.72}$$

$$\mathbf{X} = \begin{pmatrix} \mathbf{x}_1^T \\ \mathbf{x}_2^T \\ \vdots \\ \mathbf{x}_n^T \end{pmatrix}_{n\times p} \tag{3.73}$$

It is worth noting from (3.73) that the rows of $\mathbf{X}$ constitute a random sample (independent and identically distributed), but the columns of $\mathbf{X}$ do not have this property. For the random matrix $\mathbf{X}$, the rows represent observations, while the columns represent features (e.g., variables). Stated another way, each element of matrix $\mathbf{X}$, $(x)_{\ell r}$, represents the value for the r^{th} variable on the ℓ^{th} observation. Similar to the random vector $\mathbf{x}_\ell$, we can also define the vector of sample means as follows:

$$\overline{\mathbf{x}} = \begin{pmatrix} \overline{x}_1 \\ \overline{x}_2 \\ \vdots \\ \overline{x}_p \end{pmatrix}_{p\times 1} , \tag{3.74}$$

where $\overline{x}_r$ is the sample mean for the r^{th} variable and is defined as $\overline{x}_r = \dfrac{1}{n}\sum_{\ell=1}^{n} x_{\ell r}$. The sample covariance between the r^{th} and v^{th} variables, s_{rv}, can be expressed as:

$$s_{rv} = \frac{1}{n}\sum_{\ell=1}^{n}(x_{\ell r} - \overline{x}_r)(x_{\ell v} - \overline{x}_v) \tag{3.75}$$

Using (3.75), we can express the sample variance of the r^{th} variable, s_{rr} as:

$$s_{rr} = \frac{1}{n}\sum_{\ell=1}^{n}(x_{\ell r} - \overline{x}_r)^2 \tag{3.76}$$

We can also denote the sample covariance matrix, $\mathbf{S}$, by its matrix elements using (3.75) - (3.76).

$$\mathbf{S}_{p\times p} = (s)_{rv}, \tag{3.77}$$

where $(s)_{rv}$ is the matrix element in the r^{th} row and the v^{th} column of the matrix $\mathbf{S}$. Another useful random quantity for multivariate analysis is the matrix of the sum of squares and cross-products. The random matrix, $\mathbf{A}_\ell$, can be defined in terms of the random vector $\mathbf{x}_\ell$ from (3.72) as:

$$\mathbf{A}_\ell = \mathbf{x}_\ell \mathbf{x}_\ell^T = \begin{pmatrix} x_{\ell 1} \\ x_{\ell 2} \\ \vdots \\ x_{\ell p} \end{pmatrix}_{p\times 1} \begin{pmatrix} x_{\ell 1} & x_{\ell 2} & \dots & x_{\ell p} \end{pmatrix}_{1\times p}$$

$$= \begin{pmatrix} x_{\ell 1}^2 & x_{\ell 1}x_{\ell 2} & \dots & x_{\ell 1}x_{\ell p} \\ x_{\ell 2}x_{\ell 1} & x_{\ell 2}^2 & \dots & x_{\ell 2}x_{\ell p} \\ \vdots & \vdots & \ddots & \vdots \\ x_{\ell p}x_{\ell 1} & x_{\ell p}x_{\ell 2} & \dots & x_{\ell p}^2 \end{pmatrix}_{p\times p} \tag{3.78}$$

Summing $\mathbf{A}_\ell$ across all values of ℓ gives us the matrix of the sum of squares and cross-products, $\mathbf{A}$, as follows:

$$\begin{aligned} \mathbf{A}_{p\times p} &= \sum_{\ell=1}^{n} \mathbf{A}_\ell \\ &= \begin{pmatrix} \sum\limits_{\ell=1}^{n} x_{\ell 1}^2 & \sum\limits_{\ell=1}^{n} x_{\ell 1}x_{\ell 2} & \dots & \sum\limits_{\ell=1}^{n} x_{\ell 1}x_{\ell p} \\ \sum\limits_{\ell=1}^{n} x_{\ell 2}x_{\ell 1} & \sum\limits_{\ell=1}^{n} x_{\ell 2}^2 & \dots & \sum\limits_{\ell=1}^{n} x_{\ell 2}x_{\ell p} \\ \vdots & \vdots & \ddots & \vdots \\ \sum\limits_{\ell=1}^{n} x_{\ell p}x_{\ell 1} & \sum\limits_{\ell=1}^{n} x_{\ell p}x_{\ell 2} & \dots & \sum\limits_{\ell=1}^{n} x_{\ell p}^2 \end{pmatrix} \end{aligned} \tag{3.79}$$

We also note that when $\overline{\mathbf{x}}$ as defined in (3.74) is equal to $\mathbf{0}$, then $n\mathbf{S} = \mathbf{A}$. If we assume that $\mathbf{x}_\ell$ from (3.72) is distributed as $\mathcal{N}_p(\mathbf{0}, \mathbf{\Sigma})$, then it can be shown that $\mathbf{A}$ as shown in (3.79) is known to follow a $Wishart(f, \mathbf{\Sigma})$ distribution [45]. For the Wishart distribution, f is the degrees of freedom parameter and $\mathbf{\Sigma}$ refers to the scale parameter.

Reviewing the developments for the univariate case in section 3.1, we note that the MGF for the sum of squares of a random variable generated from a k-component mixture of Gaussian distributions is as shown in (3.30). As demonstrated in Appendix A, this MGF corresponds to a mixture of $Gamma\left(\frac{1}{2}, 2\sigma_j^2\right)$ distributions.

Now we will extend the calculation of the MGF to the multivariate case by adapting the work of Anderson (2003) for a single multivariate Gaussian distribution to that of a k-component mixture of Gaussian distributions. The detailed calculations are shown in Appendix B. We note that the MGF in (B.16) appears to be the k-component mixture of Wishart MGFs (each with a different scale matrix). Similar to the result demonstrated in Appendix A, we might surmise that the MGF in (B.16) corresponds to a mixture of Wisharts PDF. To verify this conjecture, we will first assume that the matrix $\mathbf{A}^*$ has a mixture of Wisharts distribution. Due to the one-to-one correspondence between a distribution and its MGF (if it exists), we can then determine this distribution's MGF, and see if it is equivalent to that shown in (B.16). Based on the mixture of Wisharts distributional assumption, we will also assume as true that $\mathbf{A}^*$ has the following probability distribution:

$$f\left(\mathbf{A}^*\right) = \sum_{j=1}^{k} w_j \left\{2^{(n_j p/2)} \Gamma_p\left(\frac{n_j}{2}\right) \det\left(\mathbf{\Sigma}_j\right)^{n_j/2}\right\}^{-1} \det\left(\mathbf{A}^*\right)^{(n_j-p-1)/2} \operatorname{etr}\left(-\frac{1}{2}\mathbf{\Sigma}_j{}^{-1}\mathbf{A}^*\right), \tag{3.80}$$

where $\mathbf{A}^*$ is a random symmetric matrix that is positive definite $(\mathbf{A}^* > \mathbf{0})$, $n_j \geq p-1$, and $\operatorname{etr}(\cdot) = \exp(\operatorname{tr}(\cdot))$. Also in (3.80), $\Gamma_p(\cdot)$ is the multivariate gamma function and is defined as follows by Gupta and Nagar [46]:

Definition 1. *Multivariate gamma function.* The multivariate gamma function denoted by $\Gamma_p(b)$ is defined as

$$\Gamma_p(b) = \int_{\mathbf{B}>0} \operatorname{etr}(-\mathbf{B}) \det(\mathbf{B})^{b-\frac{1}{2}(p+1)} \, d\mathbf{B}, \tag{3.81}$$

where $Re(b) > \frac{1}{2}(p-1)$, and the integral is over the space of $p \times p$ symmetric positive definite matrices.

Based on the developments shown in Appendix C, we note that (B.16) and (C.47) are identical. Because the relationship between a MGF and PDF is 1:1, we can conclude that the MGF shown in (B.16) corresponds to the mixture of Wisharts PDF. We can also summarize the main points of this PDF as follows:

$$f\left(\mathbf{A}^*\right) = \sum_{j=1}^{k} w_j \text{Wishart}\left(f_j, \mathbf{\Sigma}_j\right), \tag{3.82}$$

where:

$\mathbf{A}^* =$ the matrix of the sum of squares and the sum of cross-products

$\mathbf{\Sigma}_j =$ covariance matrix from the j^{th} component Wishart distribution

$f_j =$ the degrees of freedom for the j^{th} component Wishart distribution

Similar to the development in section 3.1 for the univariate case, we wish to approximate the mixture of Wishart distributions in (3.80) with a single Wishart distribution. Let us assume that the random matrix $\mathbf{A} \sim Wishart\left(f, \mathbf{\Sigma}\right)$. As in the univariate case, the multivariate extension will utilization the matching of first and second central moments. Therefore, we will first develop expressions for the first and second moments of the random matrix $\mathbf{A}$. Further, let us refer to random matrix $\mathbf{A}$ by its individual matrix elements: $\mathbf{A} = (a)_{rc}$, which indicates the matrix element in the r^{th} row and c^{th} column of $\mathbf{A}$. Using the individual matrix elements, we can define the expected value of the random matrix $\mathbf{A}$ as:

$$\mathbf{E}(\mathbf{A}) = \mathbf{E}\left(a\right)_{rc} \ \forall\ r, c \tag{3.83}$$

Let us first calculate the expected values of the elements on the main diagonal of the matrix $\mathbf{A}$ starting with $(a)_{11}$. For these developments, we will assume that $\mathbf{x}_i \sim \mathcal{N}_p\left(\mathbf{0}, \mathbf{\Sigma}\right)$.

$$\begin{aligned}\mathbf{E}\left[(a)_{11}\right] = \mathbf{E}\left(\sum_{\ell=1}^{n} x_{\ell 1}^2\right) &= \sum_{\ell=1}^{n} \mathbf{E}\left(x_{\ell 1}^2\right) \\ &= \sum_{\ell=1}^{n} \left[\mathbf{Var}\left(x_{\ell 1}\right) + \left[\mathbf{E}\left(x_{\ell 1}\right)\right]^2\right] \\ &= \sum_{\ell=1}^{n} \mathbf{Var}\left(x_{\ell 1}\right) \end{aligned} \tag{3.84}$$

The last equality in (3.84) is based on the fact that $\mathbf{x}_\ell$ follows a multivariate normal distribution as specified above, and, therefore, $x_{\ell 1} \sim \mathcal{N}\left(0, \sigma_{11}^2\right)$ where $\sigma_{11}^2 = (\sigma)_{11}$ (matrix element in the 1^{st} row and 1^{st} column of $\mathbf{\Sigma}$). Therefore, based on (3.84) and the fact that the $\mathbf{x}_\ell$s are independent and identically distributed, we have the following:

$$\begin{aligned}
\mathbf{E}\left[(a)_{11}\right] &= n\sigma_{11}^2 \\
\mathbf{E}\left[(a)_{22}\right] &= n\sigma_{22}^2 \\
\vdots &= \vdots \\
\mathbf{E}\left[(a)_{pp}\right] &= n\sigma_{pp}^2
\end{aligned} \tag{3.85}$$

Now, let us calculate the expectations of the off-diagonal elements of the random matrix $\mathbf{A}$. We will first start with the expectation of $(a)_{12}$.

$$\begin{aligned}
\mathbf{E}\left[(a)_{12}\right] = \mathbf{E}\left(\sum_{\ell=1}^{n} x_{\ell 1} x_{\ell 2}\right) &= \sum_{\ell=1}^{n} \mathbf{E}\left(x_{\ell 1} x_{\ell 2}\right) \\
&= \sum_{\ell=1}^{n} \left[\mathbf{Cov}\left(x_{\ell 1}, x_{\ell 2}\right) + \mathbf{E}\left(x_{\ell 1}\right)\mathbf{E}\left(x_{\ell 2}\right)\right] \\
&= \sum_{\ell=1}^{n} \mathbf{Cov}\left(x_{\ell 1}, x_{\ell 2}\right)
\end{aligned} \tag{3.86}$$

The last equality in (3.86) follows once again from the fact that $\mathbf{x}_\ell$ follows a multivariate normal distribution as specified above, and, as a result, $\mathbf{E}\left(x_{\ell 1}\right) = \mathbf{E}\left(x_{\ell 2}\right) = 0$. Therefore, using (3.86) and the fact that $\mathbf{Cov}\left(x_{\ell r}, x_{\ell c}\right) = (\sigma)_{rc}$, we have the following set of identities:

$$\begin{aligned}
\mathbf{E}\left[(a)_{12}\right] &= n\sigma_{12} \\
\mathbf{E}\left[(a)_{13}\right] &= n\sigma_{13} \\
\vdots &= \vdots \\
\mathbf{E}\left[(a)_{1p}\right] &= n\sigma_{1p} \\
\vdots &= \vdots \\
\mathbf{E}\left[(a)_{p-1,p}\right] &= n\sigma_{p-1,p}
\end{aligned} \tag{3.87}$$

Using (3.85) and (3.87) we can now write the expectation of the random matrix $\mathbf{A}$ as:

$$\mathbf{E}(\mathbf{A}) = n \begin{pmatrix} \sigma_{11}^2 & \sigma_{12} & \dots & \sigma_{1p} \\ \sigma_{21} & \sigma_{22}^2 & \dots & \sigma_{2p} \\ \vdots & \vdots & \ddots & \vdots \\ \sigma_{p1} & \sigma_{p2} & \dots & \sigma_{pp}^2 \end{pmatrix}_{p \times p} \tag{3.88}$$

$$= n\boldsymbol{\Sigma}, \tag{3.89}$$

where $\sigma_{rc} = \sigma_{cr}, r \neq c$. In this document, we will also use the following notation: $\boldsymbol{\Sigma} = (\sigma)_{rc}$. Under this nomenclature, when $r = c$, $(\sigma)_{rc} = (\sigma)_{rr} = \sigma_{rr}^2$. Also, when $r \neq c$, $(\sigma)_{rc} = \sigma_{rc}$. Next, we will employ a similar tactic utilized in (3.84) - (3.89) to calculate $\mathbf{Cov}(\mathbf{A})$. Once again, we will first concentrate on the main diagonal elements for the random matrix $\mathbf{A}$. Using (3.84), we have:

$$\begin{aligned}
\mathbf{Cov}\left[(a)_{11}, (a)_{11}\right] &= \mathbf{Cov}\left[\sum_{\ell=1}^{n} x_{\ell 1}^2, \sum_{\ell=1}^{n} x_{\ell 1}^2\right] \\
&= \mathbf{E}\left[\sum_{\ell=1}^{n} x_{\ell 1}^2, \sum_{\ell=1}^{n} x_{\ell 1}^2\right] - \mathbf{E}\left[\sum_{i\ell 1}^{n} x_{\ell 1}^2\right] \mathbf{E}\left[\sum_{\ell=1}^{n} x_{\ell 1}^2\right] \\
&= \mathbf{E}\left[\sum_{\ell=1}^{n} x_{\ell 1}^2\right]^2 - \left[\mathbf{E}\left(\sum_{\ell=1}^{n} x_{\ell 1}^2\right)\right]^2 \\
&= \mathbf{E}\left[\sum_{\ell=1}^{n} x_{\ell 1}^2\right]^2 - \left[\sum_{\ell=1}^{n} \mathbf{E}\left(x_{\ell 1}^2\right)\right]^2 \\
&= \mathbf{E}\left[\sum_{\ell=1}^{n} x_{\ell 1}^2\right]^2 - \left[n\mathbf{Var}\left(x_{11}\right)\right]^2 \qquad (3.90)
\end{aligned}$$

The last equality follows from the fact that the x_ℓs are independent and identically distributed. Completing the development in (3.90), we can use the following expansions:

$$\sum_{\ell=1}^{n} x_{\ell 1}^2 = \left(x_{11}^2 + x_{21}^2 + \cdots + x_{n1}^2\right) \tag{3.91}$$

$$\begin{aligned}
\left(\sum_{\ell=1}^{n} x_{\ell 1}^2\right)^2 &= \left(x_{11}^2 + x_{21}^2 + \cdots + x_{n1}^2\right) \times \left(x_{11}^2 + x_{21}^2 + \cdots + x_{n1}^2\right) \\
&= x_{11}^2 \left(x_{11}^2 + x_{21}^2 + \cdots + x_{n1}^2\right) + \\
&\quad x_{21}^2 \left(x_{11}^2 + x_{21}^2 + \cdots + x_{n1}^2\right) + \\
&\quad \vdots \\
&\quad + x_{n1}^2 \left(x_{11}^2 + x_{21}^2 + \cdots + x_{n1}^2\right)
\end{aligned} \tag{3.92}$$

Taking the expectation of (3.92), we now have:

$$\begin{aligned}
\mathbf{E}\left[\left(\sum_{\ell=1}^{n} x_{\ell 1}^2\right)^2\right] &= \mathbf{E}\left[x_{11}^2 \left(x_{11}^2 + x_{21}^2 + \cdots + x_{n1}^2\right)\right] + \\
&\quad \mathbf{E}\left[x_{21}^2 \left(x_{11}^2 + x_{21}^2 + \cdots + x_{n1}^2\right)\right] + \\
&\quad \vdots \\
&\quad + \mathbf{E}\left[x_{n1}^2 \left(x_{11}^2 + x_{21}^2 + \cdots + x_{n1}^2\right)\right]
\end{aligned} \tag{3.93}$$

Now, working with the first expectation on the right-hand side of (3.93), we have the following:

$$\mathbf{E}\left[x_{11}^2 x_{11}^2 + x_{11}^2 x_{21}^2 + \cdots + x_{11}^2 x_{n1}^2\right] = \mathbf{E}\left[\left(x_{11}^2\right)^2\right] + \mathbf{E}\left[x_{11}^2\right]\mathbf{E}\left[x_{21}^2\right] + \cdots + \mathbf{E}\left[x_{11}^2\right]\mathbf{E}\left[x_{n1}^2\right] \tag{3.94}$$

$$= \mathbf{E}\left[\left(x_{11}^2\right)^2\right] + \mathbf{E}\left[x_{11}^2\right]\mathbf{E}\left[x_{11}^2\right] + \cdots + \mathbf{E}\left[x_{11}^2\right]\mathbf{E}\left[x_{11}^2\right] \tag{3.95}$$

$$\begin{aligned}
&= \mathbf{E}\left[\left(x_{11}^2\right)^2\right] + (n-1)\left[\mathbf{E}\left(x_{11}^2\right)\right]^2 \\
&= \mathbf{E}\left[x_{11}^4\right] + (n-1)\left[\mathbf{Var}\left(x_{11}\right)\right]^2 \\
&= \mathbf{E}\left[x_{11}^4\right] + (n-1)\,\sigma_{11}^4
\end{aligned} \tag{3.96}$$

We note that (3.94) follows from the fact that the rows of the $\mathbf{X}$ matrix are statistically independent, and (3.95) follows from the fact that the $\mathbf{x}_\ell$s are identically distributed. Thus, if we continue with all the expectations from (3.93), we will have:

$$\mathbf{E}\left[\left(\sum_{\ell=1}^{n} x_{\ell 1}^2\right)^2\right] = n\left[\mathbf{E}\left(x_{11}^4\right) + (n-1)\,\sigma_{11}^4\right] \tag{3.97}$$

Now, combining the results of (3.97) with (3.90) we have the following:

$$\begin{aligned}
\mathbf{Cov}\left[(a)_{11}, (a)_{11}\right] &= n\left[\mathbf{E}\left(x_{11}^4\right) + (n-1)\,\sigma_{11}^4\right] - \left[n\sigma_{11}^2\right]^2 \\
&= n\mathbf{E}\left[x_{11}^4\right] + n\,(n-1)\,\sigma_{11}^4 - n^2\sigma_{11}^4 \\
&= n\mathbf{E}\left[x_{11}^4\right] - n\sigma_{11}^4 \\
&= n\left[\mathbf{E}\left(x_{11}^4\right) - \sigma_{11}^4\right]
\end{aligned} \tag{3.98}$$

Now, let us calculate the covariance for the off-diagonal elements of the random matrix $\mathbf{A}$. For this next step, we will first calculate the covariance between matrix elements $(a)_{11}$ and $(a)_{12}$.

$$\begin{aligned}
\mathbf{Cov}\left[(a)_{11}, (a)_{12}\right] &= \mathbf{Cov}\left[\sum_{\ell=1}^{n} x_{\ell 1}^2, \sum_{\ell=1}^{n} x_{\ell 1}x_{\ell 2}\right] \\
&= \mathbf{E}\left[\sum_{\ell=1}^{n} x_{\ell 1}^2 \sum_{\ell=1}^{n} x_{\ell 1}x_{\ell 2}\right] - \mathbf{E}\left[\sum_{\ell=1}^{n} x_{\ell 1}^2\right]\mathbf{E}\left[\sum_{\ell=1}^{n} x_{\ell 1}x_{\ell 2}\right] \\
&= \mathbf{E}\left[\sum_{\ell=1}^{n} x_{\ell 1}^2 \sum_{\ell=1}^{n} x_{\ell 1}x_{\ell 2}\right] - \left[\sum_{\ell=1}^{n} \mathbf{E}\left(x_{\ell 1}^2\right)\right]\left[\sum_{\ell=1}^{n} \mathbf{E}\left(x_{\ell 1}x_{\ell 2}\right)\right] \\
&= \mathbf{E}\left[\sum_{\ell=1}^{n} x_{\ell 1}^2 \sum_{\ell=1}^{n} x_{\ell 1}x_{\ell 2}\right] - \left[n\mathbf{Var}\left(x_{11}\right)\right]\left[n\mathbf{Cov}\left(x_{11}, x_{12}\right)\right] \\
&= \mathbf{E}\left[\sum_{\ell=1}^{n} x_{\ell 1}^2 \sum_{\ell=1}^{n} x_{\ell 1}x_{\ell 2}\right] - \left[n\sigma_{11}^2\right]\left[n\sigma_{12}\right]
\end{aligned} \tag{3.99}$$

Working with the expectation on the right-hand side of (3.99), we have:

$$\mathbf{E}\left[\sum_{\ell=1}^{n} x_{\ell 1}^2 \sum_{\ell=1}^{n} x_{\ell 1}x_{\ell 2}\right] = \mathbf{E}\left[\left(x_{11}^2 + x_{21}^2 + \cdots + x_{n1}^2\right)\left(x_{11}x_{12} + x_{21}x_{22} + \cdots + x_{n1}x_{n2}\right)\right] \tag{3.100}$$

$$\begin{aligned} &= \mathbf{E}\left[x_{11}^2\left(x_{11}x_{12} + x_{21}x_{22} + \cdots + x_{n1}x_{n2}\right)\right] + \\ &\quad \mathbf{E}\left[x_{21}^2\left(x_{11}x_{12} + x_{21}x_{22} + \cdots + x_{n1}x_{n2}\right)\right] + \\ &\quad \vdots \\ &\quad + \mathbf{E}\left[x_{n1}^2\left(x_{11}x_{12} + x_{21}x_{22} + \cdots + x_{n1}x_{n2}\right)\right] \end{aligned} \tag{3.101}$$

Working with the first expectation on the right-hand side of (3.101), we have the following:

$$\begin{aligned} \mathbf{E}\left[x_{11}^2\left(x_{11}x_{12} + x_{21}x_{22} + \cdots + x_{n1}x_{n2}\right)\right] &= \mathbf{E}\left[x_{11}^3x_{12} + x_{11}^2\left(x_{21}x_{22} + x_{31}x_{32} + \cdots + x_{n1}x_{n2}\right)\right] \\ &= \mathbf{E}\left[x_{11}^3x_{12}\right] + \\ &\quad \mathbf{E}\left[x_{11}^2\right]\left(\mathbf{E}\left[x_{21}x_{22}\right] + \mathbf{E}\left[x_{31}x_{32}\right] + \cdots + \mathbf{E}\left[x_{n1}x_{n2}\right]\right) \\ &= \mathbf{E}\left[x_{11}^3x_{12}\right] + \\ &\quad \mathbf{E}\left[x_{11}^2\right]\left(\mathbf{Cov}\left[x_{21}x_{22}\right] + \mathbf{Cov}\left[x_{31}x_{32}\right] + \cdots + \right. \\ &\quad \left.\mathbf{Cov}\left[x_{n1}x_{n2}\right]\right) \qquad (3.102) \\ &= \mathbf{E}\left[x_{11}^3x_{12}\right] + \mathbf{E}\left[x_{11}^2\right]\left(\left[n-1\right]\mathbf{Cov}\left[x_{21}, x_{22}\right]\right) \qquad (3.103) \\ &= \mathbf{E}\left[x_{11}^3x_{12}\right] + \mathbf{Var}\left[x_{11}\right]\left(\left[n-1\right]\mathbf{Cov}\left[x_{21}, x_{22}\right]\right) \qquad (3.104) \\ &= \mathbf{E}\left[x_{11}^3x_{12}\right] + \sigma_{11}^2\left(\left[n-1\right]\sigma_{21}\right) \qquad (3.105) \end{aligned}$$

The development on the right-hand side of (3.103) follows from the fact that there are $(n-1)$ covariance terms on the right-hand side of (3.102), and that, once again, the rows of the $\mathbf{X}$ matrix are identically distributed. Substituting the corresponding values from (3.88) into the right-hand side of (3.104) returns (3.105). We note that the remaining expectation on the right-hand side of (3.105) is a higher-ordered term. One way to calculate this expectation is through the use of moment generating functions. Because we are assuming that $\mathbf{x}_\ell \sim \mathcal{N}(\mathbf{0}, \boldsymbol{\Sigma})$ and using well-established properties of the multivariate Gaussian distribution, we know that the joint probability distribution of x_{11} and x_{12} is as follows:

$$f(x_{11}, x_{12}) = \mathcal{N}_2 \left(\begin{pmatrix} 0 \\ 0 \end{pmatrix}, \begin{pmatrix} \sigma_{11}^2 & \sigma_{12} \\ \sigma_{21} & \sigma_{22}^2 \end{pmatrix} \right), \text{where } \sigma_{12} = \sigma_{21} \tag{3.106}$$

The probability distribution in (3.106) is well-recognized as the bivariate Gaussian distribution. For this distribution it is also well-established that its MGF is:

$$\begin{aligned}
M(\mathbf{t}) &= \exp\left(\frac{1}{2}\mathbf{t}^T \boldsymbol{\Sigma} \mathbf{t}\right), \text{where } \mathbf{t} = \begin{pmatrix} t_{11} \\ t_{21} \end{pmatrix}, \ \boldsymbol{\Sigma} = \begin{pmatrix} \sigma_{11}^2 & \sigma_{12} \\ \sigma_{21} & \sigma_{22}^2 \end{pmatrix} \\
&= \exp\left(\frac{1}{2} \begin{pmatrix} t_{11} & t_{21} \end{pmatrix} \begin{pmatrix} \sigma_{11}^2 & \sigma_{12} \\ \sigma_{21} & \sigma_{22}^2 \end{pmatrix} \begin{pmatrix} t_{11} \\ t_{21} \end{pmatrix}\right) \\
&= \exp\left(\frac{1}{2} \begin{pmatrix} t_{11}\sigma_{11}^2 + t_{21}\sigma_{21} & t_{11}\sigma_{12} + t_{21}\sigma_{22}^2 \end{pmatrix} \begin{pmatrix} t_{11} \\ t_{21} \end{pmatrix}\right) \\
&= \exp\left(\frac{1}{2}\left[t_{11}(t_{11}\sigma_{11}^2 + t_{21}\sigma_{21}) + t_{21}(t_{11}\sigma_{12} + t_{21}\sigma_{22}^2)\right]\right) \\
&= \exp\left(\frac{1}{2}\left[t_{11}^2\sigma_{11}^2 + t_{11}t_{21}\sigma_{21} + t_{21}t_{11}\sigma_{12} + t_{21}^2\sigma_{22}^2\right]\right) \\
&= \exp\left(\frac{1}{2}\left[t_{11}^2\sigma_{11}^2 + 2t_{11}t_{21}\sigma_{12} + t_{21}^2\sigma_{22}^2\right]\right)
\end{aligned} \tag{3.107}$$

The last equality in (3.107) follows from the symmetry of $\boldsymbol{\Sigma}$. The MGF in (3.107) can be used to calculate the moments of the random variables x_{11} and x_{12}.

Specifically, we are first interested in calculating $\mathbf{E}\left[x_{11}^3 x_{12}\right]$ from (3.105). To begin, we shall first calculate various partial derivatives for the MGF in (3.107).

$$\frac{\partial M(\mathbf{t})}{\partial t_{11}} = \exp\left(\frac{1}{2}\left[t_{11}^2\sigma_{11}^2 + 2t_{11}t_{21}\sigma_{12} + t_{21}^2\sigma_{22}^2\right]\right)\left(\frac{1}{2}\left[2t_{11}\sigma_{11}^2 + 2t_{21}\sigma_{12}\right]\right) \tag{3.108}$$

$$\begin{aligned}\frac{\partial M(\mathbf{t})}{\partial t_{11}^2} &= \exp\left(\frac{1}{2}\left[t_{11}^2\sigma_{11}^2 + 2t_{11}t_{21}\sigma_{12} + t_{21}^2\sigma_{22}^2\right]\right)\left(\frac{1}{2}\left[2\sigma_{11}^2\right]\right) + \\ &\quad \exp\left(\frac{1}{2}\left[t_{11}^2\sigma_{11}^2 + 2t_{11}t_{21}\sigma_{12} + t_{21}^2\sigma_{22}^2\right]\right)\left(\frac{1}{2}\left[2t_{11}\sigma_{11}^2 + 2t_{21}\sigma_{12}\right]\right)^2 \\ &= \exp\left(\frac{1}{2}\left[t_{11}^2\sigma_{11}^2 + 2t_{11}t_{21}\sigma_{12} + t_{21}^2\sigma_{22}^2\right]\right)\sigma_{11}^2 + \\ &\quad \exp\left(\frac{1}{2}\left[t_{11}^2\sigma_{11}^2 + 2t_{11}t_{21}\sigma_{12} + t_{21}^2\sigma_{22}^2\right]\right)\left[t_{11}\sigma_{11}^2 + t_{21}\sigma_{12}\right]^2 \\ &= \exp\left(\frac{1}{2}\left[t_{11}^2\sigma_{11}^2 + 2t_{11}t_{21}\sigma_{12} + t_{21}^2\sigma_{22}^2\right]\right)\left(\sigma_{11}^2 + \left[t_{11}\sigma_{11}^2 + t_{21}\sigma_{12}\right]^2\right) \end{aligned} \tag{3.109}$$

$$\begin{aligned}\frac{\partial M(\mathbf{t})}{\partial t_{11}^3} &= \exp\left(\frac{1}{2}\left[t_{11}^2\sigma_{11}^2 + 2t_{11}t_{21}\sigma_{12} + t_{21}^2\sigma_{22}^2\right]\right)\left(2\sigma_{11}^2\left[t_{11}\sigma_{11}^2 + t_{21}\sigma_{12}\right]\right) + \\ &\quad \left(\sigma_{11}^2 + \left[t_{11}\sigma_{11}^2 + t_{21}\sigma_{12}\right]^2\right)\exp\left(\frac{1}{2}\left[t_{11}^2\sigma_{11}^2 + 2t_{11}t_{21}\sigma_{12} + t_{21}^2\sigma_{22}^2\right]\right) \times \\ &\quad \left(\frac{1}{2}\left[2t_{11}\sigma_{11}^2 + 2t_{21}\sigma_{12}\right]\right) \\ &= \exp\left(\frac{1}{2}\left[t_{11}^2\sigma_{11}^2 + 2t_{11}t_{21}\sigma_{12} + t_{21}^2\sigma_{22}^2\right]\right) \times \\ &\quad \left(2\sigma_{11}^2\left[t_{11}\sigma_{11}^2 + t_{21}\sigma_{12}\right] + \left(\sigma_{11}^2 + \left[t_{11}\sigma_{11}^2 + t_{21}\sigma_{12}\right]^2\right)\left[t_{11}\sigma_{11}^2 + t_{21}\sigma_{12}\right]\right) \\ &= \exp\left(\frac{1}{2}\left[t_{11}^2\sigma_{11}^2 + 2t_{11}t_{21}\sigma_{12} + t_{21}^2\sigma_{22}^2\right]\right) \times \\ &\quad \left(2\sigma_{11}^2\left[t_{11}\sigma_{11}^2 + t_{21}\sigma_{12}\right] + \sigma_{11}^2\left[t_{11}\sigma_{11}^2 + t_{21}\sigma_{12}\right] + \left[t_{11}\sigma_{11}^2 + t_{21}\sigma_{12}\right]^3\right) \end{aligned} \tag{3.110}$$

$$\begin{aligned}\frac{\partial M(\mathbf{t})}{\partial t_{11}^3\partial t_{21}} &= \exp\left(\frac{1}{2}\left[t_{11}^2\sigma_{11}^2 + 2t_{11}t_{21}\sigma_{12} + t_{21}^2\sigma_{22}^2\right]\right)\left(3\sigma_{11}^2\sigma_{12} + 3\sigma_{12}\left[t_{11}\sigma_{11}^2 + t_{21}\sigma_{12}\right]^2\right) + \\ &\quad \left(3\sigma_{11}^2\left[t_{11}\sigma_{11}^2 + t_{21}\sigma_{12}\right] + \left[t_{11}\sigma_{11}^2 + t_{21}\sigma_{12}\right]^3\right) \times \\ &\quad \exp\left(\frac{1}{2}\left[t_{11}^2\sigma_{11}^2 + 2t_{11}t_{21}\sigma_{12} + t_{21}^2\sigma_{22}^2\right]\right)\left(\frac{1}{2}\left[2t_{11}\sigma_{12} + 2t_{21}\sigma_{22}^2\right]\right) \end{aligned} \tag{3.111}$$

Using (3.111), we can calculate $\mathbf{E}(x_{11}^3 x_{12})$ as follows:

$$\mathbf{E}(x_{11}^3 x_{12}) = \left.\frac{\partial M(\mathbf{t})}{\partial t_{11}^3 \partial t_{21}}\right|_{\substack{t_{11}=0 \\ t_{21}=0}} = 3\sigma_{11}^2 \sigma_{21} \tag{3.112}$$

Therefore, we can now substitute (3.111) into (3.105) to obtain:

$$\begin{aligned}
\mathbf{E}\left[x_{11}^2 \left(x_{11}x_{12} + x_{21}x_{22} + \cdots + x_{n1}x_{n2}\right)\right] &= \mathbf{E}\left[x_{11}^3 x_{12}\right] + \sigma_{11}^2 \left(\left[n-1\right]\sigma_{21}\right) \\
&= 3\sigma_{11}^2\sigma_{21} + \sigma_{11}^2 \left(\left[n-1\right]\sigma_{21}\right) \\
&= 3\sigma_{11}^2\sigma_{21} + (n-1)\sigma_{11}^2\sigma_{21} \qquad (3.113)
\end{aligned}$$

Now, taking (3.113) and using the property that the rows of the $\mathbf{X}$ matrix are identically distributed, we have the following:

$$\mathbf{E}\left[\sum_{\ell=1}^{n} x_{\ell 1}^2 \sum_{\ell=1}^{n} x_{\ell 1}x_{\ell 2}\right] = n\left[3\sigma_{11}^2\sigma_{21} + (n-1)\sigma_{11}^2\sigma_{21}\right] \tag{3.114}$$

Finally, substituting (3.114) into (3.99), we have the following expression:

$$\begin{aligned}
\mathbf{Cov}\left[(a)_{11}, (a)_{12}\right] &= \mathbf{Cov}\left[\sum_{\ell=1}^{n} x_{\ell 1}^2, \sum_{\ell=1}^{n} x_{\ell 1}x_{\ell 2}\right] \\
&= n\left[3\sigma_{11}^2\sigma_{21} + (n-1)\sigma_{11}^2\sigma_{21}\right] - \left[n\sigma_{11}^2\right]\left[n\sigma_{21}\right] \\
&= 3n\sigma_{11}^2\sigma_{21} + n(n-1)\sigma_{11}^2\sigma_{21} - n^2\sigma_{11}^2\sigma_{21} \\
&= 3n\sigma_{11}^2\sigma_{21} - n\sigma_{11}^2\sigma_{21} \\
&= 2n\sigma_{11}^2\sigma_{21} \qquad (3.115)
\end{aligned}$$

Now, returning to (3.98) we note that we still need to obtain the expression for $\mathbf{E}\left[x_{11}^4\right]$ so we can finalize the expression for $\mathbf{Cov}\left[(a)_{11}, (a)_{11}\right]$. To assist in the calculation of the remaining expectation, let us return to (3.110) to next calculate the fourth partial derivative of the MGF in (3.107) with respect to t_{11}.

$$
\begin{aligned}
\frac{\partial M(\mathbf{t})}{\partial t_{11}^4} = \exp\left(\frac{1}{2}\left[t_{11}^2\sigma_{11}^2 + 2t_{11}t_{21}\sigma_{12} + t_{21}^2\sigma_{22}^2\right]\right)&\left(2\sigma_{11}^4 + \sigma_{11}^4 + 3\sigma_{11}^2\left[t_{11}\sigma_{11}^2 + t_{21}\sigma_{12}\right]^2\right) + \\
&\left(2\sigma_{11}^2\left[t_{11}\sigma_{11}^2 + t_{21}\sigma_{12}\right] + \sigma_{11}^2\left[t_{11}\sigma_{11}^2 + t_{21}\sigma_{12}\right] + \left[t_{11}\sigma_{11}^2 + t_{21}\sigma_{12}\right]^3\right) \times \\
&\exp\left(\frac{1}{2}\left[t_{11}^2\sigma_{11}^2 + 2t_{11}t_{21}\sigma_{12} + t_{21}^2\sigma_{22}^2\right]\right)\left(\frac{1}{2}\left[2t_{11}\sigma_{11}^2 + 2t_{21}\sigma_{12}\right]\right) \\
= \exp&\left(\frac{1}{2}\left[t_{11}^2\sigma_{11}^2 + 2t_{11}t_{21}\sigma_{12} + t_{21}^2\sigma_{22}^2\right]\right) \times \\
&\left(3\sigma_{11}^4 + 3\sigma_{11}^2\left[t_{11}\sigma_{11}^2 + t_{21}\sigma_{12}\right]^2 + \left[3\sigma_{11}^2\left(t_{11}\sigma_{11}^2 + t_{21}\sigma_{12}\right) + \left(t_{11}\sigma_{11}^2 + t_{21}\sigma_{12}\right)^3\right]\right) \times \\
&\left[t_{11}\sigma_{11}^2 + t_{21}\sigma_{12}\right]
\end{aligned} \tag{3.116}
$$

Therefore, similar to what was done in (3.112), we now have:

$$
\mathbf{E}(x_{11}^4) = \left.\frac{\partial M(\mathbf{t})}{\partial t_{11}^4}\right|_{\substack{t_{11}=0 \\ t_{21}=0}} = 3\sigma_{11}^4 \tag{3.117}
$$

Finally, substituting (3.117) into (3.98) we have:

$$
\begin{aligned}
\mathbf{Cov}\left[(a)_{11}, (a)_{11}\right] &= n\left(\mathbf{E}\left[x_{11}^4\right] - \sigma_{11}^4\right) \\
&= n\left(3\sigma_{11}^4 - \sigma_{11}^4\right) \\
&= 2n\sigma_{11}^4
\end{aligned} \tag{3.118}
$$

Applying the developments in (3.84) - (3.118) to the entire $\mathbf{A}$ matrix, we now have the following general expression:

$$
\mathbf{Cov}\left[\mathbf{A}\right] = \mathbf{Cov}\left[(a)_{uv}, (a)_{st}\right] = n\left[(\sigma)_{us}(\sigma)_{vt} + (\sigma)_{ut}(\sigma)_{vs}\right], \tag{3.119}
$$

where $\mathbf{\Sigma} = (\sigma)_{uv}$, $1 \leq u \leq v \leq p,\ 1 \leq s \leq t \leq p$. In (3.119), n is referred to as the degrees of freedom. Now, based on the results in (3.89) and (3.119), we can also show the expectations and variances for the finite mixture of Wisharts distribution shown in (3.80). Using notation from earlier,

$$
f(\mathbf{A}^*) = \sum_{j=1}^{k} w_j W_p\left(f_j, \mathbf{\Sigma}_j\right), \text{where} \tag{3.120}
$$

$$\mathbf{A}^* = \text{the matrix of sum of squares and the sum of cross-products}$$

$$w_j = \text{the mixture weight for the } j^{th} \text{ component Wishart distribution}$$

$$\mathbf{\Sigma}_j = \text{the covariance matrix for the } j^{th} \text{ component Wishart distribution}$$

$$f_j = \text{the degrees of freedom for the } j^{th} \text{ component Wishart distribution}$$

Based on (3.5), Theorem 15 in Appendix D, and developments similar to those demonstrated in (3.84) - (3.89), we can express the expectation of $\mathbf{A}^*$ as

$$\mathbf{E}\left(\mathbf{A}^*\right) = \sum_{j=1}^{k} w_j f_j \mathbf{\Sigma}_j, \tag{3.121}$$

or on an element-by-element basis for $\mathbf{\Sigma}_j$,

$$\mathbf{E}\left[\left(\mathbf{A}^*\right)_{uv}\right] = \sum_{j=1}^{k} w_j f_j \left(\sigma\right)_{uv}^{j}, \; 1 \le u \le v \le p \text{ where} \tag{3.122}$$

$$\left(\sigma\right)_{uv}^{j} = \text{the element in the } u^{th} \text{ row and } v^{th} \text{ column of } \mathbf{\Sigma}_j$$

It is also important to note that in (3.122), p refers to the dimension of the covariance matrices. If all the f_j in (3.122) are all equal, then the expectation in (3.122) becomes:

$$\mathbf{E}\left[\left(\mathbf{A}^*\right)_{uv}\right] = f \sum_{j=1}^{k} w_j \left(\sigma\right)_{uv}^{j}, \; 1 \le u \le v \le p \tag{3.123}$$

For calculating the variances and covariances for the finite mixture of Wisharts distribution, we can apply a similar approach as demonstrated in (3.12), (3.90) - (3.118) as well as Theorem 15 in Appendix D. Similar to $\mathbf{A}$, let us define $\mathbf{A}^*$ as:

$$\mathbf{A}^* = \begin{pmatrix} \sum_{\ell=1}^{n} x_{\ell 1}^{*2} & \sum_{\ell=1}^{n} x_{\ell 1}^{*} x_{\ell 2}^{*} & \cdots & \sum_{\ell=1}^{n} x_{\ell 1}^{*} x_{\ell p}^{*} \\ \sum_{\ell=1}^{n} x_{\ell 2}^{*} x_{\ell 1}^{*} & \sum_{\ell=1}^{n} x_{\ell 2}^{*2} & \cdots & \sum_{\ell=1}^{n} x_{\ell 2}^{*} x_{\ell p}^{*} \\ \vdots & \vdots & \ddots & \vdots \\ \sum_{\ell=1}^{n} x_{\ell p}^{*} x_{\ell 1}^{*} & \sum_{\ell=1}^{n} x_{\ell p}^{*} x_{\ell 2}^{*} & \cdots & \sum_{\ell=1}^{n} x_{\ell p}^{*2} \end{pmatrix}, \tag{3.124}$$

where $\mathbf{x}_\ell^* \sim \sum_{j=1}^{k} w_j \mathcal{N}_p(\mathbf{0}, \mathbf{\Sigma}_j)$. Similar to the univariate case in section 5.1, we note that the random matrix $\mathbf{A}^*$ is assumed to have been generated heterogeneously. That is, each matrix element of $\mathbf{A}^*$ is assumed to have some of its values generated from each component distribution (Wishart). Stated another way, the $\mathbf{x}_\ell$ vectors from (3.72) are assumed to have been generated from each of the k-component multivariate Gaussian distrbutions. As a result, $\mathbf{A}^*$ can also be expressed as a weighted sum of Wishart-distributed random matrices. That is:

$$\mathbf{A}^* = \sum_{j=1}^{k} w_j W_j, \tag{3.125}$$

where $W_j \sim W_p(f_j, \mathbf{\Sigma}_j)$. We can also state (3.125) on an element-by-element basis as:

$$(a^*)_{uv} = \sum_{j=1}^{k} w_j \, (a)_{uv}^{j}, \tag{3.126}$$

where, as before, $1 \leq u \leq v \leq p$.

$$\begin{aligned}
\mathbf{Cov}\,(\mathbf{A}^*) &= \mathbf{Cov}\left[(a^*)_{uv}, (a^*)_{st}\right] && (3.127)\\
&= \mathbf{E}\left[(a^*)_{uv}(a^*)_{st}\right] - \mathbf{E}\left[(a^*)_{uv}\right]\mathbf{E}\left[(a^*)_{st}\right] && (3.128)\\
&= \mathbf{E}\left[(a^*)_{uv}(a^*)_{st}\right] - \left[\sum_{j=1}^{k} w_j f_j\,(\sigma)_{uv}^{j}\right]\left[\sum_{j=1}^{k} w_j f_j\,(\sigma)_{st}^{j}\right] && (3.129)\\
&= \mathbf{E}\left[\left\{\sum_{j=1}^{k} w_j\,(a)_{uv}^{j}\right\}\left\{\sum_{j=1}^{k} w_j\,(a)_{st}^{j}\right\}\right] - \left[\sum_{j=1}^{k} w_j f_j\,(\sigma)_{uv}^{j}\right]\left[\sum_{j=1}^{k} w_j f_j\,(\sigma)_{st}^{j}\right] && (3.130)\\
&= \mathbf{E}\left[\sum_{j=1}^{k} w_j^2\,(a)_{uv}^{j}\,(a)_{st}^{j} + \sum_{j=1}^{k}\sum_{m\neq j}^{k} w_j w_m\,(a)_{uv}^{j}\,(a)_{st}^{m}\right] - \\
&\quad \left[\sum_{j=1}^{k} w_j f_j\,(\sigma)_{uv}^{j}\right]\left[\sum_{j=1}^{k} w_j f_j\,(\sigma)_{st}^{j}\right] && (3.131)
\end{aligned}$$

Let $u = v = s = t$ and working from the right-hand side of (3.131):

$$\begin{aligned}
\mathbf{Cov}\left[(a^*)_{uu}, (a^*)_{uu}\right] &= \mathbf{E}\left[\sum_{j=1}^{k} w_j^2 \left\{(a)_{uu}^j\right\}^2\right] + \\
&\quad \mathbf{E}\left[\sum_{j=1}^{k}\sum_{m \neq j}^{k} w_j w_m (a)_{uu}^j (a)_{uu}^m\right] - \\
&\quad \left[\sum_{j=1}^{k} w_j f_j (\sigma)_{uu}^j\right]^2 && (3.132) \\
&= \sum_{j=1}^{k} w_j^2 \mathbf{E}\left\{(a)_{uu}^j\right\}^2 + \\
&\quad \sum_{j=1}^{k}\sum_{m \neq j}^{k} w_j w_m \mathbf{E}\,(a)_{uu}^j\, \mathbf{E}\,(a)_{uu}^m - \\
&\quad \left[\sum_{j=1}^{k} w_j f_j (\sigma)_{uu}^j\right]^2 && (3.133) \\
&= \sum_{j=1}^{k} w_j^2 \mathbf{E}\left\{\sum_{\ell=1}^{f_j} \left(x_{\ell u}^2\right)^j\right\}^2 + \\
&\quad \sum_{j=1}^{k}\sum_{m \neq j}^{k} w_j w_m \mathbf{E}\left\{\sum_{\ell=1}^{f_j} \left(x_{\ell u}^2\right)^j\right\} \mathbf{E}\left\{\sum_{\ell=1}^{f_m} \left(x_{\ell u}^2\right)^m\right\} - \\
&\quad \left[\sum_{j=1}^{k} w_j f_j (\sigma)_{uu}^j\right]^2 && (3.134) \\
&= \sum_{j=1}^{k} w_j^2 \mathbf{E}\left\{\sum_{\ell=1}^{f_j} \left(x_{\ell u}^2\right)^j\right\}^2 + \\
&\quad \sum_{j=1}^{k}\sum_{m \neq j}^{k} w_j w_m \sum_{\ell=1}^{f_j} \mathbf{E}\left(x_{\ell u}^2\right)^j \sum_{\ell=1}^{f_m} \mathbf{E}\left(x_{\ell u}^2\right)^m - \\
&\quad \left[\sum_{j=1}^{k} w_j f_j (\sigma)_{uu}^j\right]^2 && (3.135)
\end{aligned}$$

Using (3.97) and applying to the right-hand side of (3.135), we have:

$$
\begin{aligned}
&= \sum_{j=1}^{k} w_j^2 \left\{ f_j \left[\mathbf{E} \left(x_{uu}^4 \right)^j + (f_j - 1) \left(\sigma_{uu}^4 \right)^j \right] \right\} + \\
&\sum_{j=1}^{k} \sum_{m \neq j}^{k} w_j w_m f_j \left(\sigma \right)_{uu}^j f_m \left(\sigma \right)_{uu}^m - \\
&\left[\sum_{j=1}^{k} w_j f_j \left(\sigma \right)_{uu}^j \right]^2 \qquad (3.136) \\
&= \sum_{j=1}^{k} w_j^2 \left\{ f_j \left[3 \left(\sigma_{uu}^4 \right)^j + (f_j - 1) \left(\sigma_{uu}^4 \right)^j \right] \right\} + \\
&\sum_{j=1}^{k} \sum_{m \neq j}^{k} w_j w_m f_j f_m \left(\sigma \right)_{uu}^j \left(\sigma \right)_{uu}^m - \\
&\left[\sum_{j=1}^{k} w_j f_j \left(\sigma \right)_{uu}^j \right]^2 \qquad (3.137) \\
&= \sum_{j=1}^{k} w_j^2 \left\{ 2 f_j \left(\sigma_{uu}^4 \right)^j + f_j^2 \left(\sigma_{uu}^4 \right)^j \right\} + \\
&\sum_{j=1}^{k} \sum_{m \neq j}^{k} w_j w_m f_j f_m \left(\sigma \right)_{uu}^j \left(\sigma \right)_{uu}^m - \\
&\left[\sum_{j=1}^{k} w_j f_j \left(\sigma \right)_{uu}^j \right]^2 \qquad (3.138) \\
&= \sum_{j=1}^{k} w_j^2 f_j \left(\sigma_{uu}^4 \right)^j \left[2 + f_j \right] + \\
&\sum_{j=1}^{k} \sum_{m \neq j}^{k} w_j w_m f_j f_m \left(\sigma \right)_{uu}^j \left(\sigma \right)_{uu}^m - \\
&\left[\sum_{j=1}^{k} w_j^2 f_j^2 \left(\sigma_{uu}^4 \right)^j + \sum_{j=1}^{k} \sum_{m \neq j}^{k} w_j w_m f_j f_m \left(\sigma \right)_{uu}^j \left(\sigma \right)_{uu}^m \right] \qquad (3.139) \\
&= \sum_{j=1}^{k} w_j^2 f_j \left[2 \left(\sigma_{uu}^4 \right)^j \right] \qquad (3.140) \\
&= \sum_{j=1}^{k} w_j^2 f_j \left[\left(\sigma \right)_{us}^j \left(\sigma \right)_{vt}^j + \left(\sigma \right)_{ut}^j \left(\sigma \right)_{vs}^j \right] \qquad (3.141)
\end{aligned}
$$

Now let $u = s, v < t$ and working with the right-hand side of (3.131) we have:

$$
\begin{aligned}
\mathbf{Cov}\left[(a^*)_{us}, (a^*)_{vt}\right] = {} & \mathbf{E}\left[\sum_{j=1}^{k} w_j^2 (a)_{us}^j (a)_{vt}^j\right] + \\
& \mathbf{E}\left[\sum_{j=1}^{k} \sum_{m \neq j}^{k} w_j w_m (a)_{us}^j (a)_{vt}^m\right] - \\
& \left[\sum_{j=1}^{k} w_j f_j (\sigma)_{us}^j\right] \times \\
& \left[\sum_{j=1}^{k} w_j f_j (\sigma)_{vt}^j\right] \qquad (3.142)\\
= {} & \sum_{j=1}^{k} w_j^2 \mathbf{E}\left\{(a)_{us}^j (a)_{vt}^j\right\} + \\
& \sum_{j=1}^{k} \sum_{m \neq j}^{k} w_j w_m \mathbf{E}(a)_{us}^j \mathbf{E}(a)_{vt}^m - \\
& \left[\sum_{j=1}^{k} w_j f_j (\sigma)_{us}^j\right] \times \\
& \left[\sum_{j=1}^{k} w_j f_j (\sigma)_{vt}^j\right] \qquad (3.143)\\
= {} & \sum_{j=1}^{k} w_j^2 \mathbf{E}\left\{\sum_{\ell=1}^{f_j} \left(x_{\ell s}^2\right)^j \sum_{\ell=1}^{f_j} (x_{\ell s} x_{\ell t})^j\right\} + \\
& \sum_{j=1}^{k} \sum_{m \neq j}^{k} w_j w_m \mathbf{E}\left\{\sum_{\ell=1}^{f_j} \left(x_{\ell s}^2\right)^j\right\} \mathbf{E}\left\{\sum_{\ell=1}^{f_m} (x_{\ell s} x_{\ell t})^m\right\} - \\
& \left[\sum_{j=1}^{k} w_j f_j (\sigma)_{us}^j\right] \times \\
& \left[\sum_{j=1}^{k} w_j f_j (\sigma)_{vt}^j\right] \qquad (3.144)
\end{aligned}
$$

Now using (3.99) - (3.105) and substituting into the right-hand-side of (3.144), we have the following:

$$
\begin{aligned}
\mathbf{Cov}\left(\mathbf{A}^*\right) &= \sum_{j=1}^{k} w_j^2 \left[f_j \left\{ \mathbf{E}\left(x_{us}^3 x_{vt}\right)^j + (f_j - 1)\, \mathbf{Var}\left(x_{us}\right)^j \mathbf{Cov}\left[\left(x_{us}\right)^j, \left(x_{vt}\right)^j \right] \right\} \right] + \\
& \sum_{j=1}^{k} \sum_{m \neq j}^{k} w_j w_m f_j \left(\sigma\right)_{us}^j f_m \left(\sigma\right)_{vt}^m - \left[\sum_{j=1}^{k} w_j f_j \left(\sigma\right)_{us}^j \right] \left[\sum_{j=1}^{k} w_j f_j \left(\sigma\right)_{vt}^j \right] \qquad (3.145) \\
&= \sum_{j=1}^{k} w_j^2 \left[f_j \left\{ 3 \left(\sigma\right)_{us}^j \left(\sigma\right)_{vt}^j + (f_j - 1) \left(\sigma\right)_{us}^j \left(\sigma\right)_{vt}^j \right\} \right] + \\
& \sum_{j=1}^{k} \sum_{m \neq j}^{k} w_j w_m f_j f_m \left(\sigma\right)_{us}^j \left(\sigma\right)_{vt}^m - \\
& \left[\sum_{j=1}^{k} w_j^2 f_j^2 \left(\sigma\right)_{us}^j \left(\sigma\right)_{vt}^j + \sum_{j=1}^{k} \sum_{m \neq j}^{k} w_j w_m f_j f_m \left(\sigma\right)_{us}^j \left(\sigma\right)_{vt}^m \right] \qquad (3.146) \\
&= \sum_{j=1}^{k} w_j^2 \left[2 f_j \left(\sigma\right)_{us}^j \left(\sigma\right)_{vt}^j + f_j^2 \left(\sigma\right)_{us}^j \left(\sigma\right)_{vt}^j \right] - \sum_{j=1}^{k} w_j^2 f_j^2 \left(\sigma\right)_{us}^j \left(\sigma\right)_{vt}^j \qquad (3.147) \\
&= \sum_{j=1}^{k} w_j^2 f_j \left(\sigma\right)_{us}^j \left(\sigma\right)_{vt}^j \left[2 + f_j\right] - \sum_{j=1}^{k} w_j^2 f_j^2 \left(\sigma\right)_{us}^j \left(\sigma\right)_{vt}^j \qquad (3.148) \\
&= \sum_{j=1}^{k} w_j^2 f_j \left[2 \left(\sigma\right)_{us}^j \left(\sigma\right)_{vt}^j \right] \qquad (3.149) \\
&= \sum_{j=1}^{k} w_j^2 f_j \left[\left(\sigma\right)_{us}^j \left(\sigma\right)_{vt}^j + \left(\sigma\right)_{ut}^j \left(\sigma\right)_{vs}^j \right] \qquad (3.150)
\end{aligned}
$$

We note that (3.150) holds either when $u = s = v < t$ or $u = s = t < v$. Therefore, combining the results in (3.141) and (3.150) we have:

$$
\mathbf{Cov}\left[\mathbf{A}^*\right] = \mathbf{Cov}\left[\left(a^*\right)_{uv}, \left(a^*\right)_{st}\right] = \sum_{j=1}^{k} {w_j}^2 f_j \left[\left(\sigma\right)_{us}^j \left(\sigma\right)_{vt}^j + \left(\sigma\right)_{ut}^j \left(\sigma\right)_{vs}^j \right], \qquad (3.151)
$$

where $1 \leq u \leq v \leq p, 1 \leq s \leq t \leq 1$. Now, based on the moments in (3.122) and (3.151) we can apply the approximation method demonstrated in section 3.1 to the multivariate case.

Let us assume that we are going to approximate $f(\mathbf{A}^*)$ as shown in (3.80) with $f(\mathbf{A})$ as follows:

$$f\left(\mathbf{A}\right)=\{2^{(gp/2)}\Gamma_p\left(\frac{g}{2}\right)\det\left(\mathbf{\Omega}\right)^{g/2}\}^{-1}\det\left(\mathbf{A}\right)^{(g-p-1)/2}\operatorname{etr}\left(-\frac{1}{2}\mathbf{\Omega}^{-1}\mathbf{A}\right) \tag{3.152}$$

$$=W_p\left(g,\mathbf{\Omega}\right), \tag{3.153}$$

where $\mathbf{A}>\mathbf{0}, g>(p-1)$, and $\mathbf{A}$ is a random symmetric matrix that is positive definite. By using (3.89) and (3.121) to equate the expectations of $\mathbf{A}$ and $\mathbf{A}^*$ we have:

$$\mathbf{E}\left(\mathbf{A}\right)=\mathbf{E}\left(\mathbf{A}^*\right)$$

$$g\mathbf{\Omega}=\sum_{j=1}^{k}w_jf_j\mathbf{\Sigma}_j$$

$$\mathbf{\Omega}=\left(\frac{1}{g}\right)\sum_{j=1}^{k}w_jf_j\mathbf{\Sigma}_j \tag{3.154}$$

The equation in (3.154) can also be written in a matrix element-by-element basis as:

$$\left(\omega\right)_{uv}=\left(\frac{1}{g}\right)\sum_{j=1}^{k}w_jf_j\left(\sigma\right)_{uv}^{j} \tag{3.155}$$

Similarly, we can turn our attention to equating covariances for $\mathbf{A}$ and $\mathbf{A}^*$:

$$\mathbf{Cov}\left(\mathbf{A}\right)=\mathbf{Cov}\left(\mathbf{A}^*\right)$$

$$g\left[\left(\omega\right)_{us}\left(\omega\right)_{st}+\left(\omega\right)_{ut}\left(\omega\right)_{vs}\right]=\sum_{j=1}^{k}{w_j}^2f_j\left[\left(\sigma\right)_{us}^{j}\left(\sigma\right)_{vt}^{j}+\left(\sigma\right)_{ut}^{j}\left(\sigma\right)_{vs}^{j}\right]$$

$$g=\frac{\sum_{j=1}^{k}{w_j}^2f_j\left[\left(\sigma\right)_{us}^{j}\left(\sigma\right)_{vt}^{j}+\left(\sigma\right)_{ut}^{j}\left(\sigma\right)_{vs}^{j}\right]}{\left[\left(\omega\right)_{us}\left(\omega\right)_{st}+\left(\omega\right)_{ut}\left(\omega\right)_{vs}\right]} \tag{3.156}$$

But as we can see from equations (3.155) - (3.156), there are more variances, covariances, and expectations than the number of unknowns. Each of the $\frac{1}{2}p(p+1)$ equations for the covariance parameters all involve g. Thus, the degrees of freedom parameter for the approximating distribution cannot be uniquely estimated.

One alternative approach to this problem was suggested by Tan and Gupta [49] who utilized a scalar representation of the covariance matrix (e.g., determinant). From multivariate statistical theory we know that the determinant of a covariance matrix is also known as the generalized variance [45]. Another well-known scalar representation of the covariance matrix is the trace (sum of diagonal elements). The trace of a covariance matrix is also known as the total variance [45]. In addition, a less well-known scalar summary of a covariance matrix is the p-th root of the determinant. We also note that these scalar summaries of a matrix can also be expressed in terms of the eigenvalues of the matrix [45]. Let $\lambda_i, i = 1, \ldots, p$, represent the eigenvalues of a given covariance matrix $\mathbf{V}$ of dimension p. Then, we have:

$$\det(\mathbf{V}) = \prod_{i=1}^{p} \lambda_i \tag{3.157}$$

$$\operatorname{tr}(\mathbf{V}) = \sum_{i=1}^{p} \lambda_i \tag{3.158}$$

$$[\det(\mathbf{V})]^{(1/p)} = \left(\prod_{i=1}^{p} \lambda_i\right)^{(1/p)} \tag{3.159}$$

We may note that (3.159) is the geometric mean of the eigenvalues of the covariance matrix. Based on the potential summary measures in (3.157) - (3.159), we are proposing distributional approximation methods based on the matrix determinant, the matrix trace, the p-th root of the matrix determinant, and a multivariate adaptation of the univariate results in secton 3.1. Each specific method is described below:

<u>Matrix Determinant</u>

We wish to restate (3.156) by equating generalized variances instead of covariances. For this development, it is helpful to express the covariance matrices in matrix form instead of on an element-by-element basis. Let $\mathbf{\Sigma}^* =$ indicate the covariance matrix having the same form as in (3.151). Therefore, equating generalized variances:

$$\det\left[\mathbf{Cov}\left(\mathbf{A}\right)\right] = \det\left[\mathbf{Cov}\left(\mathbf{A}^{*}\right)\right] \tag{3.160}$$

$$\det\left[g\boldsymbol{\Sigma}^{\Omega}\right] = \det\left[\sum_{j=1}^{k} w_j^2 f_j \boldsymbol{\Sigma}_j^{*}\right] \tag{3.161}$$

$$\left(\frac{1}{g^p}\right)\det\left(\boldsymbol{\Sigma}^{\Omega}\right) = \det\left(\sum_{j=1}^{k} w_j^2 f_j \boldsymbol{\Sigma}_j^{*}\right) \tag{3.162}$$

$$g = \left(\frac{\det\left(\boldsymbol{\Sigma}^{\Omega}\right)}{\det\left(\sum_{j=1}^{2} w_j^2 f_j \boldsymbol{\Sigma}^{*}\right)}\right)^{1/p} \tag{3.163}$$

We note that the left-hand side of (3.161) follows from the following relationships:

$$\begin{aligned}
\mathbf{Cov}\left(\mathbf{A}\right) &= \mathbf{Cov}\left((a)_{uv}, (a)_{st}\right) \\
&= g\left[(\sigma)_{us}(\sigma)_{vt} + (\sigma)_{ut}(\sigma)_{vs}\right] && (3.164) \\
&= g\left[\left(\frac{\sum_{j=1}^{k} w_j f_j (\sigma)_{us}^{j}}{g}\right)\left(\frac{\sum_{j=1}^{k} w_j f_j (\sigma)_{us}^{j}}{g}\right) + \left(\frac{\sum_{j=1} w_j f_j (\sigma)_{ut}^{j}}{g}\right)\left(\frac{\sum_{j=1}^{k} w_j f_j (\sigma)_{vs}^{j}}{g}\right)\right] && (3.165) \\
&= g\boldsymbol{\Sigma}^{\Omega} && (3.166)
\end{aligned}$$

Also, (3.162) follows from the properties of determinants, where p is the dimension of the covariance matrices. By examining (3.163) we note that the numerator on the right-hand side is a function of the mean for the mixture distribution. We also note that the right-hand side of (3.163) is a ratio of geometric means of the eigenvalues of the particular covariance matrices shown in the numerator and denominator. Finally, we note that (3.165) follows from the substitution of (3.155) into (3.164).

Matrix Trace

As introduced earlier, another scalar representation of the covariance matrix is the matrix trace. The trace ignores the covariance terms and is simply the sum of the covariance matrix elements on the main diagonal (variances). For this criteria, we will be matching on total variances as opposed to generalized varaiances in the determinant-based criteria. For this approach, we equate total variances as follows:

$$\operatorname{tr}\left[\mathbf{Cov}\left(\mathbf{A}\right)\right] = \operatorname{tr}\left[\mathbf{Cov}\left(\mathbf{A}^{*}\right)\right] \tag{3.167}$$

$$\operatorname{tr}\left[g\mathbf{\Sigma}^{\Omega}\right] = \operatorname{tr}\left[\sum_{j=1}^{k} w_j^2 f_j \mathbf{\Sigma}_j^{*}\right] \tag{3.168}$$

$$\left(\frac{1}{g}\right)\operatorname{tr}\left(\mathbf{\Sigma}^{\Omega}\right) = \operatorname{tr}\left(\sum_{j=1}^{k} w_j^2 f_j \mathbf{\Sigma}_j^{*}\right) \tag{3.169}$$

$$g = \frac{\operatorname{tr}\left(\mathbf{\Sigma}^{\Omega}\right)}{\operatorname{tr}\left(\sum_{j=1}^{k} w_j^2 f_j \mathbf{\Sigma}_j^{*}\right)} \tag{3.170}$$

We first note that the covariance matrices in (3.170) are the same form as shown in (3.163) and (3.164) - (3.166). We also note that (3.169) follows from the properties of the trace of a matrix.

<u>Multivariate Extension of Univariate Method</u>

By reviewing the univariate approximation method in section (3.1), we note that the scale parameter for the approximating univariate distribution (β) in (3.66) is a ratio of the variance and mean from the mixture of distributions case. Similarly, we note that the location parameter for the approximating univariate distribution (α) in (3.64) is the mean from the mixture of distributions case divided by β. Further, we note in (3.67) α from the approximating distribution can be expressed as the square of the mean of the mixture of distributions case divided by the variance from the mixture of distributions case. Thus, based on these relationships, another possible approximation method would be to extend the univariate results in section 3.1 to the multivariate case by performing calculations on a matrix element-by-element basis. Based on previously developed calculations such reflected in (3.119) and (3.151), we have the shown the expressions for $\mathbf{Cov}\left(\mathbf{A}\right)$ and $\mathbf{Cov}\left(\mathbf{A}^{*}\right)$.

1. Let us approximate $f\left(\mathbf{A}^{*}\right) = \sum_{j=1}^{k} w_j W_p\left(f_j, \mathbf{\Sigma}_j\right)$ by using $f\left(\mathbf{A}\right) = W_p\left(g, \mathbf{\Omega}\right)$.

2. Based on a matrix element-by-element basis:

$$(\tilde{g})_{uv} = \frac{\left(\sum_{j=1}^{k} w_j f_j (\sigma)_{uv}^{j}\right)^2}{\sum_{j=1}^{k} w_j^2 f_j \left[(\sigma)_{us}^{j} (\sigma)_{vt}^{j} + (\sigma)_{ut}^{j} (\sigma)_{vs}^{j}\right]}, \tag{3.171}$$

where $1 \leq u = s \leq v \leq p$.

3.

$$\tilde{g} = \frac{2 \sum_{u \leq v} (g)_{uv}}{p(p+1)} \tag{3.172}$$

4. Based on a matrix element-by-element basis:

$$(\omega)_{uv} = \frac{\sum_{j=1}^{k} w_j f_j (\sigma)_{uv}^{j}}{\tilde{g}}, \tag{3.173}$$

where $1 \leq u \leq v \leq p$. If $\tilde{g} < (p-1)$, then $\tilde{g} = \sum_{j=1}^{k} w_j f_j$.

3.2.1 Data simulations - multivariate case

One way to quantify the difference or distance between matrices is to use matrix norms. Because matrix norms are defined in terms of vector norms, it is often stated that the matrix norm is subordinate to or induced by the vector norm. First, some common vector and matrix norms are presented. For a given vector $\mathbf{x}_\ell$ as shown in (3.72), the vector 1-norm is defined as:

$$||\mathbf{x}_\ell||_1 = \sum_{r=1}^{p} |x_{\ell r}| \tag{3.174}$$

For a given matrix $\mathbf{X}$ as shown in (3.73), the matrix 1-norm is defined as:

$$||\mathbf{X}||_1 = \max_{\ell} \left(\sum_{r=1}^{p} |x_{\ell r}|\right) \tag{3.175}$$

Another way to state the matrix 1-norm in (3.175) is that it is the maximum of the column sums of the matrix $\mathbf{X}$. Next, we will proceed to the vector ∞-norm and the matrix ∞-norm. The vector ∞-norm is defined as follows:

$$||\mathbf{x}_\ell||_\infty = \max_p |x_{\ell r}| \tag{3.176}$$

Induced by the vector ∞-norm is the matrix ∞-norm:

$$||\mathbf{X}||_\infty = \max_r \left(\sum_{\ell=1}^{n} |x_{\ell r}| \right) \tag{3.177}$$

Another way to state the matrix ∞-norm is that it is the maximum of the row sums of $\mathbf{X}$. For a given symmetric matrix, the matrix 1-norm and the matrix ∞-norm will be identical. n An additional norm that is useful for statistical applications is the vector 2-norm as well as the matrix 2-norm. The vector 2-norm is defined as:

$$||\mathbf{x}_\ell||_2 = \sqrt{\sum_{\ell=1}^{n} |x_\ell|^2} \tag{3.178}$$

Subordinate to the vector 2-norm is the matrix 2-norm:

$$||\mathbf{X}||_2 = \sqrt{\lambda_{\max}\left(\mathbf{B}^T\mathbf{B}\right)}, \tag{3.179}$$

where λ indicates an eigenvalue and $\mathbf{B}^T\mathbf{B}$ is positive semi-definite. Another way to state the matrix 2-norm is that it is the square root of the largest eigenvalue of the matrix $\mathbf{B}$. The matrix 2-norm is also known as the spectral norm. Finally, another matrix norm that may be useful s the Frobenius norm. This matrix norm is defined as:

$$||\mathbf{X}||_F = \sqrt{\sum_{\ell=1}^{n}\sum_{r=1}^{p} |x_{\ell r}|^2} = \sqrt{tr\left(\mathbf{X}^T\mathbf{X}\right)} = \sqrt{\sum_{r=1}^{p} \lambda_r} \tag{3.180}$$

Another way to state the Frobenius norm is that it is the sum of the squared singular values for the matrix $\mathbf{X}$. The matrix norms will be used to evaluate the adequacy of the approximation method by comparing matrix norms for the mixture of Wishart distributions and the approximated Wishart distribution.

3.2.2 Wishart simulation considerations

Based on the Wishart approximation methods shown in the previous section, it can be noted that the estimate for the degrees of freedom parameter, g, for the approximating distribution may be fractional. In determining whether or not this is justified, we can utilize the following definition which indicates for a Wishart distribution the degrees of freedom parameter belongs to a Gindikin set.

Definition 2. *Gindikin set.* Suppose we have the random matrix $\mathbf{A}$ as defined in (5.79). Further, let us assume that this random matrix has the following Laplace transform:

$$\mathbf{E}\left[\exp\left(\operatorname{tr}\left[\boldsymbol{\Theta}\mathbf{A}\right]\right)\right] = \det\left(\mathbf{I}_p - \boldsymbol{\Sigma}\boldsymbol{\Theta}\right)^{-q},$$

where $\boldsymbol{\Sigma}$ is a $p \times p$ positive-definite matrix and $\boldsymbol{\Theta}$ is a symmetric $p \times p$ matrix. A Gindikin set is the set of values for q such that $q = \{1, 2, \ldots, p-1\} \cup (p-1, \ldots, \infty)$ [50] - [51].

Therefore, by focusing on the non-singular case, $q \in (p-1, \ldots, \infty)$, we note that the Wishart distribution degrees of freedom parameter can take any value witin this interval, including fractional values. Thus, based on the definition, we can generate Wishart random deviates for any q belonging to the Gindikin set. So based on definition 4, it would appear we should be able to generate Wishart random deviates with fractional degrees of freedom. However, as noted by Xiao et al. [52], there may be some special considerations when simulating a Wishart random deviate with fractional degrees of freedom. First, some software packages do not take such a scenario into account. Second, if fractional degrees of freedom are treated as integer-valued, Xiao et al. demonstrated that the impact on results can be quite noticeable. However, it should be noted that this was determined just for a single matrix and a more thorough treatment is also of interest. To have a better understanding of this result, it may be helpful to first illustrate the relationship between the Wishart distribution and the matrix-variate gamma distribution. First, we will define the matrix-variate gamma distribution.

Definition 3. *Matrix-variate gamma distribution (Gupta and Nagar, 2000).* Let $\mathbf{B}$ represent a $p \times p$ symmetric positive-definite random matrix. If $\mathbf{B}$ has the following pdf:

$$f(\mathbf{B}) = \left\{\Gamma_p(\alpha)\det(\mathbf{\Psi})^{-\alpha}\right\}^{-1}\det(\mathbf{B})^{\alpha-(1/2)(p+1)}\operatorname{etr}(-\mathbf{\Psi}\mathbf{B}), Re(\alpha) > \frac{1}{2}(p-1)$$

then $\mathbf{B}$ is said to have a matrix-variate gamma distribution with parameters α and $\mathbf{\Psi}$.

We note that $\Gamma_p(\alpha)$ is the multivariate gamma function as definted in (3.81). We may notice some similarities between Definition 3 and the Wishart pdf shown in (3.152). Let us assume that in Definition 3, $\alpha = f/2$ and $\mathbf{\Psi} = \left(\frac{1}{2}\right)\mathbf{\Sigma}^{-1}$. Substituting, we have:

$$\begin{aligned}
f(\mathbf{B}) &= \left\{\Gamma_p\left(\frac{f}{2}\right)\det\left(\left[\frac{1}{2}\right]\mathbf{\Sigma}^{-1}\right)^{-f/2}\right\}^{-1}\det(\mathbf{B})^{f/2-(1/2)(p+1)}\operatorname{etr}\left(-\left[\frac{1}{2}\right]\mathbf{\Sigma}^{-1}\mathbf{B}\right)\\
&= \left\{\Gamma_p\left(\frac{f}{2}\right)\det\left(\left[\frac{1}{2}\right]\mathbf{\Sigma}^{-1}\right)^{-f/2}\right\}^{-1}\det(\mathbf{B})^{(1/2)(f-p-1)}\operatorname{etr}\left(-\left[\frac{1}{2}\right]\mathbf{\Sigma}^{-1}\mathbf{B}\right)\\
&= \left\{\Gamma_p\left(\frac{f}{2}\right)\left[\left(\frac{1}{2}\right)^{p}\right]^{-f/2}\det\left(\mathbf{\Sigma}^{-1}\right)^{-f/2}\right\}^{-1}\det(\mathbf{B})^{(1/2)(f-p-1)}\operatorname{etr}\left(-\left[\frac{1}{2}\right]\mathbf{\Sigma}^{-1}\mathbf{B}\right)\\
&= \left\{\Gamma_p\left(\frac{f}{2}\right)2^{(pf)/2}\det(\mathbf{\Sigma})^{f/2}\right\}^{-1}\det(\mathbf{B})^{(1/2)(f-p-1)}\operatorname{etr}\left(-\left[\frac{1}{2}\right]\mathbf{\Sigma}^{-1}\mathbf{B}\right), \qquad (3.181)
\end{aligned}$$

which is equivalent to the $Wishart(f, \mathbf{\Sigma})$ distribution as shown in (3.152). Stating another way, if $\mathbf{B} \sim W_p(f, \mathbf{\Sigma})$, then $\mathbf{B} \sim G_p\left(\frac{f}{2}, \left(\frac{1}{2}\right)\mathbf{\Sigma}^{-1}\right)$, where $G_p(.)$ is the matrix-variate gamma distribution where the symmetric random matrix is of dimension p. Therefore, for simulating a Wishart random matrix, it would appear we can utilize the gamma distribution. While the Wishart and matrix-variate gamma distributions may be considered analytically equivalent, we do notice some differences when simulations are performed using existing software packages. For example, in the *R* language, the "rWishart" function can be used to simulate Wishart random matrices for a given degrees of freedom and covariance matrix [53]. One way to examaine the *R* source code for this function is to utilize a resource such as *https://svn.r-project.org/R/trunk* which lists *R* source code for a variety of functions and libraries. The *R* source code for the "rWishart" function is shown in Appendix E. As can be seen in the attached syntax, the *R* simulation is based on the chi-squared distribution with integer degrees of freedom. Further review of Appendix E indicates that the Wishart matrices are simulated based on the Bartlett decomposition [54] - [55]. Briefly, this procedure can be described as follows:

Definition 4. *Bartlett decomposition*: Let $\mathbf{W}$ represent a $p \times p$ random Wishart matrix to be simulated from a $W_p(f, \mathbf{\Sigma})$ distribution. Further let $\mathbf{W} = \mathbf{L}\mathbf{B}\mathbf{B}^T\mathbf{L}^T$, where $\mathbf{L}$ is the Cholesky factor of $\mathbf{\Sigma}$ and $\mathbf{B}$ is a triangular (lower) matrix defined as:

- $(\mathbf{B})_{uu} = b_{uu}$, where $b_{uu} \sim \chi_{f-u+1}$, for $1 \leq u \leq p$
- $(\mathbf{B})_{uv} = b_{uv}$, where $b_{uv} \sim \mathcal{N}(0, 1)$, for $1 \leq v < u \leq p$

The Cholesky factor, $\mathbf{L}$, is a lower triangular matrix such that $\mathbf{\Sigma} = \mathbf{L}\mathbf{L}^T$. Therefore, $\mathbf{W} = \mathbf{L}\mathbf{B}\mathbf{B}^T\mathbf{L}^T$ is defined as the Bartlett decomposition.

For the Bartlett decomposition (Definition 4), we note that b_{uu} follows the chi distribution. This distribution is defined as follows:

Definition 5. *Chi distribution*: Let us assume that the random variable x has the following pdf:

$$\begin{aligned} f(x; v) &= \frac{x^{v-1} \exp^{-x^2/2}}{2^{(v/2)-1}\Gamma\left(\frac{v}{2}\right)}, x \geq 0 \\ &= 0, \text{otherwise} \end{aligned}$$

Then the random variable x is assumed to follow the chi distribution.

In definition 5, we note that v refers to the degrees of freedom. Further, it can be shown, that if $x \sim \chi_v$, then $x^2 \sim \chi^2_v$. This result can be derived as follows: $y = x^2 \rightarrow \sqrt{y} = x \rightarrow \frac{\partial x}{\partial y} = \frac{1}{2\sqrt{y}}$. Substituting into the pdf in Definition 5, we now have:

$$\begin{aligned} f(y; v) &= \frac{\left(y^{1/2}\right)^{v-1} \exp^{-\left(y^{1/2}\right)^2/2}}{2^{(v/2)-1}\Gamma\left(\frac{v}{2}\right)} \left(\frac{1}{2y^{1/2}}\right) \\ &= \left(y^{1/2}\right)^{v-2} \frac{\exp^{-y/2}}{2^{v/2}\Gamma\left(\frac{v}{2}\right)} \\ &= \frac{y^{(v/2)-1} \exp^{-y/2}}{2^{v/2}\Gamma\left(\frac{v}{2}\right)}, \end{aligned}$$

which is the pdf for a chi-squared distribution with v degrees of freedom. Therefore, for performing simulations using the Bartlett decomposition as shown in Definition 4, we can utilize the square root of chi-squared random deviates.

We know from introductory statistical coursework that the chi-squared distribution is a special case of the gamma distribution. Because the Wishart distribution can be written as a matrix-variate gamma distribution, we may be interested in exploring situations where the gamma and chi-squared distributions are similar. This can give us evidence for when existing Wishart simulation approaches are relevant, and when alternative methods may be needed. For illustration, let us assume that $\mathbf{A} \sim W_p(f, \mathbf{\Sigma})$, where $\mathbf{A}$ is a symmetric random matrix. Therefore, we know that $f \geq p$ (non-singular case). The elements of $\mathbf{A}$ are sums of squares or sums of cross products. The degrees of freedom parameter, f, refers to the number of elements for each sum reflected in $\mathbf{A}$. We note that f also refers to the number of replicates sampled from the multivariate normal distribution. Therefore, for a symmetric random matrix of dimension p, simulation approaches based on the chi-squared distribution would assume that the matrix diagonal elements are chi-squared random variables with at least p degrees of freedom. For a simulation approach based on the gamma distribution, the matrix diagonal elements are gamma random variables with the location parameter, α, at least equal to $p/2$, and the scale parameter equal to 2. The comparisons between chi-squared and gamma distributions for increasing dimensions of the covariance matrix are shown in the Figure 2.

As the dimension of the covariance matrix increases, the difference between the chi-squared distribution using integer degrees of freedom and the gamma distribution with fractional degrees of freedom becomes less noticeable. However, when the dimension of the covariance matrix is not large (e.g., < 50), simulation differences do not appear to be negligible. To further evaluate this scenario, we will simulate 3×3 random matrices using 2

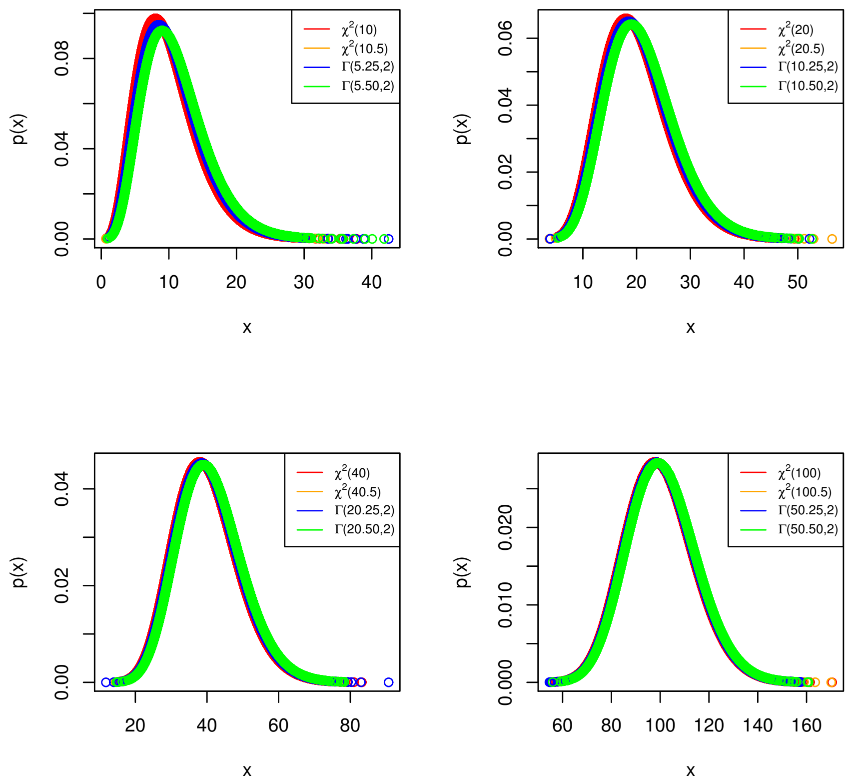

Figure 2: Comparison of Chi-Squared and Gamma Distributions for Increasing Dimensions of the Covariance Matrix

approaches. Let us assume that the matrices are distributed as $W_3\left(4.3, \mathbf{\Sigma}\right)$, where

$$\mathbf{\Sigma} = \begin{pmatrix} 2 & 0.1 & 0.1 \\ 0.1 & 3 & 0.1 \\ 0.1 & 0.1 & 2 \end{pmatrix} \tag{3.182}$$

To examine the impact of fractional degrees of freedom as well as the use of gamma distributions versus chi-squared distributions, we are interested in contrasting the following approaches:

1. Chi-squared distributions with integer degrees of freedom
2. Gamma distributions which allow for fractional degrees of freedom

Because of the similarity between integer and fractional degrees of freedom for the chi-squared distributions in the previous example, we are limiting the multivariate evaluation to the itemized list immediately above. To utilize an approach using integer degrees of freedom for chi-squared random variables, such as the R function "rWishart" shown in Appendix E, we will apply the integer floor function to the fractional degrees of freedom parameter. We will also utilize an approach using gamma distributions with fractional degrees of freedom (2.). For simulating Wishart random matrices using Gamma distributions with fractional degrees of freedom, we have identified two potential existing computer-based options using R and MATLAB. For R, the source is the "rWishart" library released in late 2017. This should not be confused with the "rWishart" R function from the "stats" library mentioned previously. The "rWishart" library includes the "rFractionalWishart" function which is used to generate Wishart random matrices with fractional (potentially) degrees of freedom. We also discovered user-provided source code for MATLAB utilizing gamma distributions to simulate Wishart random matrices with (potentially) fractional degrees of freedom [56]. Because MATLAB is not open-source, this source code is attempting to replicate the MATLAB function "wishrnd.m" which is used to simulate Wishart random matrices. Because of these different implementation methods, we are interested in evaluating if the random matrices generated are, in actuality, Wishart random matrices. The following theorem from Gupta and Nagar (1999) may be useful.

Theorem 1. If the random matrix $\mathbf{W}^{\Psi} \sim W_p\left(f, \mathbf{\Sigma}\right)$,then $\dfrac{\mathbf{c}^T\mathbf{W}^{\Psi}\mathbf{c}}{\mathbf{c}^T\mathbf{\Sigma}\mathbf{c}} \sim \chi^2_{(f)}, \forall \mathbf{c}_{p\times 1} \neq \mathbf{0}$.

For this illustration, we will apply the theorem to the Wishart random matrices from the "rFractionalWishart" function in R. We will utilize the following vectors: $\mathbf{c}_1 = (1,1,1)^T, \mathbf{c}_2 = (1,0,0)^T, \mathbf{c}_3 = (0,1,0)^T$, and $\mathbf{c}_4 = (0,0,1)^T$. Finally, we will define $\mathbf{v}_i = \dfrac{\mathbf{c}_i^T\mathbf{W}^{\Psi}\mathbf{c}_i}{\mathbf{c}_i^T\mathbf{\Sigma}\mathbf{c}_i}, i = 1, \ldots, 4$. For each i, we will generate 10,000 samples from the Wishart distribution. For each $\mathbf{v}_i$, we will calculate the sample mean and variance, as well as Q-Q plots to compare with the chi-squared assumptions from the theorem. The results are shown in the Figure 3: If the "rFractionalWishart" procedure is truly generating random

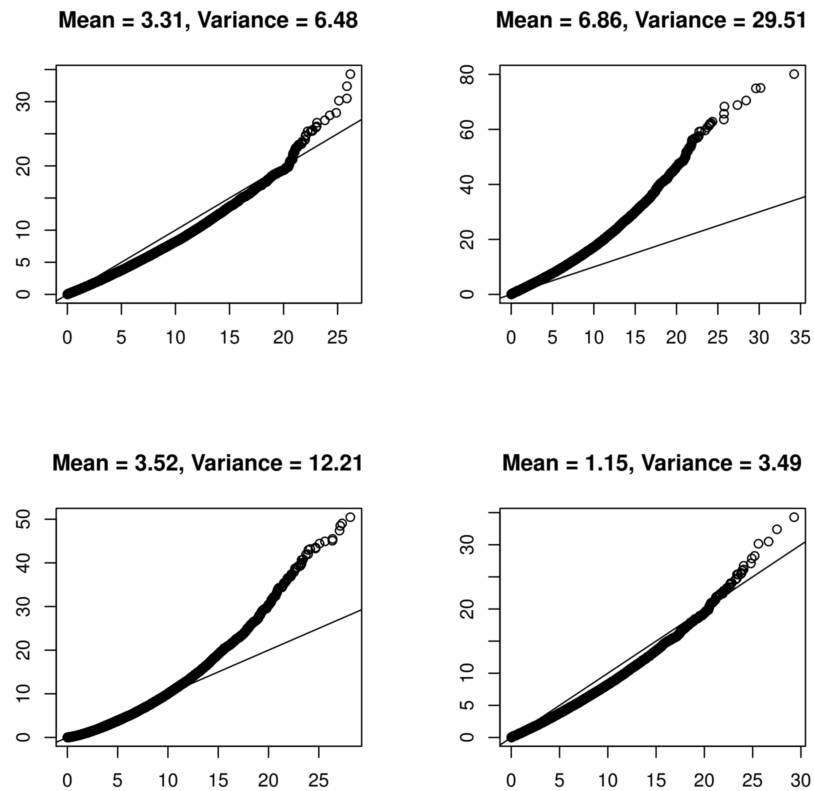

Figure 3: Q-Q plots with Sample Mean and Sample Variance: rFractionalWishart

matrices having a Wishart distribution, we would expect the sample mean to be equal to the degrees of freedom and the sample variance to be equal to twice the degrees of freedom. Thus, for this example we would expect the sample mean to be approximately 4.3 and the sample variance to be approximately 8.6. This is clearly not the case and the Q-Q plots clearly indicate that the calculated values of the test statistic to not appear to follow the chi-squared distribution. Therefore, the "rFractionalWishart"procedure in R should not be used to generate Wishart random matrices.

For the purposes of this dissertation, we will generate Wishart random matrices using a modification of the Bartlett decomposition which utilizes gamma distributions with fractional degrees of freedom. This method is defined as follows:

Definition 6. *Bartlett decomposition - II*: Let $\mathbf{W}^*$ represent a $p \times p$ random Wishart matrix to be simulated from a $W_p(f^*, \mathbf{\Sigma})$ distribution, where f^* may be fractional. Further let $\mathbf{W}^* = \mathbf{L}\mathbf{G}\mathbf{G}^T\mathbf{L}^T$, where $\mathbf{L}$ is the Cholesky factor of $\mathbf{\Sigma}$ and $\mathbf{G}$ is a triangular (lower) matrix defined as:

- $(\mathbf{G})_{uu} = g_{uu}$, where $g_{uu} \sim \textit{Generalized Gamma}(f^* - u + 1, \sqrt{2}, 2)$ and $g_{uu}^2 \sim \textit{Gamma}([f^* - u + 1]/2, 2)$, for $1 \leq u \leq p$
- $(\mathbf{G})_{uv} = g_{uv}$, where $g_{uv} \sim \mathcal{N}(0, 1)$, for $1 \leq v < u \leq p$

The Cholesky factor, $\mathbf{L}$, is a lower triangular matrix such that $\mathbf{\Sigma} = \mathbf{L}\mathbf{L}^T$. Therefore, $\mathbf{W}^* = \mathbf{L}\mathbf{G}\mathbf{G}^T\mathbf{L}^T$ is defined as the Bartlett decomposition - II.

For the Bartlett decomposition - II (Definition 6), we note that g_{uu} follows the generalized gamma distribution. This distribution is defined as follows:

Definition 7. *Generalized Gamma distribution*: Let us assume that the random variable x has the following pdf:

$$
\begin{aligned}
f(x; \alpha, \beta, \gamma) &= \frac{(\gamma/\beta^\alpha)\, x^{\alpha-1} \exp^{-(x/\beta)^\gamma}}{\Gamma\left(\dfrac{\alpha}{\gamma}\right)}, x \geq 0 \\
&= 0, \text{otherwise}
\end{aligned}
$$

Then the random variable x is assumed to follow the generalized gamma distribution.

It can be shown that if $x \sim \textit{Generalized Gamma}(\alpha, \beta, \gamma)$, then $x^2 \sim \textit{Gamma}(\alpha, \beta)$ under certain conditions when $\gamma = 2$. This result can be derived as follows: $y = x^2 \rightarrow \sqrt{y} =$

$x \rightarrow \dfrac{\partial x}{\partial y} = \dfrac{1}{2\sqrt{y}}$. Substituting into the pdf in Definition 7, we now have:

$$
\begin{aligned}
f(y;\alpha,\beta,\gamma) &= \frac{(\gamma/\beta^{\alpha})\left(y^{1/2}\right)^{\alpha-1}\exp^{-\left(y^{1/2}/\beta\right)^{\gamma}}}{\Gamma\left(\dfrac{\alpha}{\gamma}\right)}\left(\frac{1}{2y^{1/2}}\right) \\
&= \frac{\gamma\left(y^{1/2}\right)^{\alpha-2}\exp\left(-y^{1/2}/\beta\right)^{\gamma}}{2\beta^{\alpha}\Gamma\left(\dfrac{\alpha}{\gamma}\right)} \\
&= \frac{\gamma y^{(\alpha/2)-1}\exp\left(-y^{1/2}/\beta\right)^{\gamma}}{2\beta^{\alpha}\Gamma\left(\dfrac{\alpha}{\gamma}\right)}
\end{aligned}
$$

We note that if we let $\gamma = 2$ and $\beta = \sqrt{2}$, then $y \sim \Gamma(\alpha/2, 2)$. Therefore, for performing simulations using the Bartlett decomposition - II as shown in Definition 7, we can utilize the square root of gamma random deviates. As we did previously for the "'rFractionalWishart" function in R, we also can test the Wishart assumption; does the method actually generate Wishart random matrices? Q-Q plots with annotated sample mean and sample variance values are shown in Figure 4: Based on the values for the sample mean and sample variance shown in the figure, we would conclude that the assumption that the random matrices follow a Wishart distribution appears to be reasonable. In addition, the Q-Q plots would seem to indicate that the Wishart distribution appears to be reasonable; the test statistics do appear to follow a chi-squared distribution. So, using the Bartlett decomposition with gamma random deviates appears to be reasonable and will be utilized going forward.

As a side note, examination of the source code for the "rFractionalWishart" function in R identified the reason for the generated random matrices not following the Wishart distribution. The syntax does utilize the Bartlett decomposition - II, but the specification of the location parameter from the gamma distribution was incorrect. That is, the "rFractionalWishart" function uses $degrees\ of\ freedom - i + 1/2$ as the location parameter, and the correct specification is $(degrees\ of\ freedom - i + 1)/2$. In addition, for the generation of random matrices, we would like to compare differences between chi-squared distributions

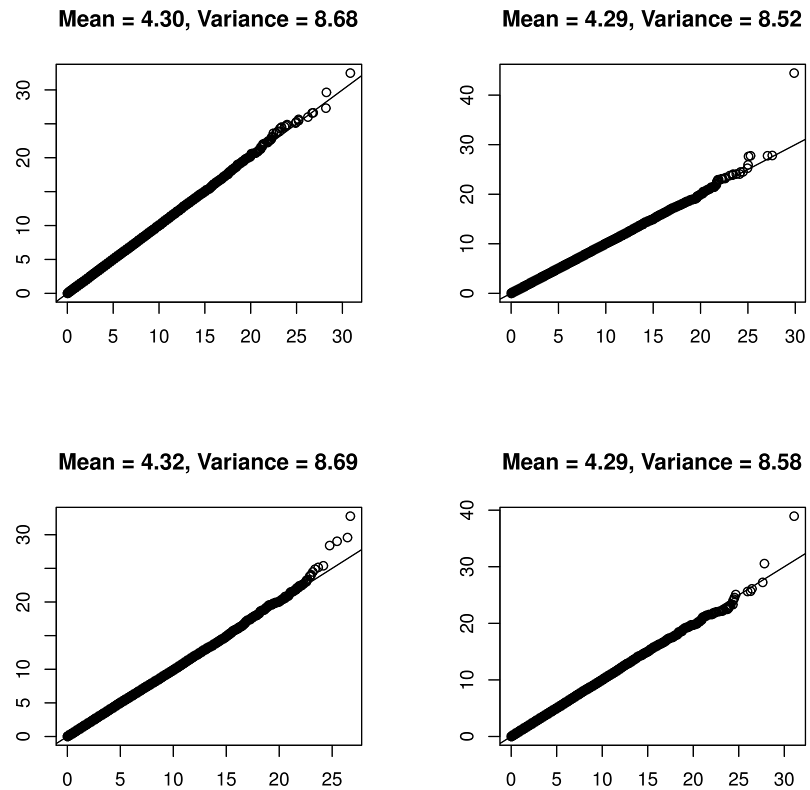

Figure 4: Q-Q plots with Sample Mean and Sample Variance - Bartlett Decomposition - II

with integer degrees of freedom and gamma distributions with fractional degrees of freedom for varying degrees of freedom. Figures 5-7 illustrate these comparisons for increasing fractional degrees of freedom: Similarly to what was seen in the univariate setting, as the degrees of freedom increase, the difference between using integer versus fractional degrees of freedom becomes negligible using the Frobenius norm criteria, differences remain using either the spectral norm or 1-norm criteria. Interestingly, the results for the spectral norm appear to be inconsistent; this criteria did better with the smallest value of degrees of freedom considered (4.3) than it did for larger values of the degrees of freedom parameter (20.3, 50.3). Overall, these results would indicate that depending on the evaluation criteria used,

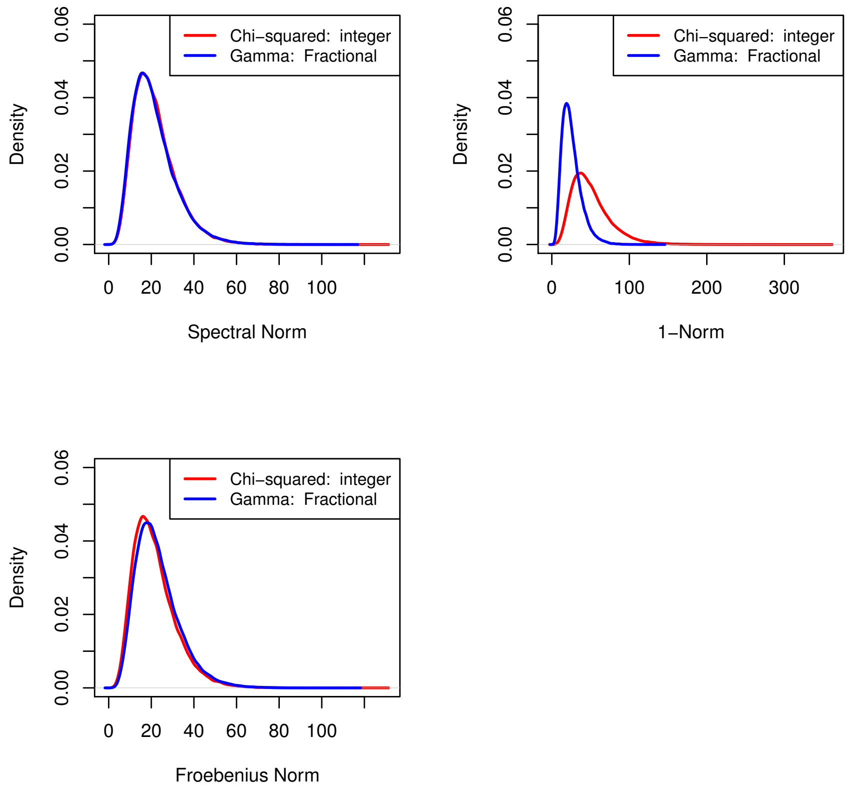

Figure 5: Comparison of Matrix Norms for Random Matrix Generation for Chi-Squared (Integer Degrees of Freedom) and Gamma (Fractional Degrees of Freedom) Assumptions: Degrees of freedom = 4.3

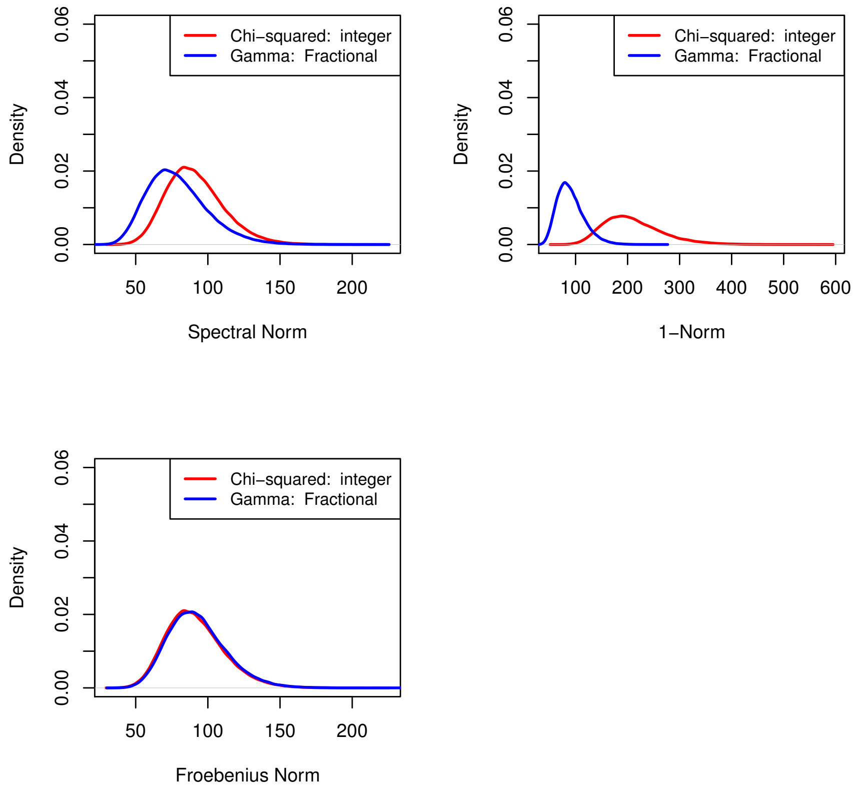

Figure 6: Comparison of Matrix Norms for Random Matrix Generation for Chi-Squared (Integer Degrees of Freedom) and Gamma (Fractional Degrees of Freedom) Assumptions: Degrees of freedom = 20.3

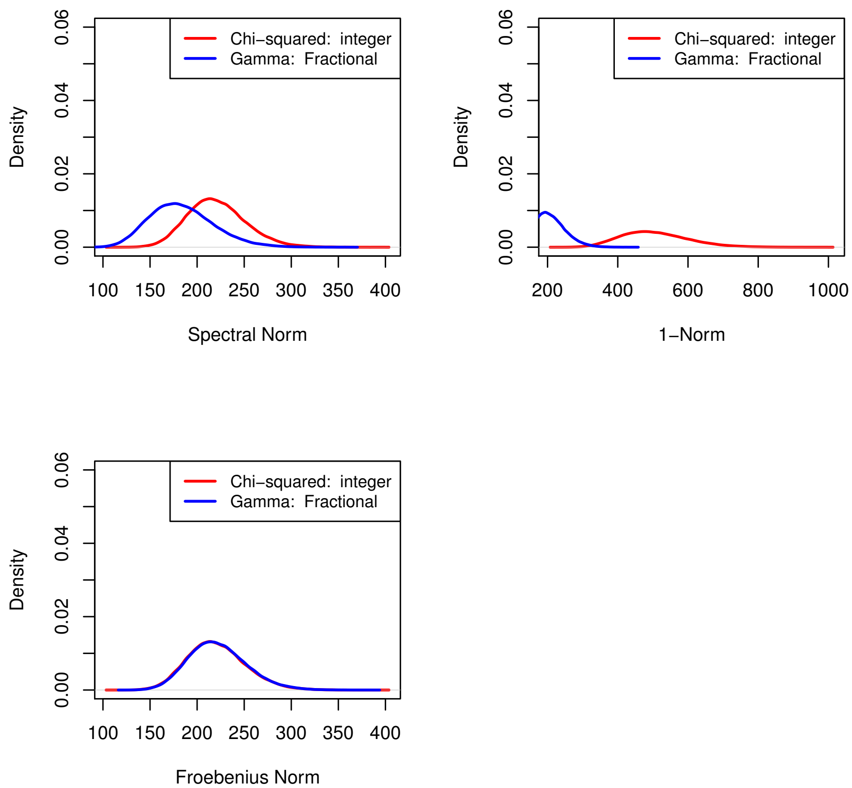

Figure 7: Comparison of Matrix Norms for Random Matrix Generation for Chi-Squared (Integer Degrees of Freedom) and Gamma (Fractional Degrees of Freedom) Assumptions: Degrees of freedom = 50.3

there may be an impact on simulations using that utilize fractional degrees of freedom versus those that assume all degrees of freedom are integer valued.

3.2.3 Simulation Results

For the multivariate simulation, we generated data from a 2-component mixture of Wisharts distribution. The covariance matrices for each component Wishart distribution were of dimension 5 with 7 degrees of freedom. Three separate covariance structures were considered for each estimation method: unstructured, Toeplitz, and banded. The matrices were generated as follows:

- Unstructured

$$\mathbf{\Sigma}_1 = \begin{pmatrix} 1.0 & 0.1 & 0.1 & 0.1 & 0.1 \\ 0.1 & 6.26 & 0.26 & 0.26 & 0.26 \\ 0.1 & 0.26 & 9.02 & 0.32 & 0.32 \\ 0.1 & 0.26 & 0.32 & 12.28 & 0.38 \\ 0.1 & 0.26 & 0.32 & 0.38 & 16.04 \end{pmatrix}$$

$$\mathbf{\Sigma}_2 = \begin{pmatrix} 1.05 & 0.1 & 0.1 & 0.1 & 0.1 \\ 0.1 & 6.35 & 0.26 & 0.26 & 0.26 \\ 0.1 & 0.26 & 8.9 & 0.32 & 0.32 \\ 0.1 & 0.26 & 0.32 & 12.35 & 0.38 \\ 0.1 & 0.26 & 0.32 & 0.38 & 16.2 \end{pmatrix}$$

- Toeplitz

$$\mathbf{\Sigma}_1 = \begin{pmatrix} 1.0 & 0.1 & 0.26 & 0.32 & 0.40 \\ 0.1 & 1.0 & 0.1 & 0.26 & 0.32 \\ 0.26 & 0.1 & 1.0 & 0.1 & 0.26 \\ 0.32 & 0.26 & 0.1 & 1.0 & 0.1 \\ 0.40 & 0.32 & 0.26 & 0.1 & 1.0 \end{pmatrix}$$

$$\mathbf{\Sigma}_2 = \begin{pmatrix} 1.15 & 0.1 & 0.26 & 0.32 & 0.40 \\ 0.1 & 1.15 & 0.1 & 0.26 & 0.32 \\ 0.26 & 0.1 & 1.15 & 0.1 & 0.26 \\ 0.32 & 0.26 & 0.1 & 1.15 & 0.1 \\ 0.40 & 0.32 & 0.26 & 0.1 & 1.15 \end{pmatrix}$$

- Banded

$$\mathbf{\Sigma}_1 = \begin{pmatrix} 1.0 & 0.1 & 0.0 & 0.0 & 0.0 \\ 0.1 & 6.26 & 0.26 & 0.0 & 0.0 \\ 0.0 & 0.26 & 9.02 & 0.32 & 0.0 \\ 0.0 & 0.0 & 0.32 & 12.28 & 0.38 \\ 0.0 & 0.0 & 0.0 & 0.38 & 16.04 \end{pmatrix}$$

$$\mathbf{\Sigma}_2 = \begin{pmatrix} 1.05 & 0.1 & 0.0 & 0.0 & 0.0 \\ 0.1 & 6.35 & 0.26 & 0.0 & 0.0 \\ 0.0 & 0.26 & 8.9 & 0.32 & 0.0 \\ 0.0 & 0.0 & 0.32 & 12.35 & 0.38 \\ 0.0 & 0.0 & 0.0 & 0.38 & 16.2 \end{pmatrix}$$

The mixing proportions were 0.60 and 0.40, respectively. Comparisons between the mixture distribution and the approximating distribution were made using the following matrix norms: 1-norm, Froebenius norm, and spectral norm. The number of replicates for the various matrix norms were 10,000. For all methods, the approximate Wishart distribution's scale matrix was calculated as in (3.155); the methods differ only in their estimation of the degrees of freedom parameter.

In Figures 8-16, matrix norms are compared between the mixture distribution and the applicable approximation method for each type of covariance structure. In addition to the visual comparison, we would like to quantitatively compare the matrix norms by calculating the average squared difference between the matrix norm from the mixture distribution and the matrix norm from the approximating distribution. The results of these calculations are shown in Tables 2-4. Based on a review of the plots in Figures 8-16 and Tables 2-4, we note that the trace-based criteria appears to perform the best in terms of lowest average squared error. Further, for the trace-based criteria, the banded covariance structure appeared to perform the best in terms of average squared error, followed by the Toeplitz covariance structure, and finally the unstructured covariance. It is worth noting that the same pattern was not seen for the element-by-element approach; in terms of average squared error the banded covariance structure performed the best followed by the unstructured covariance, and then the Toeplitz covariance structure. In terms of the matrix norms utilized for this simulation, the 1-norm appeared to perform the worst and the Frobenius norm appears to perform the best. Because the 1-norm criteria is based on maximum column (row) sums, this approach may perform better when the dimension of the covariance matrix is much larger than was evaluated in this dissertation.

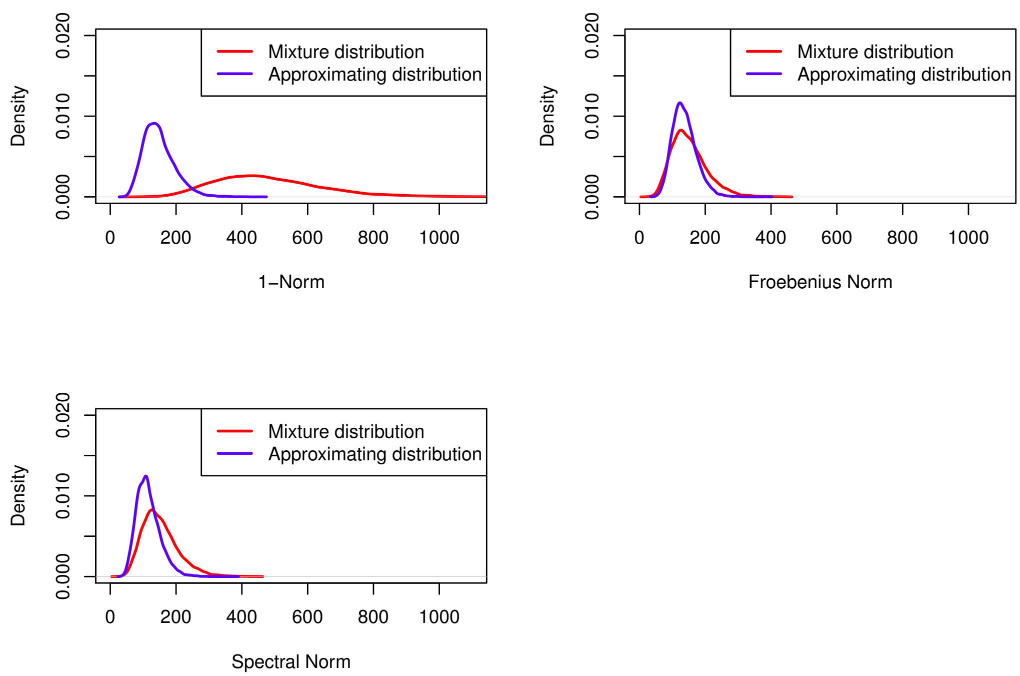

Figure 8: Comparison of Matrix Norms for the Determinant-based Estimation Method: Unstructured Covariance Matrix

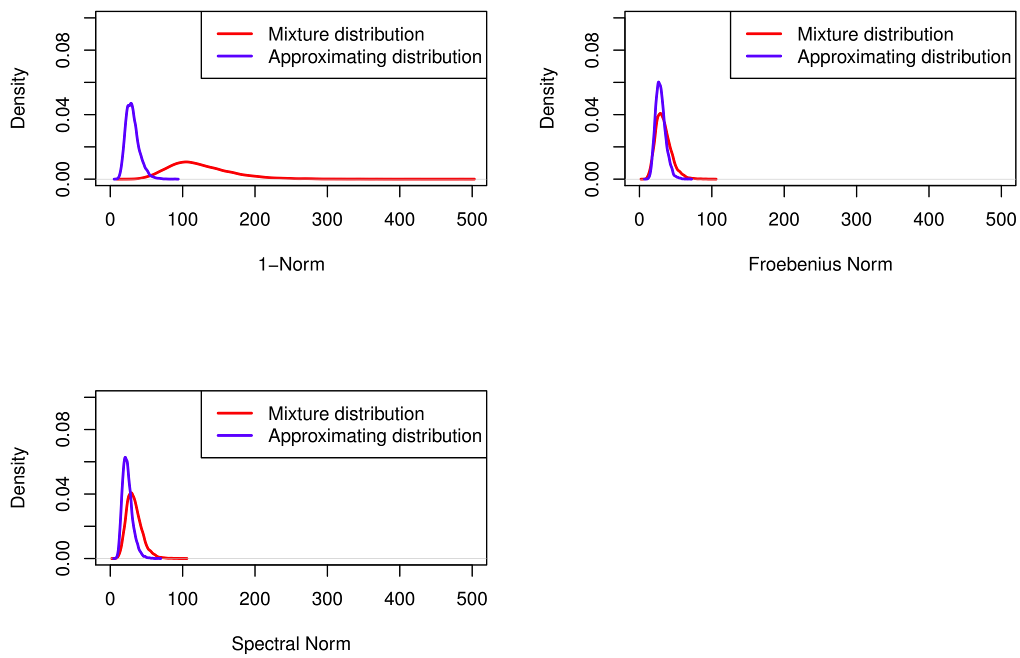

Figure 9: Comparison of Matrix Norms for the Determinant-based Estimation Method: Toeplitz Covariance Matrix

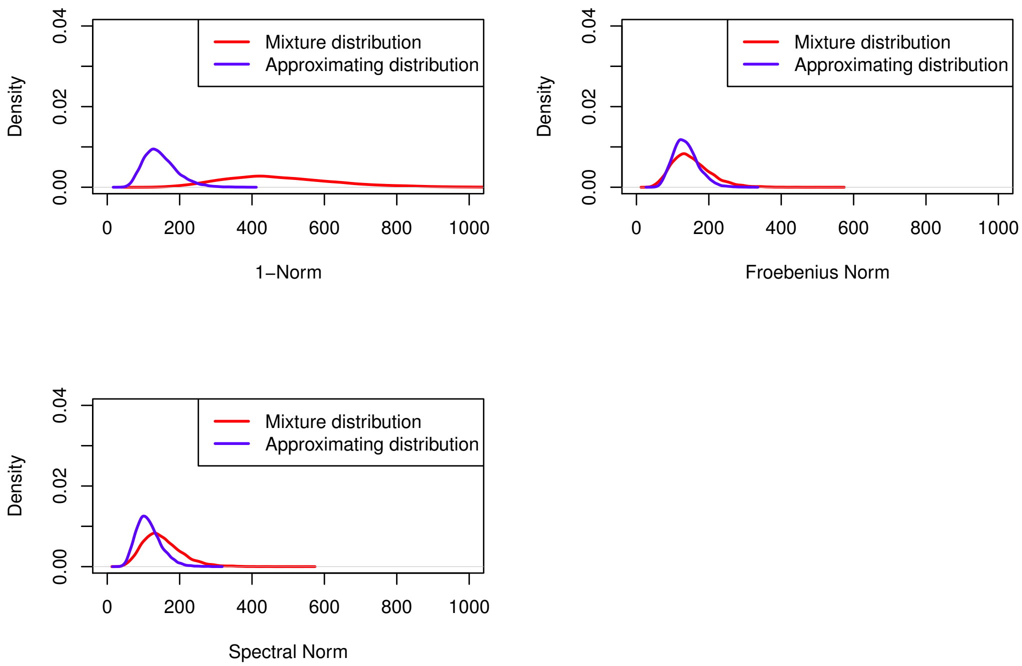

Figure 10: Comparison of Matrix Norms for the Determinant-based Estimation Method: Banded Covariance Matrix

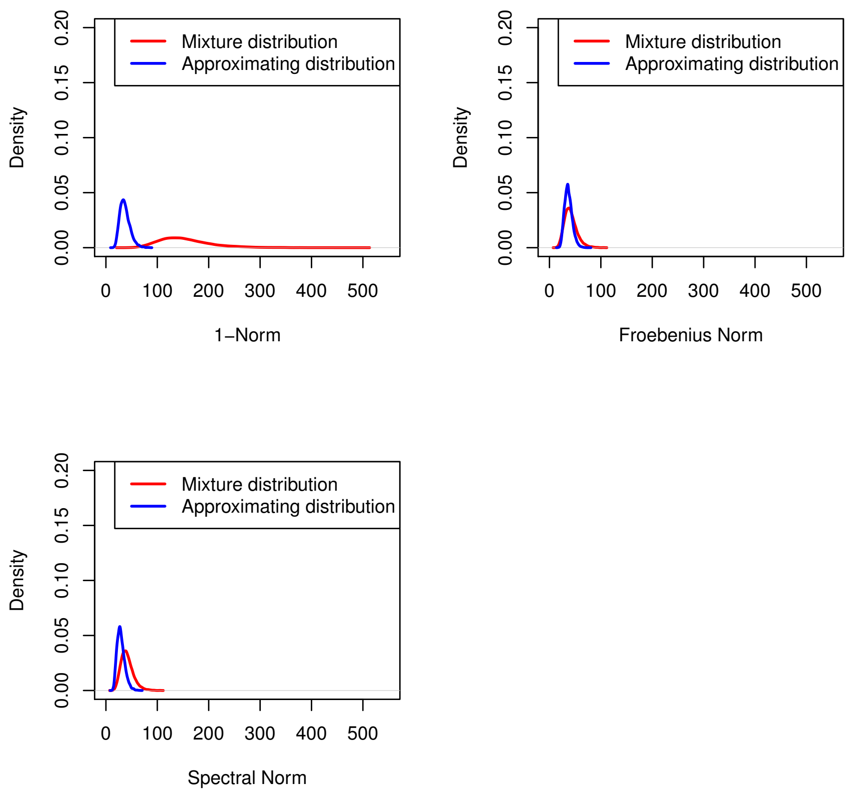

Figure 11: Comparison of Matrix Norms for the Trace-based Estimation Method: Unstructured Covariance Matrix

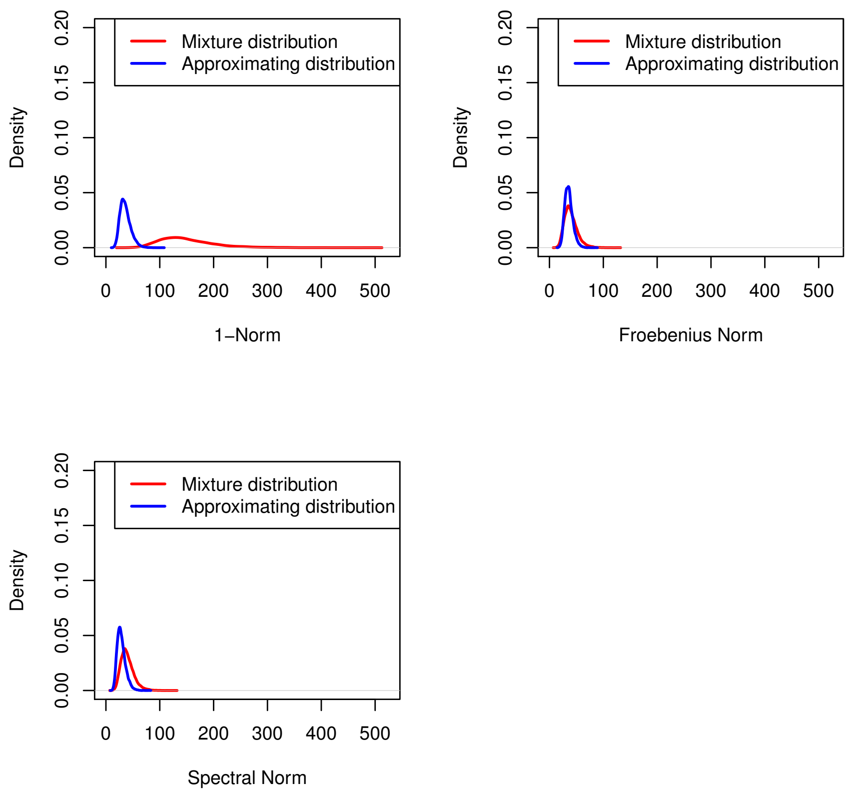

Figure 12: Comparison of Matrix Norms for the Trace-based Estimation Method: Toeplitz Covariance Matrix

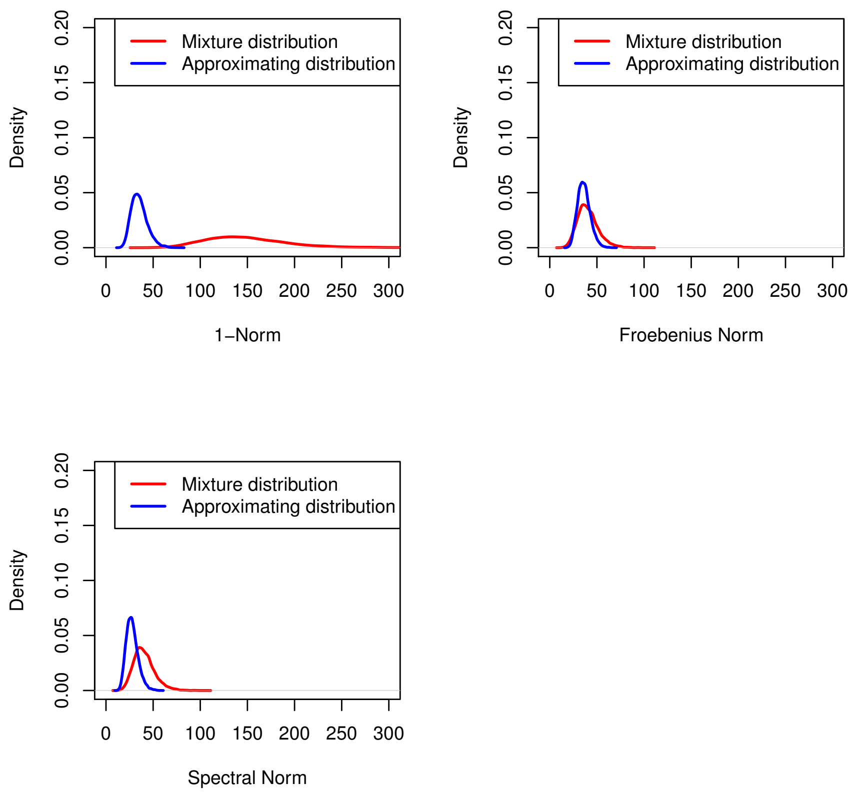

Figure 13: Comparison of Matrix Norms for the Trace-based Estimation Method: Banded Covariance Matrix

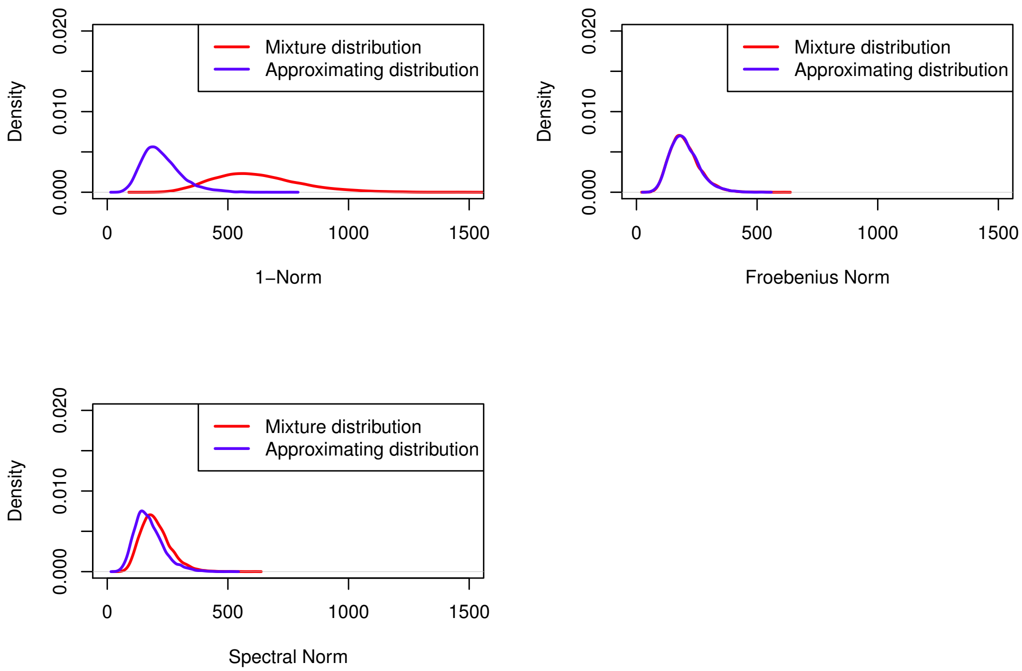

Figure 14: Comparison of Matrix Norms for the Matrix Element-based Estimation Method: Unstructured Covariance Matrix

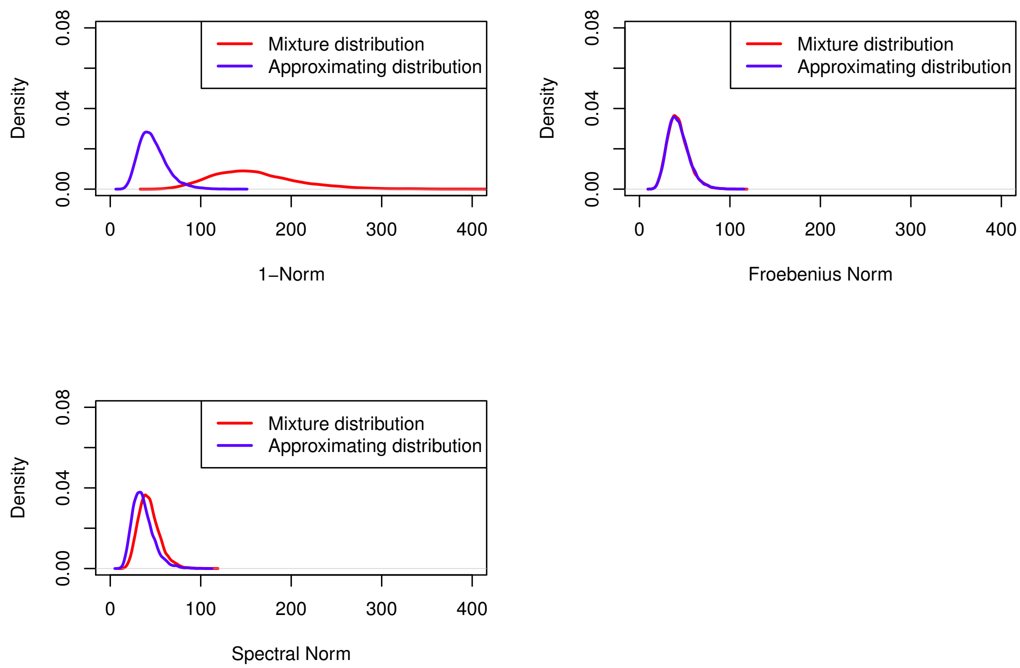

Figure 15: Comparison of Matrix Norms for the Matrix Element-based Estimation Method: Toeplitz Covariance Matrix

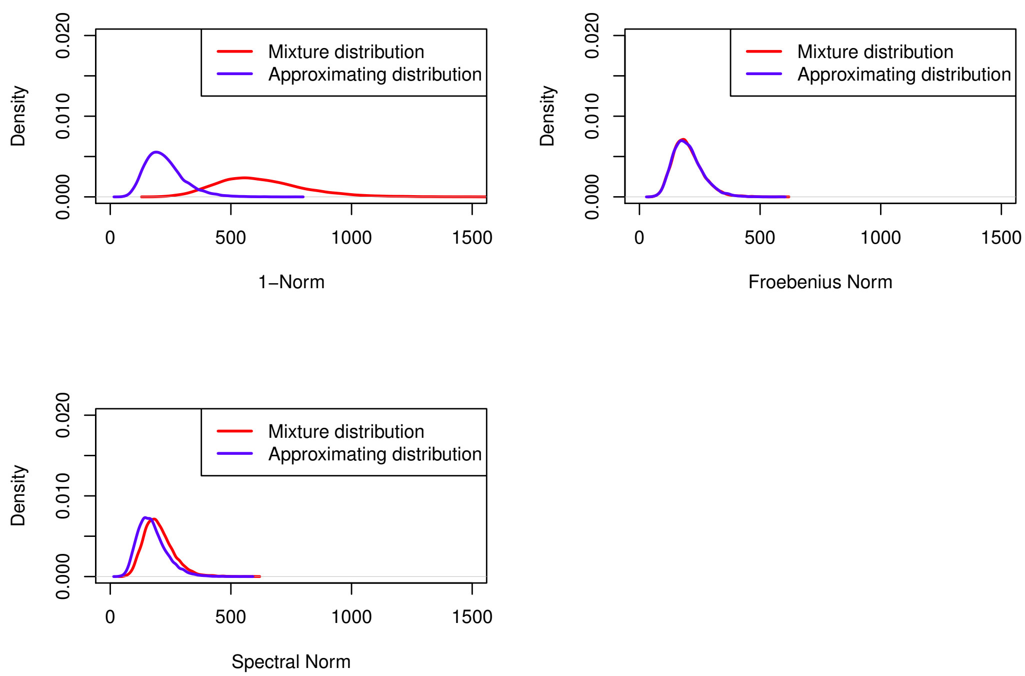

Figure 16: Comparison of Matrix Norms for the Matrix Element-based Estimation Method: Banded Covariance Matrix

Table 2: Unstructured Covariance - Comparison of Average Squared Error Between Mixture Distribution Matrix Norms and Approximating Distribution Matrix Norms

Estimator	Average Squared Error
Determinant:	
1-norm	226,086.60
Frobenius norm	5,828.00
Spectral norm	8,038.02
Trace:	
1-norm	15,605.46
Frobenius norm	214.64
Spectral norm	333.43
Matrix Element-by-Element:	
1-norm	13,481.11
Frobenius	269.15
Spectral norm	300.71

Table 3: Toeplitz Covariance - Comparison of Average Squared Error Between Mixture Distribution Matrix Norms and Approximating Distribution Matrix Norms

Estimator	Average Squared Error
Determinant:	
1-norm	14,999.06
Frobenius norm	234.87
Spectral norm	337.82
Trace:	
1-norm	14,530.12
Frobenius norm	205.95
Spectral norm	311.53
Matrix Element-by-Element:	
1-norm	15,108.50
Frobenius	281.31
Spectral norm	318.13

Table 4: Banded Covariance - Comparison of Average Squared Error Between Mixture Distribution Matrix Norms and Approximating Distribution Matrix Norms

Estimator	Average Squared Error
Determinant:	
1-norm	225,741.60
Frobenius norm	5,607.88
Spectral norm	7,768.13
Trace:	
1-norm	14,169.01
Frobenius norm	178.23
Spectral norm	305.13
Matrix Element-by-Element:	
1-norm	11,953.74
Frobenius norm	225.26
Spectral norm	256.91

4.0 FUTURE DIRECTIONS

4.1 PROPORTIONATE CONTRIBUTION OF EIGENVALUES

When the dimension of a problem is somewhat large, we often seek a way to reduce the dimensionality of the analysis by eliminating variables that do not contribute significantly in a statistical sense. One way this can be accomplished is by examining the eigenvalues of the sample covariance matrix, and determining which eigenvalues are greater than some pre-determined threshold. However, this determination is normally made on an ad hoc basis (e.g., scree plot). A more formal determination could be accomplished using a hypothesis test for the proportionate contribution of a set of eigenvalues:

$$H_0 : \Psi = \frac{\lambda_1 + \cdots + \lambda_s}{\lambda_1 + \cdots + \lambda_p}, \tag{4.1}$$

where $\lambda_1, \ldots, \lambda_p$ denote the population eigenvalues, Ψ is less than some pre-determined threshold, and $s \leq p$. Under the assumption that the sample covariance matrix is distributed as $\sim W_p\left(g, \mathbf{\Sigma}\right)$, Mardia (1979) demonstrated that $\hat{\Psi}$ is asymptotically normal. However, this is based on the assumption of the joint distribution of the random variables following a multivariate Gaussian distribution. The work in this dissertation could possibly be used to extend the proportionate contribution of eigenvalues hypothesis test to scenarios where the joint distribution of the random variables is a mixture of multivariate Gaussian distributions, through the use of an approximating $\sim W_p\left(g, \mathbf{\Sigma}\right)$ distribution. It should be noted that his approach is most applicable when the data is not considered to be high-dimensional ($p < n$).

When the data is considered to be high-dimensional ($p > n$), we could appeal to the use of a singular Wishart distribution as proposed by others [58]-[59]. Through the use of an approximating singular Wishart distribution, it may be possible to extend the proportionate contribution of eigenvalues hypothesis test to the high-dimensional setting when ad hoc methods may be more challenging due to the increased dimensionality of the problem.

APPENDIX A

DERIVING THE DISTRIBUTION OF THE SAMPLE VARIANCE BASED ON A K-COMPONENT FINITE MIXTURE OF $\mathcal{N}(0, \sigma_J^2)$ DISTRIBUTIONS

From (3.30), we have the following MGF:

$$M_{y_\ell^2}(t) = \sum_{j=1}^{k} w_j \left(1 - 2\sigma_j^2 t\right)^{-1/2}, t < \frac{1}{2\sigma_j^2}$$

Because $y_\ell^2 > 0$, we know that the PDF corresponding to the MGF will be equal to 0 when $y_\ell^2 < 0$. Therefore,

$$M_{y_\ell^2}(-t) = \int_0^\infty e^{-ty_\ell^2} g\left(y_\ell^2\right) dy_\ell^2 \tag{A.1}$$

$$= L[g\left(y_\ell^2\right)], \tag{A.2}$$

where $L[\cdot]$ is the Laplace transform. For simplicity, let $y_\ell^2 = s$. Therefore, $g\left(y_\ell^2\right) = g(s)$ can be calculated as:

$$g(s) = L^{-1}[M_s(-t)] \tag{A.3}$$

$$= \frac{1}{2\pi i} \int_{a-i\infty}^{a+i\infty} e^{ts} M_s(-t)\, dt, \tag{A.4}$$

where $i = \sqrt{-1}$ and $L^{-1}[\cdot]$ is the inverse Laplace transform [60]. Thus, $g(s)$ can be written as:

$$g(s) = \frac{1}{2\pi i} \int_{a-i\infty}^{a+i\infty} e^{ts} \sum_{j=1}^{k} w_j \left(1 + 2\sigma_j^2 t\right)^{-1/2} dt \tag{A.5}$$

Because k is finite, we can re-write (A.5) as:

$$g(s) = \frac{1}{2\pi} \sum_{j=1}^{k} w_j \int_{a-i\infty}^{a+i\infty} e^{ts} \left(1 + 2\sigma_j^2 t\right)^{-1/2} dt \tag{A.6}$$

For evaluation of the integral in (A.6), we can appeal to the field of complex analysis. Before proceeding further, some definitions and theorems from complex analysis will be presented [61]. For what follows, z will represent a complex variable, $z = c + id$, where c and d are real and $i = \sqrt{-1}$.

Definition 8. *Multiple-valued functions.* Let $f(z)$ be a function of the complex variable z. If only one value of $f(z)$ corresponds to each value of z, then $f(z)$ is a single-valued function of z. If more than one value of $f(z)$ corresponds to each value of z, then $f(z)$ is a multiple-valued function of z. A multiple-valued function may also be thought of as a set of single-valued functions, each known as a branch of the function.

Definition 9. *Derivative.* If $f(z)$ is single-valued in some region $\mathcal{R}$ of the z plane, the derivative of $f(z)$ is defined as

$$f'(z) = \lim_{\triangle z \to 0} \frac{f(z + \triangle z) - f(z)}{\triangle z} \tag{A.7}$$

provided that the limit exists independent of the way in which $\triangle z \to 0$.

Definition 10. *Analytic Function.* If $f'(z)$ exists at all points z of a region, $\mathcal{R}$, then $f(z)$ is said to be analytic in $\mathcal{R}$ and is consequently known as an analytic function in $\mathcal{R}$.

Definition 11. *Singular points.* A point at which $f(z)$ fails to be analytic is known as a singular point or singularity of $f(z)$. Types of singularities are as follows:

1. *Isolated singularities.* The point $z = z_0$ is called an isolated singularity of $f(z)$ if a $\delta > 0$ can be found such that the circle $|z - z_0| = \delta$ encloses no other singular point other than z_0.

2. *Poles.* If z_0 is an isolated singularity and a positive integer $\tilde{n}$ can be found such that $\lim_{z \to z_0} (z - z_0)^{\tilde{n}} f(z) = B \neq 0$ then $z = z_0$ is called a pole of order $\tilde{n}$. A simple pole has order 1.

3. *Branch points.* For multiple-valued functions, a branch point is a non-isolated singularity

because a multiple-valued function is not continuous, and thus, not analytic in a deleted neighborhood of a branch point.

4. *Removable singularity.* An isolated singular point, z_0, is a removable singularity of $f(z)$ if $\lim_{z \to z_0} f(z)$ exists.

5. *Essential singularity.* An isolated singularity that is not a pole or a removable singularity is referred to as an essential singularity.

Definition 12. *Curve.* Let $\varphi(v)$ and $\psi(v)$ be real functions of the real variable v assumed continuous in $v_1 \leq v \leq v_2$. The equations $z = c + id \Rightarrow \varphi(v) + i\psi(v) = z(v)$ then define a continuous curve or arc in the z plane joining points $a_1 = z(v_1)$ and $a_2 = z(v_2)$. If $v_1 \neq v_2$ while $z(v_1) = z(v_2)$, the endpoints coincide and the curve is said to be closed. If $\varphi(v)$ and $\psi(v)$ have continuous derivatives in $v_1 \leq v \leq v_2$ the curve is also referred to as a smooth curve or arc. A curve that is composed of a finite number of smooth arcs is called a piecewise smooth curve or contour.

Definition 13. *Simply/Multiply-connected regions.* A region $\mathcal{R}$ is called simply-connected if any simple closed curve which lies in $\mathcal{R}$ can be shrunk to a point without leaving $\mathcal{R}$. If this is not true, then the region $\mathcal{R}$ is multiply-connected.

Theorem 2. *Jordan curve theorem.* A Jordan curve is a closed curve that divides the plane into 2 regions having the curve as a common boundary. The region that is bounded (such that all points of it satisfy $|z| < M$ where M is some positive constant) is the interior of the curve while the other region is known as the exterior of the curve.

Definition 14. *Transversal of a closed path.* The boundary C of a region is said to be transversed in the positive direction if an object traveling in this direction (and perpendicular to the plane) has the region to the left. Thus, $\oint_C f(z)\, dz$ is used to indicate integration of $f(z)$ around the boundary C in the positive direction. In the case of a circular region, the positive direction is the counter-clockwise direction. The integral around C is often called a contour integral.

Theorem 3. *Cauchy-Goursat theorem.* Let $f(z)$ be an analytic function of z in the region $\mathcal{R}$ and on its boundary C. Then

$$\oint_C f(z)\,dz = 0 \tag{A.8}$$

Definition 15. *Contour integration.* When a real integral is challenging to evaluate directly, one approach is to appeal to the methods of contour integration. For this method, the real integral is evaluated in the complex plane by integrating around a suitably chosen contour in the complex plane. The contour is chosen so that it encloses the real valued integral and so that the contour does not include any non-isolated singularity (e.g., branch point). Further, the contour is chosen to be an analytic function.

Based on that brief background, let us evaluate the integral in (A.6) via the techniques of contour integration. To review, we wish to evaluate

$$\int_{a-i\infty}^{a+i\infty} e^{ts}\left(1+2\sigma_j^2 t\right)^{-1/2} dt \tag{A.9}$$

via contour integration. Based on the complex variable z, the contour integral becomes

$$\oint_C \frac{e^{zs}}{\sqrt{1+2\sigma_j^2 z}}dz, \tag{A.10}$$

where C is the particular contour chosen in the complex plane. Because of the square root in the denominator of the integrand in (A.10), we note that the integrand is a multiple-valued function. We also note that a branch point (non-isolated singularity) exists at $z = -1/2\sigma_j^2, \sigma_j^2 > 0$. Therefore, the contour chosen must not include this branch point. No other singularities are identified for this particular function. Let the contour denoted by C in (A.10) be shown in Figure 17. The contour reflected in Figure 17 is also known as a Bromwich contour [62].

Therefore, the previously referenced contour C is represented by the region $ABD-EHJ-KLN-A$ in Figure 1. In this figure, EH and KL actually lie on the real axis but have been separated for visual purposes. Also, HJK is a circle of radius ε and BDE and LNA are arcs from a circle of radius R.

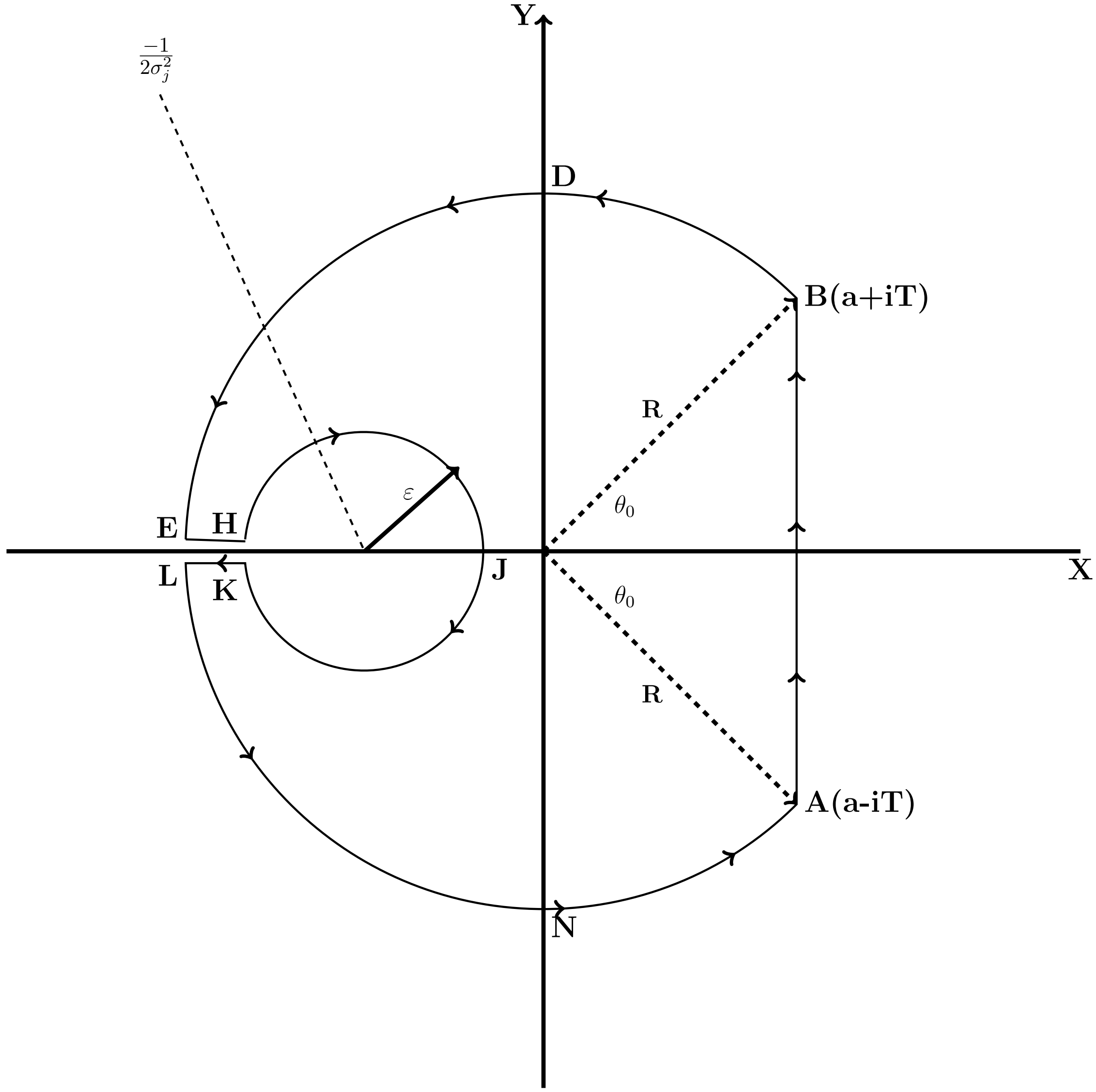

Figure 17: Contour for integral in (A.10)

By the Cauchy-Goursat theorem, we have:

$$\oint_C \frac{e^{zs}}{\sqrt{1+2\sigma_j^2 z}}\,dz = 0 \tag{A.11}$$

By use of the Cauchy-Goursat theorem and by showing that certain sub-integrals of the particular contour will go to zero under certain limiting conditions, the integral in (A.11) can be evaluated directly.

On BD and NA, $z = Re^{i\theta}$, where θ goes from θ_0 to $\pi/2$ and $3\pi/2$ to $2\pi - \theta_0$, respectively. Similarly on DE and LN, $z = Re^{i\theta}$, where θ goes from $\pi/2$ to π and π to $3\pi/2$, respectively. On EH, $\left(1+2\sigma_j^2 z\right) = ue^{\pi i} \Rightarrow \sqrt{1+2\sigma_j^2 z} = \sqrt{u}e^{\pi i/2} = \sqrt{u}[\cos(\pi/2) + i\sin(\pi/2)] = i\sqrt{u}$. On KL, $\left(1+2\sigma_j^2 z\right) = ue^{-\pi i} \Rightarrow \sqrt{1+2\sigma_j^2 z} = \sqrt{u}e^{-\pi i/2} = \sqrt{u}[\cos(-\pi/2) + i\sin(-\pi/2)] = -i\sqrt{u}$.

In both of these cases:

$$\begin{aligned} z &= \frac{-u-1}{2\sigma_j^2} \\ dz &= \frac{-1}{2\sigma_j^2}\,du \end{aligned} \tag{A.12}$$

Along EH, z goes from $-R$ to $\left(\frac{-1}{2\sigma_j^2} - \varepsilon\right)$. Therefore for $z = -R$:

$$\begin{aligned} z &= \frac{-u-1}{2\sigma_j^2} \\ -R &= \frac{-u-1}{2\sigma_j^2} \\ -2\sigma_j^2 R &= -u-1 \\ u &= 2\sigma_j^2 R - 1 \end{aligned} \tag{A.13}$$

Similarly, for $z = \left(\frac{-1}{2\sigma_j^2} - \varepsilon\right)$:

$$\begin{aligned}
z &= \frac{-u-1}{2\sigma_j^2} \\
\frac{-1}{2\sigma_j^2} - \varepsilon &= \frac{-u-1}{2\sigma_j^2} \\
-1 - 2\sigma_j^2\varepsilon &= -u - 1 \\
u &= 2\sigma_j^2\varepsilon
\end{aligned} \tag{A.14}$$

Thus, along EH, u goes from $2\sigma_j^2 R - 1$ to $2\sigma_j^2\varepsilon$ and along KL, u goes from $2\sigma_j^2\varepsilon$ to $2\sigma_j^2 R - 1$. On HJK, $z + (1/2\sigma_j^2) = \varepsilon e^{i\phi}$ where ϕ goes from $-\pi$ to π. For this arc, $z = \varepsilon e^{i\phi} - \frac{1}{2\sigma_j^2} \Rightarrow dz = \varepsilon e^{i\phi} i\, d\phi$. Thus, the contour integral in (A.11) can be written as:

$$\begin{gathered}
\int_{a-iT}^{a+iT} \frac{\exp(zs)}{\sqrt{1+2\sigma_j^2 z}}\, dz + \\
\int_{\theta_0}^{\pi} \frac{\exp\left(sRe^{i\theta}\right)}{\sqrt{1+2\sigma_j^2 Re^{i\theta}}} (Re^{i\theta} i)\, d\theta + \\
\int_{2\sigma_j^2 R-1}^{2\sigma_j^2\varepsilon} \frac{\exp\left(-s\left(\frac{u+1}{2\sigma_j^2}\right)\right)}{i\sqrt{u}} \left(\frac{-1}{2\sigma_j^2}\right) du + \\
\int_{-\pi}^{\pi} \frac{\exp\left(s\left(\varepsilon e^{i\phi} - \frac{1}{2\sigma_j^2}\right)\right)}{\sqrt{1+2\sigma_j^2\left(\varepsilon e^{i\phi} - \frac{1}{2\sigma_j^2}\right)}} \varepsilon e^{i\phi} i\, d\phi + \\
\int_{2\sigma_j^2\varepsilon}^{2\sigma_j^2 R-1} \frac{\exp\left(-s\left(\frac{u+1}{2\sigma_j^2}\right)\right)}{-i\sqrt{u}} \left(\frac{-1}{2\sigma_j^2}\right) du + \\
\int_{\pi}^{2\pi-\theta_0} \frac{\exp\left(sRe^{i\theta}\right)}{\sqrt{1+2\sigma_j^2 Re^{i\theta}}} (Re^{i\theta} i)\, d\theta
\end{gathered} \tag{A.15}$$

Show that:

$$\lim_{R\to\infty} \int_{\theta_0}^{\pi} \frac{\exp\left(sRe^{i\theta}\right)}{\sqrt{1+2\sigma_j^2 Re^{i\theta}}} \left(Re^{i\theta} i\right) d\theta = 0 \tag{A.16}$$

Proof. Let us assume that:

$$\left|\frac{1}{\sqrt{1+2\sigma_j^2 Re^{i\theta}}}\right| \leq \frac{M}{R^m}, \tag{A.17}$$

where $M, m > 0$ and M is an upper bound. Working with the integral in (A.16), we have:

$$\begin{aligned}
ie^{i\theta}\exp\left(sRe^{i\theta}\right) &= ie^{i\theta}\exp\left(Rs[\cos\theta + i\sin\theta]\right) \quad \text{(Euler's formula)}\\
&= ie^{i\theta}\exp\left(Rs\cos\theta + iRs\sin\theta\right)\\
&= i\exp\left(i[\theta + Rs\sin\theta]\right)\exp\left(Rs\cos\theta\right)\\
&= i\exp\left(i[\theta + k^*]\right)\exp\left(Rs\cos\theta\right), \text{ where } k^* = Rs\sin\theta\\
&= i[\cos\left(\theta + k^*\right) + i\sin\left(\theta + k^*\right)]\exp\left(Rs\cos\theta\right)\\
&= [i\cos\left(\theta + k^*\right) - \sin\left(\theta + k^*\right)]\exp\left(Rs\cos\theta\right)
\end{aligned} \tag{A.18}$$

Now returning to the original integral in (A.16), and using properties of absolute value, we have:

$$\left|\int_{\theta_0}^{\pi}\frac{\exp\left(sRe^{i\theta}\right)}{\sqrt{1+2\sigma_j^2 Re^{i\theta}}}\left(Re^{i\theta}i\right)d\theta\right| \leq \int_{\theta_0}^{\pi}\left|\frac{\exp\left(sRe^{i\theta}\right)}{\sqrt{1+2\sigma_j^2 Re^{i\theta}}}(Re^{i\theta}i)\right|d\theta \tag{A.19}$$

Working with the r.h.s of (A.19) and substituting based on (A.18), we have:

$$\begin{aligned}
\int_{\theta_0}^{\pi}\left|\frac{\exp\left(sRe^{i\theta}\right)}{\sqrt{1+2\sigma_j^2 Re^{i\theta}}}(Re^{i\theta}i)\right|d\theta &= \int_{\theta_0}^{\pi}|i\cos\left(\theta + k^*\right) - \sin\left(\theta + k^*\right)|\left|\frac{1}{\sqrt{1+2\sigma_j^2 Re^{i\theta}}}\right| \times\\
&\qquad |\exp\left(Rs\cos\theta\right)|\,|R|d\theta\\
&= \int_{\theta_0}^{\pi}\sqrt{\cos^2\left(\theta + k^*\right) + \sin^2\left(\theta + k^*\right)}\left|\frac{1}{\sqrt{1+2\sigma_j^2 Re^{i\theta}}}\right| \times\\
&\qquad |\exp\left(Rs\cos\theta\right)||R|d\theta\\
&= \int_{\theta_0}^{\pi}\left|\frac{1}{\sqrt{2\sigma_j^2 Re^{i\theta}+1}}\right||\exp\left(Rs\cos\theta\right)||R|d\theta
\end{aligned} \tag{A.20}$$

Working with the r.h.s. of (A.20) and using (A.17), we now have:

$$\int_{\theta_0}^{\pi}\left|\frac{1}{\sqrt{2\sigma_j^2 Re^{i\theta}+1}}\right||\exp\left(Rs\cos\theta\right)||R|d\theta \leq \frac{M}{R^m}\int_{\theta_0}^{\pi}|\exp\left(Rs\cos\theta\right)||R|d\theta \tag{A.21}$$

If we split up the integral on the r.h.s. of (A.21), we have:

$$\begin{aligned}\frac{M}{R^m}\int_{\theta_0}^{\pi}|\exp\left(Rs\cos\theta\right)||R|d\theta &= \frac{M}{R^m}\int_{\theta_0}^{\pi/2}|\exp\left(Rs\cos\theta\right)||R|d\theta + \frac{M}{R^m}\int_{\pi/2}^{\pi}|\exp\left(Rs\cos\theta\right)||R|d\theta \\ &= \frac{M}{R^{m-1}}\int_{\theta_0}^{\pi/2}\exp\left(Rs\cos\theta\right)d\theta + \frac{M}{R^{m-1}}\int_{\pi/2}^{\pi}\exp\left(Rs\cos\theta\right)d\theta\end{aligned} \tag{A.22}$$

Now, using the results from (A.19) - (A.22), we have the following inequality:

$$\left|\int_{\theta_0}^{\pi}\frac{\exp\left(sRe^{i\theta}\right)}{\sqrt{1+2\sigma_j^2Re^{i\theta}}}\left(Re^{i\theta}i\right)d\theta\right| \leq \frac{M}{R^{m-1}}\int_{\theta_0}^{\pi/2}\exp\left(Rs\cos\theta\right)d\theta + \frac{M}{R^{m-1}}\int_{\pi/2}^{\pi}\exp\left(Rs\cos\theta\right)d\theta \tag{A.23}$$

Working with the r.h.s of (A.23), we can evaluate the left integral by the substitution method. Let $\theta = \pi/2 - \phi \Rightarrow d\theta = -d\phi \Rightarrow -d\theta = d\phi$. Substituting:

$$\begin{aligned}\frac{M}{R^{m-1}}\int_{\theta_0}^{\pi/2}\exp\left(Rs\cos\theta\right)d\theta &= \frac{-M}{R^{m-1}}\int_{\pi/2-\theta_0}^{0}\exp\left(Rs\cos\left(\pi/2-\phi\right)\right)d\phi \\ &= \frac{M}{R^{m-1}}\int_{0}^{\pi/2-\theta_0=\phi_0}\exp\left(Rs\sin\phi\right)d\phi\end{aligned} \tag{A.24}$$

From Figure 1, we can establish the following:

$$\begin{aligned}\cos\theta_0 &= a/R \\ \sin\phi_0 &= a/R \\ \phi_0 &= \sin^{-1}\left(a/R\right)\end{aligned}$$

This leads to the following inequality:

$$\sin\phi \leq \sin\phi_0 \leq \cos\theta_0 = a/R \tag{A.25}$$

Using (A.25) we can now establish bounds on the integral on the r.h.s. of (A.24) as follows:

$$\begin{aligned}\frac{M}{R^{m-1}}\int_{0}^{\phi_0}\exp\left(Rs\sin\phi\right)d\phi \leq \frac{M}{R^{m-1}}\int_{0}^{\phi_0}\exp\left(sa\right)d\phi &= \frac{M}{R^{m-1}}\phi_0\exp\left(sa\right) \\ &= \frac{M}{R^{m-1}}\exp\left(sa\right)\sin^{-1}\left(a/R\right)\end{aligned} \tag{A.26}$$

Therefore, as $R \to \infty$, we note that $\dfrac{M}{R^{m-1}} \exp(sa) \sin^{-1}(a/R) \to 0$ because, under $R \to \infty$, $\sin^{-1}(a/R) \approx a/R$. Thus, combining (A.24) - (A.26) we can state:

$$\frac{M}{R^{m-1}} \lim_{R\to\infty} \int_{\theta_0}^{\pi/2} \exp(Rs\cos\theta)\, d\theta = 0 \tag{A.27}$$

Next, to complete the proof, we must show that the right integral on the r.h.s. of (A.23) also goes to 0 as $R \to \infty$. For this part of the proof, we will utilize the following substitution: Let $\theta = \pi/2 + \phi \Rightarrow d\theta = d\phi$. Substituting, we have:

$$\begin{aligned}\frac{M}{R^{m-1}} \int_{\pi/2}^{\pi} \exp(Rs\cos\theta)\, d\theta &= \frac{M}{R^{m-1}} \int_{0}^{\pi/2} \exp(Rs\cos(\pi/2+\phi))\, d\phi \\ &= \frac{M}{R^{m-1}} \int_{0}^{\pi/2} \exp(-Rs\sin\phi)\, d\phi \end{aligned} \tag{A.28}$$

Let $H(\phi) = \sin\phi/\phi = \phi^{-1}\sin\phi$. Taking the first derivative of $H(\phi)$ w.r.t. ϕ, we have:

$$\begin{aligned}\frac{\partial H(\phi)}{\partial \phi} &= \phi^{-1}\cos\phi + \sin\phi(-1)\left(\phi^{-2}\right) \\ &= \frac{\cos\phi}{\phi} - \frac{\sin\phi}{\phi^2} \\ &= \frac{\phi\cos\phi - \sin\phi}{\phi^2}\end{aligned} \tag{A.29}$$

If $G(\phi) = \phi\cos\phi - \sin\phi$, we have the following:

$$\begin{aligned}\frac{\partial G(\phi)}{\partial \phi} &= \phi(-\sin\phi) + \cos\phi(1) - \cos\phi \\ &= -\phi\sin\phi + \cos\phi - \cos\phi \\ &= -\phi\sin\phi\end{aligned} \tag{A.30}$$

For $0 \le \phi \le \pi/2$, $G'(\phi) \le 0$ and is a decreasing function. Because $G(0) = 0$, $G(\phi) \le 0$. Thus, $H'(\phi) \le 0$ or $H(\phi)$ is a decreasing function. Further:

$$\lim_{\phi\to 0} H(\phi) = \lim_{\phi\to 0} \frac{\sin\phi}{\phi} = \frac{\cos\phi|_{\phi=0}}{1} = 1 \tag{A.31}$$

Thus, $H(\phi)$ decreases from 1 to $2/\pi$ as ϕ goes from 0 to $\pi/2$. Then: $1 \geq \sin\phi/\phi \geq 2/\pi \Rightarrow \phi \geq \sin\phi \geq 2\phi/\pi$. Thus, $\sin\phi \geq 2\phi/\pi$. Substituting into the integral on the r.h.s. of (A.28), we have:

$$\frac{M}{R^{m-1}} \int_0^{\pi/2} \exp\left(-Rs\sin\phi\right) d\phi \leq \frac{M}{R^{m-1}} \int_0^{\pi/2} \exp\left(-Rs\left(2\phi/\pi\right)\right) d\phi = \frac{\pi M}{2sR^m}\left(1 - \exp\left(-Rs\right)\right) \tag{A.32}$$

Applying the limit as $R \to \infty$ to the upper bound in (A.33), we have:

$$\lim_{R\to\infty} \frac{\pi M}{2sR^m}\left(1 - \exp\left(-Rs\right)\right) = 0 \tag{A.33}$$

Therefore, combining (A.19), (A.23), (A.27), and (A.33), we now have:

$$\lim_{R\to\infty}\left|\int_{\theta_0}^{\pi} \frac{\exp\left(sRe^{i\theta}\right)}{\sqrt{1+2\sigma_j^2 Re^{i\theta}}}\left(Re^{i\theta} i\right) d\theta\right| \leq \lim_{R\to\infty}\int_{\theta_0}^{\pi}\left|\frac{\exp\left(sRe^{i\theta}\right)}{\sqrt{1+2\sigma_j^2 Re^{i\theta}}}\left(Re^{i\theta} i\right)\right| d\theta = 0 \tag{A.34}$$

Finally, this then leads to the following:

$$\lim_{R\to\infty}\int_{\theta_0}^{\pi} \frac{\exp\left(sRe^{i\theta}\right)}{\sqrt{1+2\sigma_j^2 Re^{i\theta}}}\left(Re^{i\theta} i\right) d\theta = 0 \tag{A.35}$$

This completes the proof. $\square$

Next, show that:

$$\lim_{\varepsilon\to 0}\int_{-\pi}^{\pi} \frac{\exp\left(s\left(\varepsilon e^{i\phi} - \frac{1}{2\sigma_j^2}\right)\right)}{\sqrt{1+2\sigma_j^2\left(\varepsilon e^{i\phi} - \frac{1}{2\sigma_j^2}\right)}} \varepsilon e^{i\phi} i \, d\phi = 0 \tag{A.36}$$

Proof. We can rewrite (A.36) as:

$$\lim_{\varepsilon\to 0}\int_{-\pi}^{\pi}\frac{\exp\left(s\left(\varepsilon e^{i\phi}-\frac{1}{2\sigma_j^2}\right)\right)}{\sqrt{2\sigma_j^2\varepsilon e^{i\phi}}}\varepsilon e^{i\phi}i\ d\phi=\lim_{\varepsilon\to 0}\int_{-\pi}^{\pi}\frac{\exp\left(s\left(\varepsilon e^{i\phi}-\frac{1}{2\sigma_j^2}\right)\right)}{\sqrt{2\sigma_j^2}}\sqrt{\varepsilon e^{i\phi}}i\ d\phi \quad (A.37)$$

If we assume that the limit can be moved inside the integral on the r.h.s. of (A.37), we have:

$$\int_{-\pi}^{\pi}\lim_{\varepsilon\to 0}\frac{\exp\left(s\left(\varepsilon e^{i\phi}-\frac{1}{2\sigma_j^2}\right)\right)}{\sqrt{2\sigma_j^2}}\sqrt{\varepsilon e^{i\phi}}i\ d\phi \quad (A.38)$$

We note that as $\varepsilon\to 0$, the interior limit approaches 0. Because this limit uniformly converges to zero, moving the limit inside the integral is justified via the uniform convergence theorem. Therefore, the desired result has been demonstrated and the proof is complete. □

Finally, we wish to show that

$$\lim_{R\to\infty}\int_{\pi}^{2\pi-\theta_0}\frac{\exp\left(sRe^{i\theta}\right)}{\sqrt{1+2\sigma_j^2Re^{i\theta}}}(Re^{i\theta}i)d\theta \quad (A.39)$$

Proof. Using a similar development employed in (A.19) - (A.21), we have:

$$\left|\int_{\pi}^{2\pi-\theta_0}\frac{\exp\left(sRe^{i\theta}\right)}{\sqrt{1+2\sigma_j^2Re^{i\theta}}}(Re^{i\theta}i)d\theta\right|\le\frac{M}{R^m}\int_{\pi}^{2\pi-\theta_0}|\exp\left(Rs\cos\theta\right)||R|d\theta \quad (A.40)$$

Splitting up the integral on the r.h.s. of (A.40), we have:

$$\begin{aligned}\frac{M}{R^m}\int_{\pi}^{2\pi-\theta_0}|\exp\left(Rs\cos\theta\right)||R|d\theta &= \frac{M}{R^m}\int_{\pi}^{3\pi/2}|\exp\left(Rs\cos\theta\right)||R|d\theta+\\ &\quad\frac{M}{R^m}\int_{3\pi/2}^{2\pi-\theta_0}|\exp\left(Rs\cos\theta\right)||R|d\theta\\ &=\frac{M}{R^{m-1}}\int_{\pi}^{3\pi/2}\exp\left(Rs\cos\theta\right)d\theta+\\ &\quad\frac{M}{R^{m-1}}\int_{3\pi/2}^{2\pi-\theta_0}\exp\left(Rs\cos\theta\right)d\theta \end{aligned}\quad (A.41)$$

Working with the left integral on the r.h.s. of (A.41), let us make the following substitution: $\theta = \pi + \tilde{\phi} \Rightarrow d\theta = d\tilde{\phi}$. Substituting, we have:

$$\frac{M}{R^{m-1}} \int_0^{\pi/2} \exp\left(Rs \cos\left(\pi + \tilde{\phi}\right)\right) d\tilde{\phi} = \frac{M}{R^{m-1}} \int_0^{\pi/2} \exp\left(-Rs \cos\tilde{\phi}\right) d\tilde{\phi} \tag{A.42}$$

For $0 \leq \phi \leq \pi/2$, we know that $\cos\tilde{\phi} \geq 0$. Therefore, using the r.h.s. of (A.42), we now have:

$$\frac{M}{R^{m-1}} \int_0^{\pi/2} \exp\left(-Rs \cos\tilde{\phi}\right) d\tilde{\phi} \leq \frac{M}{R^{m-1}} \int_0^{\pi/2} \exp\left(-Rs(0)\right) d\tilde{\phi} = \frac{M\pi}{2R^{k-1}} \tag{A.43}$$

Because the limit of the upper bound in (A.43) as $R \to \infty$ is equal to 0, we have shown that:

$$\lim_{R\to\infty} \frac{M}{R^{m-1}} \int_\pi^{3\pi/2} \exp\left(Rs\cos\theta\right) d\theta = 0 \tag{A.44}$$

Now let us work with the right integral on the r.h.s. of (A.41). Making the same substitution utilized in (A.42), we now have:

$$\begin{aligned} \frac{M}{R^{m-1}} \int_{3\pi/2}^{2\pi-\theta_0} \exp\left(Rs\cos\theta\right) d\theta &= \frac{M}{R^{m-1}} \int_{\pi/2}^{\pi-\theta_0} \exp\left(Rs\cos\left(\pi+\tilde{\phi}\right)\right) d\tilde{\phi} \\ &= \frac{M}{R^{m-1}} \int_{\pi/2}^{\pi-\theta_0} \exp\left(-Rs\cos\tilde{\phi}\right) d\tilde{\phi} \end{aligned} \tag{A.45}$$

Now, working with the r.h.s. of (A.45), let us consider the additional substitution: $\tilde{\phi} = \pi/2 + \psi \Rightarrow d\tilde{\phi} = d\psi$. Therefore, we now have:

$$\begin{aligned} \frac{M}{R^{m-1}} \int_{\pi/2}^{\pi-\theta_0} \exp\left(-Rs\cos\tilde{\phi}\right) d\tilde{\phi} &= \frac{M}{R^{m-1}} \int_0^{\pi/2-\theta_0} \exp\left(-Rs\cos\left(\pi/2+\psi\right)\right) d\psi \\ &= \frac{M}{R^{m-1}} \int_0^{\pi/2-\theta_0} \exp\left(Rs\sin\psi\right) d\psi \end{aligned} \tag{A.46}$$

It should be noted that, in (A.24) - (A.27), an integral of the same form as that on the r.h.s. of (A.46) was shown to approach 0 as $R \to \infty$. Therefore, combining (A.40) with the results from (A.41) - (A.46), we now have:

$$\lim_{R\to\infty} \left| \int_\pi^{2\pi-\theta_0} \frac{\exp\left(sRe^{b\theta}\right)}{\sqrt{1+2\sigma_j^2 Re^{b\theta}}} (Re^{b\theta} b) d\theta \right| = 0 \tag{A.47}$$

It follows from (A.47) that:

$$\lim_{R\to\infty}\int_{\pi}^{2\pi-\theta_0}\frac{\exp\left(sRe^{i\theta}\right)}{\sqrt{1+2\sigma_j^2Re^{i\theta}}}(Re^{i\theta}i)d\theta=0 \tag{A.48}$$

Therefore, the overall result has been demonstrated and the proof is complete. $\square$

Because of the integrals which have been shown to approach zero based on certain limiting properties ($R\to\infty,\ \varepsilon\to 0$) in (A.33), (A.38), and (A.48), the contour integral in (A.15) can be written as:

$$\begin{aligned}&\int_{a-iT}^{a+iT}\frac{\exp(zs)}{\sqrt{1+2\sigma_j^2z}}dz\ -\\ &\int_{2\sigma_j^2R-1}^{2\sigma_j^2\varepsilon}\frac{\exp\left(-s\left(\frac{u+1}{2\sigma_j^2}\right)\right)}{2\sigma_j^2i\sqrt{u}}du\ -\\ &\int_{2\sigma_j^2\varepsilon}^{2\sigma_j^2R-1}\frac{\exp\left(-s\left(\frac{u+1}{2\sigma_j^2}\right)\right)}{-2\sigma_j^2i\sqrt{u}}du=0\end{aligned} \tag{A.49}$$

By reversing the limits of integration on the second integral in (A.49), we now have:

$$\begin{aligned}&\int_{a-iT}^{a+iT}\frac{\exp(zs)}{\sqrt{1+2\sigma_j^2z}}dz\ +\\ &\int_{2\sigma_j^2\varepsilon}^{2\sigma_j^2R-1}\frac{\exp\left(-s\left(\frac{u+1}{2\sigma_j^2}\right)\right)}{2\sigma_j^2i\sqrt{u}}du\ -\\ &\int_{2\sigma_j^2\varepsilon}^{2\sigma_j^2R-1}\frac{\exp\left(-s\left(\frac{u+1}{2\sigma_j^2}\right)\right)}{-2\sigma_j^2i\sqrt{u}}du=0\end{aligned} \tag{A.50}$$

Rearranging terms in (A.50), we have:

$$\begin{aligned}\int_{a-iT}^{a+iT}\frac{\exp(zs)}{\sqrt{1+2\sigma_j^2z}}dz\ &=-\int_{2\sigma_j^2\varepsilon}^{2\sigma_j^2R-1}\frac{\exp\left(-s\left(\frac{u+1}{2\sigma_j^2}\right)\right)}{2\sigma_j^2i\sqrt{u}}du\ -\\ &\int_{2\sigma_j^2\varepsilon}^{2\sigma_j^2R-1}\frac{\exp\left(-s\left(\frac{u+1}{2\sigma_j^2}\right)\right)}{2\sigma_j^2i\sqrt{u}}du\end{aligned} \tag{A.51}$$

Combining terms on the r.h.s. of (A.51), we have:

$$\int_{a-iT}^{a+iT} \frac{\exp(zs)}{\sqrt{1+2\sigma_j^2 z}} dz = -\frac{2}{i} \int_{2\sigma_j^2 \varepsilon}^{2\sigma_j^2 R-1} \frac{\exp\left(-s\left(\frac{u+1}{2\sigma_j^2}\right)\right)}{2\sigma_j^2 \sqrt{u}} du \tag{A.52}$$

Multiplying the numerator and denominator of the r.h.s. of (A.52) by $i = \sqrt{-1}$, we have:

$$\int_{a-bT}^{a+bT} \frac{\exp(zs)}{\sqrt{1+2\sigma_j^2 z}} dz = -\frac{2i}{i^2} \int_{2\sigma_j^2 \varepsilon}^{2\sigma_j^2 R-1} \frac{\exp\left(-s\left(\frac{u+1}{2\sigma_j^2}\right)\right)}{2\sigma_j^2 \sqrt{u}} du \tag{A.53}$$

Simplifying the r.h.s. of (A.53), we have:

$$\int_{a-iT}^{a+iT} \frac{\exp(zs)}{\sqrt{1+2\sigma_j^2 z}} dz = 2i \int_{2\sigma_j^2 \varepsilon}^{2\sigma_j^2 R-1} \frac{\exp\left(-s\left(\frac{u+1}{2\sigma_j^2}\right)\right)}{2\sigma_j^2 \sqrt{u}} du \tag{A.54}$$

Because the original integral from (A.6) is scaled by the factor $1/2\pi i$, multiplying both sides of (A.54) by this factor leads to:

$$\frac{1}{2\pi i} \int_{a-iT}^{a+iT} \frac{\exp(zs)}{\sqrt{1+2\sigma_j^2 z}} dz = \frac{1}{\pi} \int_{2\sigma_j^2 \varepsilon}^{2\sigma_j^2 R-1} \frac{\exp\left(-s\left(\frac{u+1}{2\sigma_j^2}\right)\right)}{2\sigma_j^2 \sqrt{u}} du \tag{A.55}$$

We note from Figure 1 that $T = \sqrt{R^2 - a^2}$ and taking limits as $R \to \infty$, $\epsilon \to 0$, we have:

$$\begin{aligned}
\underbrace{\frac{1}{2\pi i} \int_{a-i\infty}^{a+i\infty} \frac{\exp(zs)}{\sqrt{1+2\sigma_j^2 z}} dz}_{\text{original integral}} &= \frac{1}{\pi} \int_0^\infty \frac{\exp\left(-s\left(\frac{u+1}{2\sigma_j^2}\right)\right)}{2\sigma_j^2 \sqrt{u}} du \\
&= \frac{1}{\pi} \int_0^\infty \frac{\exp\left(\frac{-su}{2\sigma^2 j}\right) \exp\left(\frac{-s}{2\sigma_j^2}\right)}{2\sigma_j^2 \sqrt{u}} du \\
&= \frac{\exp\left(\frac{-s}{2\sigma_j^2}\right)}{\pi} \int_0^\infty \frac{\exp\left(\frac{-su}{2\sigma_j^2}\right)}{2\sigma_j^2 \sqrt{u}} du
\end{aligned} \tag{A.56}$$

Performing a change of variables, we have:

$$\text{Let } \frac{u}{2\sigma_j^2} = v^2$$

$$\left(\frac{1}{2\sigma_j^2}\right) du = 2v\ dv$$

$$du = \left(2\sigma_j^2\right) 2v\ dv \tag{A.57}$$

Substituting (A.57) into the integral on the r.h.s of (A.56), we have:

$$\frac{\exp\left(\frac{-s}{2\sigma_j^2}\right)}{\pi} \int_0^\infty \frac{\exp\left(\frac{-su}{2\sigma_j^2}\right)}{2\sigma_j^2\sqrt{u}} du = \frac{\exp\left(\frac{-s}{2\sigma_j^2}\right)}{\pi} \int_0^\infty \frac{e^{-v^2 s}}{v\sqrt{2\sigma_j^2}} 2v\ dv$$

$$= \left(\frac{2}{\sqrt{2\sigma_j^2}}\right) \frac{\exp\left(\frac{-s}{2\sigma_j^2}\right)}{\pi} \int_0^\infty e^{-v^2 s} dv \tag{A.58}$$

Performing another change of variables, we have:

$$\text{Let } w = v^2 s$$

$$dw = 2vs\ dv$$

$$\left(\frac{1}{2vs}\right) dw = dv \tag{A.59}$$

Substituting (A.59) into the integral on the r.h.s. of (A.58), we have:

$$\left(\frac{2}{\sqrt{2\sigma_j^2}}\right) \frac{\exp\left(\frac{-s}{2\sigma_j^2}\right)}{\pi} \int_0^\infty e^{-w} \left(\frac{1}{2s\sqrt{\frac{w}{s}}}\right) dw = \frac{\exp\left(\frac{-s}{2\sigma_j^2}\right)}{\pi\sqrt{2s\sigma_j^2}} \underbrace{\int_0^\infty e^{-w} w^{-1/2} dw}_{\Gamma(1/2)=\sqrt{\pi}}$$

$$= \frac{\exp\left(\frac{-s}{2\sigma_j^2}\right)}{\sqrt{\pi}\sqrt{2\sigma_j^2}\sqrt{s}}$$

$$= \text{Gamma}\left(\frac{1}{2}, 2\sigma_j^2\right) \tag{A.60}$$

Thus, using (A.6) and (A.60) we can state:

$$g(s) = g\left(y_\ell^2\right) = \sum_{j=1}^{k} w_j \text{Gamma}\left(\frac{1}{2}, 2\sigma_j^2\right)$$

APPENDIX B

DERIVING THE MOMENT GENERATING FUNCTION FOR A SUM OF SQUARES AND CROSS-PRODUCTS MATRIX FROM A K-COMPONENT FINITE MIXTURE OF MULTIVARIATE GAUSSIANS DISTRIBUTION

Let us suppose that the random vectors $\mathbf{z}_1, \mathbf{z}_2, \ldots, \mathbf{z}_{n_j}$ are independent and identically distributed, each with the following PDF:

$$f_k\left(\mathbf{z}_\ell\right) = \sum_{j=1}^{k} w_j f_j\left(\mathbf{z}_\ell \mid \boldsymbol{\tau}_j\right), \tag{B.1}$$

where $\mathbf{z}_\ell$ is a vector of dimension $(p \times 1)$, $\infty < \mathbf{z}_\ell < \infty$, $\boldsymbol{\tau}_j = \left(\boldsymbol{\mu}_j, \boldsymbol{\Sigma}_j\right)$, $|\boldsymbol{\Sigma}_j| > 0$, and $\ell = 1, \ldots, n_j$. Each f_j in (B.1) is a p-dimensional multivariate Gaussian distribution given by:

$$f_j\left(\mathbf{z}_\ell \mid \boldsymbol{\tau}_j\right) = \frac{1}{(2\pi)^{p/2}\left|\boldsymbol{\Sigma}_j\right|^{1/2}} \exp\left(-\frac{1}{2}\left({\mathbf{z}_\ell}^T\left(\boldsymbol{\Sigma}_j\right)^{-1}\mathbf{z}_\ell\right)\right) \tag{B.2}$$

Let the random matrix $\mathbf{A}^*$ be defined as:

$$\mathbf{A}^* = \sum_{\ell=1}^{n_j} \mathbf{z}_\ell {\mathbf{z}_\ell}^T, \tag{B.3}$$

which is a matrix similar in form to that defined in (5.79). Further, each $\mathbf{z}_\ell$ is distributed as shown in (B.1). Next, let us introduce the $p \times p$ matrix $\boldsymbol{\Theta}_j = (\theta)_{st}^j$, with $(\theta)_{st}^j = (\theta)_{ts}^j$.

In this parametrization, $(\theta)_{st}^{j}$ represents the matrix element in the s^{th} row and t^{th} column of $\boldsymbol{\Theta}_j$. Next, let us derive the moment-generating function of $\mathbf{A}^*$:

$$
\begin{aligned}
M_{\mathbf{A}^*}(\boldsymbol{\Theta}) &= \mathbf{E}\left(\exp\left[\operatorname{tr}\left(\mathbf{A}^*\boldsymbol{\Theta}\right)\right]\right) \\
&= \int_{\infty}^{\infty} \exp\left[\operatorname{tr}\left(\mathbf{A}^*\boldsymbol{\Theta}_j\right)\right] \sum_{j=1}^{k} w_j f_j\left(\mathbf{z}_\ell \mid \boldsymbol{\tau}_j\right) d\mathbf{z}_\ell && \text{(B.4)} \\
&= \sum_{j=1}^{k} w_j \int_{\infty}^{\infty} \exp\left[\operatorname{tr}\left(\mathbf{A}^*\boldsymbol{\Theta}_j\right)\right] f_j\left(\mathbf{z}_\ell \mid \boldsymbol{\tau}_j\right) d\mathbf{z}_\ell && \text{(B.5)} \\
&= \sum_{j=1}^{k} w_j \mathbf{E}\left(\exp\left[\operatorname{tr}\left(\mathbf{A}^*\boldsymbol{\Theta}_j\right)\right]\right) && \text{(B.6)}
\end{aligned}
$$

We can note that (B.5) follows from (B.4) due to the finite mixture model framework; therefore, the order of summation and integration can be interchanged. Also, (B.6) follows from (B.4) because the integrand in (B.5) is simply the expected value of $\exp\left[\operatorname{tr}\left(\mathbf{A}^*\boldsymbol{\Theta}_j\right)\right]$. Now, substituting (B.3) into (B.6), we now have:

$$
\begin{aligned}
\sum_{j=1}^{k} w_j \mathbf{E}\left(\exp\left[\operatorname{tr}\left(\mathbf{A}^*\boldsymbol{\Theta}_j\right)\right]\right) &= \sum_{j=1}^{k} w_j \mathbf{E}\left(\exp\left[\operatorname{tr}\left(\sum_{\ell=1}^{n_j} \mathbf{z}_\ell {\mathbf{z}_\ell}^T \boldsymbol{\Theta}_j\right)\right]\right) \\
&= \sum_{j=1}^{k} w_j \mathbf{E}\left(\exp\left[\operatorname{tr}\left(\sum_{\ell=1}^{n_j} {\mathbf{z}_\ell}^T \boldsymbol{\Theta}_j \mathbf{z}_\ell\right)\right]\right) && \text{(B.7)} \\
&= \sum_{j=1}^{k} w_j \mathbf{E}\left(\exp\left(\sum_{\ell=1}^{n_j} {\mathbf{z}_\ell}^T \boldsymbol{\Theta}_j \mathbf{z}_\ell\right)\right) && \text{(B.8)}
\end{aligned}
$$

Both (B.7) and (B.8) follow from the properties of the trace of a square matrix [45]. Because each $\mathbf{z}_\ell$ is independent and identically distributed, we can write (B.8) as:

$$
\begin{aligned}
\sum_{j=1}^{k} w_j \mathbf{E}\left(\exp\left(\sum_{\ell=1}^{n_j} {\mathbf{z}_\ell}^T \boldsymbol{\Theta}_j \mathbf{z}_\ell\right)\right) &= \sum_{j=1}^{k} w_j \prod_{\ell=1}^{n_j} \mathbf{E}\left(\exp\left({\mathbf{z}_\ell}^T \boldsymbol{\Theta}_j \mathbf{z}_\ell\right)\right) \\
&= \sum_{j=1}^{k} w_j \left[\mathbf{E}\exp\left(\boldsymbol{\zeta}^T \boldsymbol{\Theta}_j \boldsymbol{\zeta}\right)\right]^{n_j}, && \text{(B.9)}
\end{aligned}
$$

For the right-hand side of (B.7), $\boldsymbol{\zeta} \sim \mathcal{N}_p(\mathbf{0}, \boldsymbol{\Sigma}_\mathbf{j})$. Before proceeding further, we will utilize the following theorems from Anderson [45].

Theorem 4. (Anderson A.2.1 (2003)) Given any symmetric matrix $\mathbf{B}$, there exists an orthogonal matrix $\mathbf{C}$ such that

$$\mathbf{C}^T\mathbf{B}\mathbf{C} = \mathbf{D} = \begin{pmatrix} d_1 & 0 & \dots & 0 \\ 0 & d_2 & \dots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \dots & d_p \end{pmatrix}$$

If $\mathbf{B}$ is positive semi-definite, then $d_h \geq 0,\ h = 1, \dots, p$; if $\mathbf{B}$ is positive definite, then $d_h > 0$.

Theorem 5. (Anderson A.2.2 (2003)) Given a positive semi-definite matrix $\mathbf{B}$ and a positive definite matrix $\mathbf{A}$, there exists a non-singular matrix $\mathbf{F}$ such that

$$\mathbf{F}^T\mathbf{B}\mathbf{F} = \begin{pmatrix} \lambda_1 & 0 & \dots & 0 \\ 0 & \lambda_2 & \dots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \dots & \lambda_p \end{pmatrix}$$
$$\mathbf{F}^T\mathbf{A}\mathbf{F} = \mathbf{I}_p,$$

where $\lambda_1 \geq \dots \geq \lambda_p\ (\geq 0)$ are the eigenvalues of $\mathbf{B}$. If $\mathbf{B}$ is positive definite, $\lambda_h \geq 0$, $h = 1, \dots, p$.

Now, returning to (B.9), and utilizing Theorems 4 and 5, we have the following. For a real given $\mathbf{\Theta}_j$ matrix, there exists a non-singular $p \times p$ matrix $\mathbf{B}_j$ such that:

$$\mathbf{B}_j^T \left(\mathbf{\Sigma}_j\right)^{-1} \mathbf{B}_j = \mathbf{I}_p \tag{B.10}$$

$$\mathbf{B}_j^T \mathbf{\Theta}_j \mathbf{B}_j = \mathbf{D}_j, \tag{B.11}$$

where $\mathbf{D}_j$ is a real diagonal matrix. Returning to (B.9), we previously indicated that $\boldsymbol{\zeta} \sim \mathcal{N}_p(\mathbf{0}, \mathbf{\Sigma_j})$. Another way to state this is that $\boldsymbol{\zeta} = \left(\mathbf{\Sigma}_j\right)^{1/2}\mathbf{y}$, where $\mathbf{y} \sim \mathcal{N}_p(\mathbf{0}, \mathbf{I_p})$. Therefore, using this alternative formulation, and noting that from (B.10) that $\mathbf{B}_j = (\boldsymbol{\Sigma}_j)^{1/2}$, we can restate (B.9) as:

$$\begin{aligned}
\sum_{j=1}^{k} w_j \left[\mathbf{E}\exp\left(\boldsymbol{\zeta}^T \boldsymbol{\Theta}_j \boldsymbol{\zeta}\right)\right]^{n_j} &= \sum_{j=1}^{k} w_j \left[\mathbf{E}\exp\left(\left(\mathbf{B}_j\mathbf{y}\right)^T \left(\mathbf{B}_j^T\right)^{-1} \mathbf{D}_j \left(\mathbf{B}_j\right)^{-1} \mathbf{B}_j\mathbf{y}\right)\right]^{n_j} \\
&= \sum_{j=1}^{k} w_j \left[\mathbf{E}\exp\left(\mathbf{y}^T \mathbf{D}_j \mathbf{y}\right)\right]^{n_j} \\
&= \sum_{j=1}^{k} w_j \left[\mathbf{E}\prod_{h=1}^{p} \exp\left(\left(d_{hh}\right)_j \mathbf{y}_h^2\right)\right]^{n_j} \\
&= \sum_{j=1}^{k} w_j \left[\prod_{h=1}^{p} \mathbf{E}\left[\exp\left(\left(d_{hh}\right)_j \mathbf{y}_h^2\right)\right]\right]^{n_j}, \qquad \text{(B.12)}
\end{aligned}$$

where $(d_{hh})_j$ is the h^{th} diagonal element of $\mathbf{D}_j$. Further, the h^{th} factor in the product on the right-hand side of (5.91) is $\mathbf{E}\left[\exp\left(\left(d_{hh}\right)_j y_h^2\right)\right]$, where $y_h \sim \mathcal{N}(0,1)$. This expectation is the MGF of a chi-squared random variable with 1 degree of freedom: $\left(1 - 2\left(d_{hh}\right)_j\right)^{-1/2}$. Substituting into the right-hand side of (B.12) we now have:

$$\begin{aligned}
\sum_{j=1}^{k} w_j \left[\prod_{h=1}^{p} \mathbf{E}\left[\exp\left(\left(d_{hh}\right)_j \mathbf{y}_h^2\right)\right]\right]^{n_j} &= \sum_{j=1}^{k} w_j \left[\prod_{h=1}^{p} \left(1 - 2\left(d_{hh}\right)_j\right)^{-1/2}\right]^{n_j} \\
&= \sum_{j=1}^{k} w_j \left[\left(\det\left(\mathbf{I} - 2\mathbf{D}_j\right)\right)^{-1/2}\right]^{n_j} \qquad \text{(B.13)}
\end{aligned}$$

Recognizing that $\mathbf{I} - 2\mathbf{D}_j$ is a diagonal matrix and by using (B.10) - (B.11) we now have:

$$\begin{aligned}
\det\left(\mathbf{I} - 2\mathbf{D}_j\right) &= \det\left(\mathbf{B}_j^T \left(\boldsymbol{\Sigma}_j\right)^{-1} \mathbf{B}_j - 2\mathbf{B}_j^T \boldsymbol{\Theta}_j \mathbf{B}_j\right) \\
&= \det\left(\mathbf{B}_j^T \left(\left(\boldsymbol{\Sigma}_j\right)^{-1} - 2\boldsymbol{\Theta}_j\right) \mathbf{B}_j\right) \\
&= \det\left(\mathbf{B}_j^T\right) \det\left(\left(\boldsymbol{\Sigma}_j\right)^{-1} - 2\boldsymbol{\Theta}_j\right) \det\left(\mathbf{B}_j\right) \qquad \text{(B.14)} \\
&= \left(\det\left(\mathbf{B}_j\right)\right)^2 \det\left(\left(\boldsymbol{\Sigma}_j\right)^{-1} - 2\boldsymbol{\Theta}_j\right) \qquad \text{(B.15)}
\end{aligned}$$

Equations (B.14) - (B.15) follow from the properties of determinants. Specifically, (B.14) follows from noting that the determinant of a product of matrices of the same dimension is the product of the determinants for each individual matrix. In this case, $\mathbf{B}_j$ and $(\boldsymbol{\Sigma}_j)^{-1} - 2\boldsymbol{\Theta}_j$ are each of dimension $p \times p$. In addition, (B.15) follows from the property that a matrix and its transpose have the same determinant (and recognizing that the determinant is a scalar). From (B.10) we know that $\mathbf{B}_j = (\boldsymbol{\Sigma}_j)^{1/2} \Rightarrow (\det(\mathbf{B}_j))^2 = 1/\det\left((\boldsymbol{\Sigma}_j)^{-1}\right)$. Combining this with (B.13) and (B.15), we now have:

$$\begin{aligned}
\sum_{j=1}^{k} w_j \left[(\det(\mathbf{I} - 2\mathbf{D}_j))^{-1/2} \right]^{n_j} &= \sum_{j=1}^{k} w_j (\det(\mathbf{I} - 2\mathbf{D}_j))^{-n_j/2} \\
&= \sum_{j=1}^{k} w_j \left[\frac{\left(\det\left((\boldsymbol{\Sigma}_j)^{-1}\right)\right)^{n_j/2}}{\left(\det\left((\boldsymbol{\Sigma}_j)^{-1} - 2\boldsymbol{\Theta}_j\right)\right)^{n_j/2}} \right] \\
&= \sum_{j=1}^{k} w_j (\det(\mathbf{I} - 2\boldsymbol{\Theta}_j \boldsymbol{\Sigma}_j))^{-n_j/2} \qquad \text{(B.16)}
\end{aligned}$$

APPENDIX C

DERIVING THE MOMENT GENERATING FUNCTION FROM A K-COMPONENT FINITE MIXTURE OF WISHART DISTRIBUTIONS

Based on (3.80), let us define the MGF of $\mathbf{A}^*$.

$$\begin{aligned}
M_{\mathbf{A}^*}(\boldsymbol{\Theta}) &= \mathbf{E}\left[\operatorname{etr}\left(\mathbf{A}^*\boldsymbol{\Theta}\right)\right] \\
&= \int_{\mathbf{A}^*>0} \sum_{j=1}^{k} w_j \left\{2^{(n_j p)/2}\Gamma_p\left(\frac{n_j}{2}\right)\det\left(\boldsymbol{\Sigma}_j\right)^{n_j/2}\right\}^{-1} \det\left(\mathbf{A}^*\right)^{(n_j-p-1)/2} \times \\
&\quad \operatorname{etr}\left(-\frac{1}{2}(\boldsymbol{\Sigma}_j)^{-1}\mathbf{A}^*\right)\operatorname{etr}\left(\boldsymbol{\Theta}_j\mathbf{A}^*\right) d\mathbf{A}^* && \text{(C.1)} \\
&= \sum_{j=1}^{k} w_j \int_{\mathbf{A}^*>0} \left\{2^{(n_j p)/2}\Gamma_p\left(\frac{n_j}{2}\right)\det\left(\boldsymbol{\Sigma}_j\right)^{n_j/2}\right\}^{-1} \det\left(\mathbf{A}^*\right)^{(n_j-p-1)/2} \times \\
&\quad \operatorname{etr}\left(-\frac{1}{2}(\boldsymbol{\Sigma}_j)^{-1}\mathbf{A}^*\right)\operatorname{etr}\left(\boldsymbol{\Theta}_j\mathbf{A}^*\right) d\mathbf{A}^* && \text{(C.2)} \\
&= \sum_{j=1}^{k} w_j \left\{2^{(n_j p)/2}\Gamma_p\left(\frac{n_j}{2}\right)\det\left(\boldsymbol{\Sigma}_j\right)^{n_j/2}\right\}^{-1} \int_{\mathbf{A}^*>0} \det\left(\mathbf{A}^*\right)^{(n_j-p-1)/2} \times \\
&\quad \operatorname{etr}\left(-\frac{1}{2}(\boldsymbol{\Sigma}_j)^{-1}\mathbf{A}^*\right)\operatorname{etr}\left(\boldsymbol{\Theta}_j\mathbf{A}^*\right) d\mathbf{A}^*, && \text{(C.3)}
\end{aligned}$$

where $\boldsymbol{\Theta}$ is a $p \times p$ symmetric real matrix, $\boldsymbol{\Theta}_j$ is a $p \times p$ symmetric real matrix from the j^{th} component distribution, and $\mathbf{A}^*$ is a positive definite matrix. We note that (C.2) follows from (C.1) because m is considered to be finite. Now continuing to work with the r.h.s of (C.3), we have

$$= \sum_{j=1}^{k} w_j \left\{ 2^{(n_j p)/2} \Gamma_p \left(\frac{n_j}{2} \right) \det\left(\mathbf{\Sigma}_j\right)^{n_j/2} \right\}^{-1} \int_{\mathbf{A}^*>0} \det\left(\mathbf{A}^*\right)^{(n_j-p-1)/2} \times$$
$$\operatorname{etr}\left(-\frac{1}{2} \left[\left(\mathbf{\Sigma}_j\right)^{-1} \mathbf{A}^* - 2\mathbf{\Theta}_j \mathbf{A}^* \right] \right) d\mathbf{A}^* \tag{C.4}$$
$$= \sum_{j=1}^{k} w_j \left\{ 2^{(n_j p)/2} \Gamma_p \left(\frac{n_j}{2} \right) \det\left(\mathbf{\Sigma}_j\right)^{n_j/2} \right\}^{-1} \int_{\mathbf{A}^*>0} \det\left(\mathbf{A}^*\right)^{(n_j-p-1)/2} \times$$
$$\operatorname{etr}\left(-\frac{1}{2} \left[\left(\mathbf{\Sigma}_j\right)^{-1} \mathbf{A}^* - 2\mathbf{\Theta}_j \mathbf{\Sigma}_j \left(\mathbf{\Sigma}_j\right)^{-1} \mathbf{A}^* \right] \right) d\mathbf{A}^* \tag{C.5}$$
$$= \sum_{j=1}^{k} w_j \left\{ 2^{(n_j p)/2} \Gamma_p \left(\frac{n_j}{2} \right) \det\left(\mathbf{\Sigma}_j\right)^{n_j/2} \right\}^{-1} \int_{\mathbf{A}^*>0} \det\left(\mathbf{A}^*\right)^{(n_j-p-1)/2} \times$$
$$\operatorname{etr}\left(-\frac{1}{2} \left[\mathbf{I}_p - 2\mathbf{\Theta}_j \mathbf{\Sigma}_j \right] \mathbf{\Sigma}_j^{-1} \mathbf{A}^* \right) d\mathbf{A}^* \tag{C.6}$$

Before continuing with evaluating the integral in (C.6), some results from matrix variate distributions may be helpful.

Definition 16. *Matrix-variate Laplace Transform (Gupta and Nagar, 1999).* Let $f(\mathbf{J})$ be a function of $\mathbf{J}_{p\times p} > \mathbf{0}$ (positive definite) and let $\mathbf{L}$ be a $p \times p$ complex symmetric matrix. A complex matrix is one whose elements may be complex numbers. Then the matrix-variate Laplace transform $g(\ell)$ of $f(\mathbf{J})$ is defined as

$$g(\ell) = \int_{\mathbf{J}>\mathbf{0}} \operatorname{etr}\left(-\mathbf{LJ}\right) f\left(\mathbf{J}\right) d\mathbf{J}, \tag{C.7}$$

where the integral is assumed to be absolutely convergent in the right half-plane, $Re\left(\ell\right) > \mathbf{0}$.

Using Definition 2, a matrix-variate Laplace transform which will be useful to our continued developments is

$$g(\ell) = \int_{\mathbf{\Lambda}>\mathbf{0}} \operatorname{etr}\left(-\mathbf{\Lambda L}\right) \det\left(\mathbf{\Lambda}\right)^{b-\frac{1}{2}(p+1)} d\mathbf{\Lambda}, \tag{C.8}$$

where $\mathbf{\Lambda}, \mathbf{L}$ are $p \times p$ symmetric complex matrices. Herz (1955) demonstrated that the matrix-variate Laplace transform in (C.8) is absolutely convergent for $Re(\mathbf{L}) > \mathbf{0}$. Because we are interested in the case where $\mathbf{L}$ is real, we shall restrict further developments to the $Re(\mathbf{L}) > \mathbf{0}$ case. Before we can proceed to evaluate the integral in (C.6), we must first review and establish some results for Jacobians of matrix-variate transformations.

Definition 17. *Matrix-variate Jacobian (Gupta and Nagar, 1999).* Let $\mathbf{X}$ and $\mathbf{Y}$ be two matrices having the same number of independent elements $x_1, \ldots, x_p$ and $y_1, \ldots, y_p$, respectively. Consider the matrix transformation $\mathbf{Y} = F(\mathbf{X})$. Then the Jacobian of the transformation from $\mathbf{X}$ to $\mathbf{Y}$ is defined as:

$$J(\mathbf{X} \to \mathbf{Y}) = mod \det \begin{pmatrix} \frac{\partial x_1}{\partial y_1} & \cdots & \frac{\partial x_1}{\partial y_p} \\ \vdots & \ddots & \vdots \\ \frac{\partial x_p}{\partial y_1} & \cdots & \frac{\partial x_p}{\partial y_p} \end{pmatrix}, \tag{C.9}$$

where *mod* indicates the modulus.

Now, we will present several theorems with accompanying proofs regarding Jacobians of linear matrix-variate transformations that will be useful for evaluating the integral in equation (C.6).

Theorem 6. Linear transformation of a vector (Deemer and Olkin, 1951). For $\mathbf{y}$ and $\mathbf{x}$, each $p \times 1$ vectors and $\mathbf{K}$ a $p \times p$ matrix, define $\mathbf{y} = \mathbf{K}\mathbf{x}$. Then $J(\mathbf{x} \to \mathbf{y}) = \mathrm{mod}\det(\mathbf{K})$.

Proof. Let $y_\ell = \sum_{h=1}^{p} a_{\ell h} x_h$ and $\frac{\partial y_\ell}{\partial x_j} = k_{\ell j}$. Then, $J(\mathbf{x} \to \mathbf{y}) = \mathrm{mod}\det(\mathbf{K})$. □

Theorem 7. Linear transformation of a matrix - I (Deemer and Olkin, 1951). For $\mathbf{Y}$ and $\mathbf{X}$, each $p \times q$ matrices, and $\mathbf{K}$, a $p \times p$ matrix, define $\mathbf{Y} = \mathbf{K}\mathbf{X}$. Then $J(\mathbf{X} \to \mathbf{Y}) = \mathrm{mod}\det(\mathbf{K})^q$.

Proof. This follows from Theorem 6, because the transformation of each column of $\mathbf{Y}$ is independent of the others, and there are q such columns of $\mathbf{Y}$. Further, the Jacobian of each column transformation is $mod \det(\mathbf{K})$. □

Theorem 8. Linear transformation of a matrix - II (Deemer and Olkin, 1951). For $\mathbf{Y}$ and $\mathbf{X}$, each $p \times q$ matrices, $\mathbf{K}$, a $p \times p$ matrix, and $\mathbf{M}$, a $q \times q$ matrix, define $\mathbf{Y} = \mathbf{KXM}$. Then $J(\mathbf{X} \to \mathbf{Y}) = \mathrm{mod}\det(\mathbf{K})^q(\mathbf{M})^p$.

Proof. Let $\mathbf{U} = \mathbf{KX}$ and then $\mathbf{Y} = \mathbf{UM}$. Using Theorem 7, it follows that $J(\mathbf{X} \to \mathbf{U}) = \mathit{mod}\det(\mathbf{K})^q$ and $J(\mathbf{U} \to \mathbf{Y}) = \mathit{mod}\det(\mathbf{M})^p$. Because Jacobians are essentially partial derivatives (or functions thereof), the desired result follows from the application of the chain rule for calculating derivatives. $\square$

Lemma 1. Linear transformation of a matrix - III (Deemer and Olkin, 1951). Let $\mathbf{Y} = \mathbf{K}_n\mathbf{K}_{n-1}\cdots\mathbf{K}_1\mathbf{X}\mathbf{K}_1^T\cdots\mathbf{K}_{n-1}^T\mathbf{K}_n^T$. Then $J(\mathbf{X} \to \mathbf{Y}) = J(\mathbf{X} \to \mathbf{Y}_1)J(\mathbf{Y}_1 \to \mathbf{Y}_2)\cdots J(\mathbf{Y}_{n-1} \to \mathbf{Y})$, where $\mathbf{Y}_\ell = \mathbf{K}_\ell\mathbf{Y}_{\ell-1}\mathbf{K}_\ell^T, \mathbf{Y}_0 = \mathbf{X}, \mathbf{Y}_n = \mathbf{Y}, \ell = 1, \ldots, n$.

Proof. This follows from the application of the chain rule of differentiation as demonstrated in the proof of Theorem 8. $\square$

Theorem 9. Linear transformation of a matrix - IV (Deemer and Olkin, 1951). Let $\mathbf{Y}$ be a $p \times p$ matrix, $\mathbf{X}$ be a symmetric $p \times p$ matrix, and $\mathbf{K}$ be a $p \times p$ matrix. If $\mathbf{Y} = \mathbf{KXK}^T$, then $J(\mathbf{X} \to \mathbf{Y}) = \mathrm{mod}\det(\mathbf{K})^{p+1}$

Proof. As a first step in this proof, we shall define an elementary transformation matrix for any $\mathbf{I}_p$ matrix. An elementary matrix is a matrix obtained by applying a single elementary row transformation to $\mathbf{I}_p$. These transformations include:

1. Interchanging any 2 rows of a given matrix (e.g., $R_1 \longleftrightarrow R_2$)
2. Multiplying a single row of a given matrix by a constant (e.g., $cR_1 \to R_1$)
3. Adding a multiple of one row of a given matrix to another row (e.g., $R_3 \to R_3 + cR_1$)

These elementary row transformations also define the different types of elementary matrices. We are interested in finding $\mathit{mod}\det(\mathbf{K})^{p+1}$, where $\mathbf{K}$ is a square matrix of dimension p. Based on the well-known properties of determinants (Searle, 1982), when two rows (columns) of a matrix are interchanged, a determinant changes its sign. Because the determinant of the

$\mathbf{I}_p$ matrix equals 1, interchanging any two rows of the $\mathbf{I}_p$ matrix will result in a determinant equal to -1. The modulus of this determinant is equal to 1. Therefore, further developments of the proof will focus on the elementary row transformations delineated in 2. and 3. above. Now, let us state some properties of the elementary row transformation matrices specified by 2. and 3. above. Let us denote $\mathbf{R}_{aa}(c^*)$ as the $\mathbf{I}_p$ identity matrix with the a^{th} diagonal element replaced by c^*; this is the matrix specified in 2. above (multiplication of a single row of a given matrix by a constant). Because $\mathbf{R}_{aa}(c^*)$ is a diagonal matrix, $\det(\mathbf{R}_{aa}(c^*)) = c^*$. Because the determinant is non-zero, $\mathbf{R}_{aa}(c^*)$ is non-singular, and its inverse is the identity matrix with $(c^*)^{-1}$ as the a^{th} diagonal element.

Similarly, let us denote $\mathbf{P}_{ab}(c^*)$ as an upper (lower) triangular matrix with the diagonal elements and the appropriate off-diagonal elements equal to c^*. Therefore, $\det(\mathbf{P}_{ab}(c^*)) = 1$ and $\mathbf{P}_{ab}^{-1}(c^*) = \mathbf{P}_{ab}(-c^*)$. These elementary matrices play an important role in equivalent canonical forms as the following theorem demonstrates.

Theorem 10. Full-rank factorization (Searle, 1982). Any non-null matrix $\mathbf{S}$ of rank r is equivalent to $\mathbf{PSQ} = \begin{pmatrix} \mathbf{I}_r & \mathbf{0} \\ \mathbf{0} & \mathbf{0} \end{pmatrix}$, where $\mathbf{I}_r$ is the identity matrix of dimension r, $\mathbf{S}$ is a matrix of dimension $m \times n$, and $\mathbf{P}$ and $\mathbf{Q}$ are non-singular matrices of dimension $m \times m$ and $n \times n$, respectively. Therefore, we note that $\mathbf{S} = \mathbf{P}^{-1} \begin{pmatrix} \mathbf{I}_r & \mathbf{0} \\ \mathbf{0} & \mathbf{0} \end{pmatrix} \mathbf{Q}^{-1}$,and $\mathbf{S}$ can be seen as a product of elementary matrices, and, as was stated previously, the inverse of an elementary matrix is also an elementary matrix.

Returning to Theorem 9, which we wish to prove, we note that matrix $\mathbf{K}$ is a $p \times p$ matrix. Using Theorem 10, we can rewrite $\mathbf{K}$ as follows:

$$\begin{aligned} \mathbf{K} &= (\mathbf{E}_1\mathbf{E}_2 \cdots \mathbf{E}_n)^{-1} (\mathbf{E}_{n+1}\mathbf{E}_{n+2} \cdots \mathbf{E}_m)^{-1} \\ &= \mathbf{E}_n^{-1} \cdots \mathbf{E}_2^{-1}\mathbf{E}_1^{-1}\mathbf{E}_m^{-1} \cdots \mathbf{E}_{n+2}^{-1}\mathbf{E}_{n+1}^{-1} && \text{(C.10)} \\ &= \mathbf{F}_m\mathbf{F}_{m-1} \cdots \mathbf{F}_1, && \text{(C.11)} \end{aligned}$$

where the $\mathbf{F}$ matrices in (C.11) are of the type $\mathbf{R}_{aa}(c^*)$ or $\mathbf{P}_{ab}(c^*)$. Because $\mathbf{K}$ is a square matrix, we can rewrite the linear transformation $\mathbf{Y} = \mathbf{K}\mathbf{X}\mathbf{K}^T$ as:

$$\begin{aligned}
\mathbf{Y} &= (\mathbf{F}_m\mathbf{F}_{m-1}\cdots\mathbf{F}_1)\,\mathbf{X}\,(\mathbf{F}_m\mathbf{F}_{m-1}\cdots\mathbf{F}_1)^T \\
&= (\mathbf{F}_m\mathbf{F}_{m-1}\cdots\mathbf{F}_1)\,\mathbf{X}\left(\mathbf{F}_1^T\cdots\mathbf{F}_{m-1}^T\mathbf{F}_m^T\right)
\end{aligned} \tag{C.12}$$

By using Lemma 1, we can write the Jacobian, $J(\mathbf{X} \to \mathbf{Y})$ as:

$$J(\mathbf{X} \to \mathbf{Y}) = J(\mathbf{X} \to \mathbf{Y}_1)J(\mathbf{Y}_1 \to \mathbf{Y}_2)\cdots J(\mathbf{Y}_{m-1} \to \mathbf{Y}), \tag{C.13}$$

where $\mathbf{Y}_d = \mathbf{F}_d\mathbf{Y}_{d-1}\mathbf{F}_d^T, \mathbf{Y}_0 = \mathbf{X}, \mathbf{Y}_m = \mathbf{Y}, (d = 1, \ldots, m)$.

Let $\mathbf{G}$ represent any of the $\mathbf{F}$ matrices of the form $\mathbf{R}_{aa}(c^*)$. Therefore, the transformation $\mathbf{Y}_d = \mathbf{G}\mathbf{Y}_{d-1}\mathbf{G}^T$ implies that $y_{aa} = (c^*)^2 x_{aa},\ y_{ab} = c^* x_{ab}\ (a \neq b),$ and $y_{bc} = x_{bc}\ (b, c \neq a)$. The matrix of the partial derivatives is thus diagonal with $(p-1)$ elements c^* and one element $(c^*)^2$. Thus:

$$\begin{aligned}
J(\mathbf{Y}_{d-1} \to \mathbf{Y}_d) &= mod\ (c^*)^2 c^{(p-1)}(1) \\
&= mod\ c^{(p+1)} \\
&= mod\ \det(\mathbf{G})^{(p+1)}
\end{aligned} \tag{C.14}$$

Thus, it follows that:

$$\begin{aligned}
J(\mathbf{X} \to \mathbf{Y}) &= mod\ \left[\det(\mathbf{F}_m)^{p+1}\det(\mathbf{F}_{m-1})^{p+1}\cdots\det(\mathbf{F}_1)^{p+1}\right] \\
&= mod\ \left[\det(\mathbf{F}_m\mathbf{F}_{m-1}\cdots\mathbf{F}_1)^{p+1}\right] \\
&= mod\ \det(\mathbf{K})^{p+1}
\end{aligned} \tag{C.15}$$

Similarly, let $\mathbf{H}$ represent any of the $\mathbf{F}$ matrices of the form $\mathbf{P}_{ab}(c^*)$. Therefore, the transformation $\mathbf{Y}_d = \mathbf{H}\mathbf{Y}_{d-1}\mathbf{H}^T$ implies that $y_{aa} = x_{aa} + 2c^* x_{ab} + (c^*)^2 x_{bb}, y_{ac} = y_{ca} = x_{ac} + c^* x_{bc},\ (c \neq a),$ and $y_{bc} = x_{bc}\ (b, c \neq a)$. The matrix of partial derivatives will be an upper-triangular matrix with 1s on the main diagonal. Therefore, its determinant is equal to 1. Thus, for this type of elementary transformation, $J(\mathbf{Y}_{d-1} \to \mathbf{Y}_d) = mod\ \det(\mathbf{H})^{p+1}$. Similarly, it follows that $J(\mathbf{X} \to \mathbf{Y}) = mod\ \det(\mathbf{K})^{p+1}$, and the proof is now complete. □

Now, we will state a corollary of Theorem 9 in terms of a linear transformation involving $\mathbf{K}^{-1}$.

Corollary 1. Linear transformation of a matrix - V (Deemer and Olkin, 1951). For $\mathbf{Y}, \mathbf{X}$(symmetric), and $\mathbf{K}$ all $p \times p$ matrices, define $\mathbf{Y} = \mathbf{K}^{-1}\mathbf{X}\left(\mathbf{K}^{-1}\right)^T$. Then $J(\mathbf{X} \to \mathbf{Y}) = mod\ \det(\mathbf{K})^{-(p+1)}$.

Proof. Using a development similar to that used in the proof of the preceding theorem, $\mathbf{K}$ is a non-singular matrix. Using elementary matrices we can write $\mathbf{K}^{-1}$ as

$$\begin{aligned}
\mathbf{K}^{-1} &= \left(\mathbf{F}_m\mathbf{F}_{m-1}\cdots\mathbf{F}_1\right)^{-1} \\
&= \mathbf{F}_1^{-1}\cdots\mathbf{F}_{m-1}^{-1}\mathbf{F}_m^{-1} \\
&= \mathbf{F}_1^{*}\cdots\mathbf{F}_{m-1}^{*}\mathbf{F}_m^{*},
\end{aligned} \tag{C.16}$$

where the $\mathbf{F}^*$ matrices in (C.16) are of the type $\mathbf{R}_{aa}(c^*)^{-1}$ or $\mathbf{P}_{ab}(-c^*)$. Therefore, the linear transformation $\mathbf{Y} = \mathbf{K}^{-1}\mathbf{X}\left(\mathbf{K}^{-1}\right)^T$ can be written as

$$\begin{aligned}
\mathbf{Y} &= \left(\mathbf{F}_1^{*}\cdots\mathbf{F}_{m-1}^{*}\mathbf{F}_m^{*}\right)\mathbf{X}\left(\mathbf{F}_1^{*}\cdots\mathbf{F}_{m-1}^{*}\mathbf{F}_m^{*}\right)^T \\
&= \left(\mathbf{F}_1^{*}\cdots\mathbf{F}_{m-1}^{*}\mathbf{F}_m^{*}\right)\mathbf{X}\left[\left(\mathbf{F}_m^{*}\right)^T\left(\mathbf{F}_{m-1}^{*}\right)^T\cdots\left(\mathbf{F}_1^{*}\right)^T\right]
\end{aligned} \tag{C.17}$$

By using Lemma 1, we can write the Jacobian $J(\mathbf{X} \to \mathbf{Y})$ as:

$$J(\mathbf{X} \to \mathbf{Y}) = J(\mathbf{X} \to \mathbf{Y}_1)J(\mathbf{Y}_1 \to \mathbf{Y}_2)\cdots J(\mathbf{Y}_{m-1} \to \mathbf{Y}),$$

where $\mathbf{Y}_d = \mathbf{F}_d^{*}\mathbf{Y}_{d-1}\left(\mathbf{F}_d^{*}\right)^T (d = 1, \ldots, m), \mathbf{Y}_0 = \mathbf{X}, \mathbf{Y}_m = \mathbf{Y}$. Let $\mathbf{G}^*$ represent any of the $\mathbf{F}^*$ matrices of the form $\mathbf{R}_{aa}(c^*)^{-1}$. Therefore, the transformation $\mathbf{Y}_d = \mathbf{G}^*\mathbf{Y}_{d-1}\left(\mathbf{G}^*\right)^T$ implies that $y_{aa} = (c^*)^{-2}x_{aa},\ y_{ab} = (c^*)^{-1}x_{ab}(a \neq b),$ and, $y_{bc} = x_{bc}(b, c \neq a)$. The matrix of the partial derivatives is diagonal with $(p-1)$ elements $(c^*)^{-1}$ and one element $(c^*)^{-2}$. Thus:

$$\begin{aligned} J(\mathbf{Y}_{d-1} \to \mathbf{Y}_d) &= mod\ (c^*)^{-2}(c^*)^{-(p-1)}(1) \\ &= mod\ (c^*)^{-2-p+1} \\ &= mod\ (c^*)^{-p-1} \\ &= mod\ (c^*)^{-(p+1)} \\ &= mod\ \det(\mathbf{G}^*)^{-(p+1)} \end{aligned} \tag{C.18}$$

thus it follows that:

$$\begin{aligned} J(\mathbf{X} \to \mathbf{Y}) &= mod\left[\det\left(\mathbf{F}_1^*\right)^{p+1} \cdots \det\left(\mathbf{F}_{m-1}^*\right)^{p+1} \det\left(\mathbf{F}_m^*\right)^{p+1}\right] \\ &= mod\left[\det\left(\mathbf{F}_1^* \cdots \mathbf{F}_{m-1}^*\mathbf{F}_m^*\right)^{p+1}\right] \\ &= mod\left[\det(\mathbf{K}^{-1})^{(p+1)}\right] \\ &= mod\det(\mathbf{K})^{-(p+1)} \end{aligned} \tag{C.19}$$

Similarly, let $\mathbf{H}^*$ represent any of the $\mathbf{F}^*$ matrices of the form $\mathbf{P}_{ab}(-c^*)$. Therefore, the transformation $\mathbf{Y}_d = \mathbf{H}^*\mathbf{Y}_{d-1}\left(\mathbf{H}^*\right)^T$ implies that $y_{aa} = x_{aa} - 2c^*x_{ab} + (c^*)^2x_{bb},\ y_{ac} = y_{ca} = x_{ac} - c^*x_{bc}(c \neq a),$ and $y_{bc} = x_{bc}(b, c \neq a)$. The matrix of partial derivatives will be an upper-triangular matrix with $1s$ on the main diagonal. Therefore, its determinant is equal to 1. Thus, for this type of elementary transformation, $J(\mathbf{Y}_{d-1} \to \mathbf{Y}_d) = mod\ \det(\mathbf{H}^*)^{-(p+1)}$. This uses the fact that $\det(\mathbf{H}^{-1}) = \det(\mathbf{H})^{-1}$. Similarly, it follows that $J(\mathbf{X} \to \mathbf{Y}) = mod\ \det(\mathbf{K})^{-(p+1)}$ and the proof is now complete. □

Corollary 2. Linear transformation of a matrix - VI. For $\mathbf{Y}, \mathbf{X}(symmetric),$ and $\mathbf{K}$ all $(p \times p)$ matrices, define $\mathbf{Y} = \mathbf{K}^{1/2}\mathbf{X}\left(\mathbf{K}^{1/2}\right)^T$, where $\mathbf{K}^{1/2}$ is the positive square root of $\mathbf{K}$. Then $J(\mathbf{X} \to \mathbf{Y}) = mod\ \det(\mathbf{K})^{-\frac{1}{2}(p+1)}$.

Proof. From Theorem 9, we demonstrated that for $\mathbf{Y} = \mathbf{K}\mathbf{X}\mathbf{K}^T$, we have $J(\mathbf{X} \to \mathbf{Y}) = \mathit{mod}\, \det(\mathbf{K})^{p+1}$. Pre- and post-multiplying each side of $\mathbf{Y} = \mathbf{K}\mathbf{X}\mathbf{K}^T$ by $\mathbf{K}^{-1/2}$ and$(\mathbf{K}^{-1/2})^T$, respectively, we have:

$$\begin{aligned}\mathbf{K}^{-1/2}\mathbf{Y}\left(\mathbf{K}^{-1/2}\right)^T &= \mathbf{K}^{-1/2}\mathbf{K}\mathbf{X}\mathbf{K}^T\left(\mathbf{K}^{-1/2}\right)^T \\ &= \mathbf{K}^{1/2}\mathbf{X}\left(\mathbf{K}^{1/2}\right)^T \qquad \text{(C.20)}\end{aligned}$$

Therefore, we can prove the result of interest by finding the Jacobian of the transformation $\mathbf{Y}^* = \mathbf{K}^{-1/2}\mathbf{Y}\left(\mathbf{K}^{-1/2}\right)^T$, where $\mathbf{Y}^*$ is also a $(p \times p)$ matrix. Using a development similar to that used for the proof of Corollary 1 using elementary matrices, we will apply a similar approach to the proof of Corollary 2. From the proof:

$$\mathbf{K}^{-1} = \left(\mathbf{F}_m\mathbf{F}_{m-1}\cdots\mathbf{F}_1\right)^{-1} \qquad \text{(C.21)}$$

$$\left(\mathbf{K}^{-1}\right)^{1/2} = \left[\left(\mathbf{F}_m\mathbf{F}_{m-1}\cdots\mathbf{F}_1\right)^{-1}\right]^{1/2} \qquad \text{(C.22)}$$

$$= \left(\mathbf{F}_m\mathbf{F}_{m-1}\cdots\mathbf{F}_1\right)^{-1/2} \qquad \text{(C.23)}$$

In (C.21), all of the $\mathbf{F}$ matrices are elementary matrices. By applying the positive square root to both sides of (C.21) we obtain (C.22). As referenced earlier, the $\mathbf{F}$ matrices are assumed to be of the form $\mathbf{R}_{aa}(c)$ or $\mathbf{P}_{ab}(c)$. Because $\mathbf{R}_{aa}(c)$ is a diagonal (and symmetric) matrix, we know that $\mathbf{R}_{aa}^{1/2}(c) = \mathbf{R}_{aa}(\sqrt{c})$. Thus:

$$\begin{aligned}\mathbf{R}_{aa}(c) &= \mathbf{R}_{aa}^{1/2}(c)\left(\mathbf{R}_{aa}^{1/2}(c)\right)^T \\ &= \mathbf{R}_{aa}(\sqrt{c})\left(\mathbf{R}_{aa}(\sqrt{c})\right)^T \\ &= \mathbf{R}_{aa}^2(\sqrt{c}) \qquad \text{(C.24)}\end{aligned}$$

We note that (C.24) is due to the symmetry of $\mathbf{R}_{aa}(c)$. Similarly, we can state

$$\begin{aligned}\mathbf{R}_{aa}(c^{-1}) &= \mathbf{R}_{aa}\left(c^{-1/2}\right)\left(\mathbf{R}_{aa}\left(c^{-1/2}\right)\right)^T \\ &= \mathbf{R}_{aa}^2\left(c^{-1/2}\right) \qquad \text{(C.25)}\end{aligned}$$

Now, let us first assume that all of the $\mathbf{F}$ matrices in (C.23) are of the form $\mathbf{R}_{aa}(c)$. Then we have:

$$
\begin{aligned}
\mathbf{K}^{-1/2} &= (\mathbf{F}_m\mathbf{F}_{m-1}\cdots\mathbf{F}_1)^{-1/2} \\
&= \left[(\mathbf{F}_m\mathbf{F}_{m-1}\cdots\mathbf{F}_1)^{-1}\right]^{1/2} \\
&= \left[\left(\{\mathbf{F}_m^\psi\}^2\{\mathbf{F}_{m-1}^\psi\}^2\cdots\{\mathbf{F}_1^\psi\}^2\right)^{-1}\right]^{1/2} \\
&= \left[\{\mathbf{F}_1^\psi\}^{-2}\cdots\{\mathbf{F}_{m-1}^\psi\}^{-2}\{\mathbf{F}_m^\psi\}^{-2}\right]^{1/2} \\
&= \left[\{\mathbf{F}_1^\psi\}^{-1}\cdots\{\mathbf{F}_{m-1}^\psi\}^{-1}\{\mathbf{F}_m^\psi\}^{-1}\right],
\end{aligned}
\tag{C.26}
$$

where each $\mathbf{F}^\psi$ matrix is of the form $\mathbf{R}_{aa}(\sqrt{c})$. Therefore, using a similar development from the proof of Corollary 1, we note the $\{\mathbf{F}^\psi\}^{-1}$ matrices are all of the form $\mathbf{R}_{aa}\left(c^{-1/2}\right)$. Thus, the linear transformation $\mathbf{Y}^* = \mathbf{K}^{-1/2}\mathbf{Y}\left(\mathbf{K}^{-1/2}\right)^T$ can be written as:

$$
\begin{aligned}
\mathbf{Y}^* &= \left[\{\mathbf{F}_1^\psi\}^{-1}\cdots\{\mathbf{F}_{m-1}^\psi\}^{-1}\{\mathbf{F}_m^\psi\}^{-1}\right]\mathbf{Y}\left[\{\mathbf{F}_1^\psi\}^{-1}\cdots\{\mathbf{F}_{m-1}^\psi\}^{-1}\{\mathbf{F}_m^\psi\}^{-1}\right]^T \\
&= \left[\{\mathbf{F}_1^\psi\}^{-1}\cdots\{\mathbf{F}_{m-1}^\psi\}^{-1}\{\mathbf{F}_m^\psi\}^{-1}\right]\mathbf{Y}\left[\{\mathbf{F}_m^\psi\}^{-1}\right]^T\left[\{\mathbf{F}_{m-1}^\psi\}^{-1}\right]^T\cdots\left[\{\mathbf{F}_1^\psi\}^{-1}\right]^T
\end{aligned}
\tag{C.27}
$$

By once again using Lemma 1, we can write the Jacobian $J(\mathbf{Y}\to\mathbf{Y}^*)$ as:

$$
J(\mathbf{Y}\to\mathbf{Y}^*) = J(\mathbf{Y}\to\mathbf{Y}_1^*)J(\mathbf{Y}_1^*\to\mathbf{Y}_2^*)\cdots J(\mathbf{Y}_{m-1}^*\to\mathbf{Y}^*), \tag{C.28}
$$

where $\mathbf{Y}_d^* = \{\mathbf{F}_d^\psi\}^{-1}\mathbf{Y}_{d-1}^*\left[\{\mathbf{F}_d^\psi\}^{-1}\right]^T$ $(d = 1,\ldots,m)$, $\mathbf{Y}_0^* = \mathbf{Y}$, and $\mathbf{Y}_m^* = \mathbf{Y}^*$. Let $\mathbf{G}^\psi$ represent any of the $\{\mathbf{F}^\psi\}^{-1}$ matrices of the form $\mathbf{R}_{aa}(c^{-1/2})$. Therefore, the transformation $\mathbf{Y}_d^* = \mathbf{G}^\psi\mathbf{Y}_{d-1}^*\left(\mathbf{G}^\psi\right)^T$ implies that $y_{aa}^* = (c^{-1})y_{aa}$, $y_{ab}^* = (c^{-1/2})y_{ab}(a\neq b)$, and $y_{cb}^* = y_{bc}(b,c\neq a)$. The matrix of the partial derivatives is diagonal with $(p-1)$ elements $(c^{-1/2})$ and one element (c^{-1}). Thus:

$$
\begin{aligned}
J(\mathbf{Y}_{d-1}^*\to\mathbf{Y}_d^*) &= \mathit{mod}\ (c^{-1})(c^{-1/2})^{(p-1)}(1) \\
&= \mathit{mod}\ (c)^{-1-\frac{1}{2}(p-1)} \\
&= \mathit{mod}\ (c)^{-1-\frac{1}{2}p+\frac{1}{2}} \\
&= \mathit{mod}\ (c)^{-\frac{1}{2}(p+1)} \\
&= \mathit{mod}\ \det(\mathbf{G}^\psi)^{-\frac{1}{2}(p+1)}
\end{aligned}
\tag{C.29}
$$

Finally, it follows that:

$$\begin{aligned}
J(\mathbf{Y} \to \mathbf{Y}^*) &= mod \left[\det\left(\{\mathbf{F}_1^{\psi}\}^{-1}\right)^{(p+1)} \cdots \det\left(\{\mathbf{F}_{m-1}^{\psi}\}^{-1}\right)^{(p+1)} \det\left(\{\mathbf{F}_m^{\psi}\}^{-1}\right)^{(p+1)}\right] \\
&= mod\ \det\left(\{\mathbf{F}_1^{\psi}\}^{-1} \cdots \{\mathbf{F}_{m-1}^{\psi}\}^{-1}\{\mathbf{F}_m^{\psi}\}^{-1}\right) \\
&= mod\ \det\left[\mathbf{K}^{-1/2}\right]^{(p+1)} \\
&= mod\ \det(\mathbf{K})^{-\frac{1}{2}(p+1)} \qquad \text{(C.30)}
\end{aligned}$$

Now, let us look at the other type of elementary matrix: $\mathbf{P}_{ab}(c)$. Based on earlier statements in this section, we know the $\mathbf{P}_{ab}(c)$ matrix is an upper-triangular matrix with all entries on the main diagonal equal to 1, the entry in the a^{th} row and b^{th} column equal to c, and all other entries equal to 0. Therefore, we wish to find a positive square root matrix, say $\mathbf{B}^*$, such that $\mathbf{B}^{*2} = \mathbf{P}_{ab}(c)$. Let $\mathbf{B}^*$ be a $(p \times p)$ matrix with a general form as follows:

$$\mathbf{B}^* = \begin{pmatrix} b_{11}^* & b_{12}^* & \cdots & b_{1p}^* \\ \vdots & b_{22}^* & \vdots & \vdots \\ \vdots & \vdots & \ddots & \vdots \\ b_{p1}^* & \cdots & \cdots & b_{pp}^* \end{pmatrix} \qquad \text{(C.31)}$$

We wish to find the elements of $\mathbf{B}^*$ such that $\mathbf{B}^{*2} = \mathbf{T}$ or

$$\begin{pmatrix} b_{11}^* & b_{12}^* & \cdots & b_{1p}^* \\ \vdots & b_{22}^* & \vdots & \vdots \\ \vdots & \vdots & \ddots & \vdots \\ b_{p1}^* & \cdots & \cdots & b_{pp}^* \end{pmatrix} \begin{pmatrix} b_{11}^* & b_{12}^* & \cdots & b_{1p}^* \\ \vdots & b_{22}^* & \vdots & \vdots \\ \vdots & \vdots & \ddots & \vdots \\ b_{p1}^* & \cdots & \cdots & b_{pp}^* \end{pmatrix} = \begin{pmatrix} t_{11} & t_{12} & \cdots & t_{1p} \\ 0 & t_{22} & \vdots & \vdots \\ \vdots & 0 & \ddots & \vdots \\ 0 & \vdots & 0 & t_{pp} \end{pmatrix} \qquad \text{(C.32)}$$

Let us first examine the main diagonal elements of the matrix $\mathbf{T}$. On an element-by-element basis:

$$(t)_{aa} = {b_{aa}^*}^2 + \sum_{a \neq b} b_{ab}^* b_{ba}^* \qquad \text{(C.33)}$$

Now, let us look at the off-diagonal elements of $\mathbf{T}$. We have:

$$(t)_{ab} = \sum_{k=1}^{p} b_{ak} b_{kb} \qquad \text{(C.34)}$$

Noting the entries from matrix $\mathbf{T}$, we see that $(t)_{ab} = 0$ when $a < b$. Looking at the main elements on the main diagonal of $\mathbf{T}$, we note that for our matrix $\mathbf{P}_{ab}(c)$, we have the restriction $(t)_{11} = (t)_{22} = \cdots = (t)_{pp} = 1$. By setting all the equations for the $(t)_{aa}$ equal to each other, and noting all the cross-product terms will be eliminated, it follows that $(b^*)^2_{aa} = (t)_{aa} = 1$. Therefore, $(b^*)_{ii} = 1$. Before proceeding further, we note that our matrix $\mathbf{P}_{ab}(c)$ is of the general form of $\mathbf{T}$ with all the $(t)_{ab} = 0$ except one; the remaining one is equal to c. By substituting $(b^*)_{aa} = 1$ into the original equations in (C.33) - (C.34), and then performing a series of other substitutions, it follows that the $\mathbf{B}^*$ matrices are of the form $\mathbf{P}_{ab}(c)$. Further, this leads to the expression $\mathbf{P}^2_{ab}(c/2) = \mathbf{P}_{ab}(c)$. Thus, the positive square root matrix $\mathbf{P}_{ab}(c/2)$ is also an elementary matrix of the same type as $\mathbf{P}_{ab}(c)$. Therefore, using an approach similar to that in (C.20) - (C.26), we can write:

$$\mathbf{K}^{-1/2} = \left[\{\mathbf{F}^\gamma_1\}^{-1} \cdots \{\mathbf{F}^\gamma_{m-1}\}^{-1}\{\mathbf{F}^\gamma_m\}^{-1}\right], \tag{C.35}$$

where each $\mathbf{F}^\gamma$ matrix is of the form $\mathbf{P}_{ab}(c/2)$. Similar to the developments in (C.27), we can also write the linear transformation $\mathbf{Y}^* = \mathbf{K}^{-1/2}\mathbf{Y}\left(\mathbf{K}^{-1/2}\right)^T$ as:

$$\begin{aligned}\mathbf{Y}^* &= \left[\{\mathbf{F}^\gamma_1\}^{-1} \cdots \{\mathbf{F}^\gamma_{m-1}\}^{-1}\{\mathbf{F}^\gamma_m\}^{-1}\right] \mathbf{Y} \left[\{\mathbf{F}^\gamma_1\}^{-1} \cdots \{\mathbf{F}^\gamma_{m-1}\}^{-1}\{\mathbf{F}^\gamma_m\}^{-1}\right]^T \\ &= \left[\{\mathbf{F}^\gamma_1\}^{-1} \cdots \{\mathbf{F}^\gamma_{m-1}\}^{-1}\{\mathbf{F}^\gamma_m\}^{-1}\right] \mathbf{Y} \left[\{\mathbf{F}^\gamma_m\}^{-1}\right]^T \left[\{\mathbf{F}^\gamma_{m-1}\}^{-1}\right]^T \cdots \left[\{\mathbf{F}^\gamma_1\}^{-1}\right]^T \end{aligned} \tag{C.36}$$

Once again, using Lemma 1, we can write the Jacobian $J(\mathbf{Y} \to \mathbf{Y}^*)$ as:

$$J(\mathbf{Y} \to \mathbf{Y}^*) = J(\mathbf{Y} \to \mathbf{Y}^*_1)J(\mathbf{Y}^*_1 \to \mathbf{Y}^*_2)\cdots J(\mathbf{Y}^*_{m-1} \to \mathbf{Y}^*), \tag{C.37}$$

where $\mathbf{Y}^*_d = \{\mathbf{F}^\gamma_d\}^{-1}\mathbf{Y}^*_{d-1}\left[\{\mathbf{F}^\gamma_d\}^{-1}\right]^T$ $(d = 1, \ldots, m)$, $\mathbf{Y}^*_0 = \mathbf{Y}$, and $\mathbf{Y}^*_m = \mathbf{Y}^*$. Let $\mathbf{G}^\gamma$ represent any of the $\{\mathbf{F}^\gamma\}^{-1}$ matrices of the form $\mathbf{P}_{ab}(-c/2)$. Therefore, the transformation $\mathbf{Y}^* = \mathbf{G}^\gamma\mathbf{Y}^*_{d-1}\left(\mathbf{G}^\gamma\right)^T$ implies that $y^*_{aa} = y_{aa} - 2\left(\frac{c}{2}\right)y_{ab} + \left(-\frac{c}{2}\right)^2 y_{bb} = y_{aa} - cy_{ab} + \frac{c^2}{4}y_{bb}$, $y^*_{ac} = y^*_{ca} = y_{ac} - \left(\frac{c}{2}\right)y_{bc}(c \neq a)$, and $y^*_{bc} = y_{bc}(b, c \neq a)$. Similar to the developments demonstrated in the proof of Corollary 1, the matrix of partial derivatives will be an upper-triangular matrix with 1s on the main diagonal. Therefore, its determinant is equal to 1. Thus for this type of elementary transformation, $J(\mathbf{Y}^*_{d-1} \to \mathbf{Y}^*_d) = mod\ \det(\mathbf{G}^\gamma)^{(p+1)}$. Similarly, it follows that

$$\begin{aligned} J(\mathbf{Y} \to \mathbf{Y}^*) &= mod \left[\det\left(\{\mathbf{F}_1^\gamma\}^{-1}\right)^{p+1} \cdots \det\left(\{\mathbf{F}_{m-1}^\gamma\}^{-1}\right)^{p+1} \det\left(\{\mathbf{F}_m^\gamma\}^{-1}\right)^{p+1}\right] \\ &= mod \left[\det\left(\mathbf{F}_1^\gamma \cdots \mathbf{F}_{m-1}^\gamma \mathbf{F}_m^\gamma\right)^{-(p+1)}\right] \\ &= mod\ \det(\mathbf{K})^{-\frac{1}{2}(p+1)} \end{aligned} \tag{C.38}$$

□

Now, we can use these results in completing the derivation of the MGF for a mixture of Wishart distributions. Let us first return to the integral in (C.8):

$$g(\ell) = \int_{\mathbf{\Lambda}>\mathbf{0}} \operatorname{etr}(-\mathbf{\Lambda}\mathbf{L}) \det \mathbf{\Lambda}^{b-\frac{1}{2}(p+1)} d\mathbf{\Lambda}, \tag{C.39}$$

Let $\mathbf{J}^* = \mathbf{L}^{1/2}\mathbf{\Lambda}\mathbf{L}^{1/2}$. From Corollary 2 we know that

$$\begin{aligned} J(\mathbf{\Lambda} \to \mathbf{J}^*) &= mod\ \det(\mathbf{L})^{-\frac{1}{2}(p+1)} \\ &= \det(\mathbf{L})^{-\frac{1}{2}(p+1)} \end{aligned} \tag{C.40}$$

Also, working with $\mathbf{J}^*$ we have:

$$\mathbf{J}^* = \mathbf{L}^{1/2}\mathbf{\Lambda}\mathbf{L}^{1/2}$$

$$\mathbf{L}^{1/2}\mathbf{J}^*\mathbf{L}^{1/2} = \mathbf{L}^{1/2}\left(\mathbf{L}^{1/2}\mathbf{\Lambda}\mathbf{L}^{1/2}\right)\mathbf{L}^{1/2}$$

$$= \mathbf{L}\mathbf{\Lambda}\mathbf{L}$$

$$\mathbf{L}^{-1}\left(\mathbf{L}^{1/2}\mathbf{J}^*\mathbf{L}^{1/2}\right)\mathbf{L}^{-1} = \mathbf{L}^{-1}\left(\mathbf{L}\Lambda\mathbf{L}\right)\mathbf{L}^{-1}$$

$$\mathbf{L}^{-1/2}\mathbf{J}^*\mathbf{L}^{-1/2} = \mathbf{\Lambda} \tag{C.41}$$

$$\left(\mathbf{L}^{-1/2}\mathbf{J}^*\mathbf{L}^{-1/2}\right)\mathbf{L} = \mathbf{\Lambda}\mathbf{L}$$

$$\mathbf{L}^{-1/2}\mathbf{J}^*\mathbf{L}^{1/2} = \mathbf{\Lambda}\mathbf{L} \tag{C.42}$$

Substituting (C.40) - (C.42) into (C.38) we have:

$$\begin{aligned} g(\ell) &= \int_{\mathbf{J}^*>\mathbf{0}} \operatorname{etr}\left(-\mathbf{L}^{-1/2}\mathbf{J}^*\mathbf{L}^{1/2}\right) \det\left(\mathbf{L}^{-1/2}\mathbf{J}^*\mathbf{L}^{-1/2}\right)^{b-\frac{1}{2}(p+1)} \det(\mathbf{L})^{-\frac{1}{2}(p+1)} d\mathbf{J}^* \\ &= \det(\mathbf{L})^{-\frac{1}{2}(p+1)} \int_{\mathbf{J}^*>\mathbf{0}} \operatorname{etr}\left(-\mathbf{L}^{-1/2}\mathbf{J}^*\mathbf{L}^{1/2}\right) \det\left(\mathbf{L}^{-1/2}\mathbf{J}^*\mathbf{L}^{-1/2}\right)^{b-\frac{1}{2}(p+1)} d\mathbf{J}^* \end{aligned} \tag{C.43}$$

Because $\mathbf{L}^{-1/2}$ and $\mathbf{J}^*$ are square matrices of equal size, we have:

$$\begin{aligned} g(\ell) &= \det(\mathbf{L})^{-\frac{1}{2}(p+1)} \int_{\mathbf{J}^*>\mathbf{0}} \operatorname{etr}\left(-\mathbf{L}^{-1/2}\mathbf{J}^*\mathbf{L}^{1/2}\right) \det\left(\mathbf{L}^{-1}\right)^{b-\frac{1}{2}(p+1)} \det(\mathbf{J}^*)^{b-\frac{1}{2}(p+1)} d\mathbf{J}^* \\ &= \det(\mathbf{L})^{-\frac{1}{2}(p+1)} \det\left(\mathbf{L}^{-1}\right)^{b-\frac{1}{2}(p+1)} \int_{\mathbf{J}^*>\mathbf{0}} \operatorname{etr}\left(-\mathbf{L}^{-1/2}\mathbf{J}^*\mathbf{L}^{1/2}\right) \det(\mathbf{J}^*)^{b-\frac{1}{2}(p+1)} d\mathbf{J}^* \\ &= \det(\mathbf{L})^{-\frac{1}{2}(p+1)-b+\frac{1}{2}(p+1)} \int_{\mathbf{J}^*>\mathbf{0}} \operatorname{etr}\left(-\mathbf{L}^{-1/2}\mathbf{J}^*\mathbf{L}^{1/2}\right) \det(\mathbf{J}^*)^{b-\frac{1}{2}(p+1)} d\mathbf{J}^* \\ &= \det(\mathbf{L})^{-b} \int_{\mathbf{J}^*>\mathbf{0}} \operatorname{etr}\left(-\mathbf{L}^{-1/2}\mathbf{J}^*\mathbf{L}^{1/2}\right) \det(\mathbf{J}^*)^{b-\frac{1}{2}(p+1)} d\mathbf{J}^* \\ &= \det(\mathbf{L})^{-b} \int_{\mathbf{J}^*>\mathbf{0}} \operatorname{etr}\left(-\mathbf{J}^*\right) \det(\mathbf{J}^*)^{b-\frac{1}{2}(p+1)} d\mathbf{J}^* \\ &= \det(\mathbf{L})^{-b}\ \Gamma_p(b), \end{aligned} \tag{C.44}$$

where the last equality follows from Definition 1. Based on this result, let us return to the integral in (C.6):

$$\begin{aligned} I(\mathbf{A}^*) &= \int_{\mathbf{A}^*>0} \det\left(\mathbf{A}^*\right)^{(n_j-p-1)/2} \operatorname{etr}\left(-\frac{1}{2}\left[\mathbf{I}_p - 2\boldsymbol{\Theta}_j\boldsymbol{\Sigma}_j\right]\boldsymbol{\Sigma}_j^{-1}\mathbf{A}^*\right) d\mathbf{A}^* \\ &= \int_{\mathbf{A}^*>0} \det\left(\mathbf{A}^*\right)^{\frac{n_j}{2}-\frac{1}{2}(p+1)} \operatorname{etr}\left(-\frac{1}{2}\left[\mathbf{I}_p - 2\boldsymbol{\Theta}_j\boldsymbol{\Sigma}_j\right]\boldsymbol{\Sigma}_j^{-1}\mathbf{A}^*\right) d\mathbf{A}^* \end{aligned} \tag{C.45}$$

This integral is the same form as the one in (C.38). Therefore, it follows that the integral in (C.45) can be written as

$$I(\mathbf{A}^*) = \det\left(\frac{1}{2}\left[\mathbf{I}_p - 2\boldsymbol{\Theta}_j\boldsymbol{\Sigma}_j\right]\boldsymbol{\Sigma}_j^{-1}\right)^{-\frac{n_j}{2}} \Gamma_p\left(\frac{n_j}{2}\right) \tag{C.46}$$

Substituting (C.46) into (C.6) we now have:

$$
\begin{aligned}
M_{\mathbf{A}^*}(\boldsymbol{\Theta}) &= \sum_{j=1}^{k} w_j \left\{ 2^{(n_j p)/2} \Gamma_p \left(\frac{n_j}{2} \right) \det \left(\boldsymbol{\Sigma}_j \right)^{\frac{n_j}{2}} \right\}^{-1} \det \left(\frac{1}{2} \left[\mathbf{I}_p - 2 \boldsymbol{\Theta}_j \boldsymbol{\Sigma}_j \right] \boldsymbol{\Sigma}_j^{-1} \right)^{-\frac{n_j}{2}} \Gamma_p \left(\frac{n_j}{2} \right) \\
&= \sum_{j=1}^{k} w_j \left\{ 2^{(n_j p)/2} \det \left(\boldsymbol{\Sigma}_j \right)^{\frac{n_j}{2}} \right\}^{-1} \det \left(\frac{1}{2} \left[\mathbf{I}_p - 2 \boldsymbol{\Theta}_j \boldsymbol{\Sigma}_j \right] \boldsymbol{\Sigma}_j^{-1} \right)^{-\frac{n_j}{2}} \\
&= \sum_{j=1}^{k} w_j \left\{ \det \left(\boldsymbol{\Sigma}_j \right)^{\frac{n_j}{2}} \right\}^{-1} \det \left(\left[\mathbf{I}_p - 2 \boldsymbol{\Theta}_j \boldsymbol{\Sigma}_j \right] \boldsymbol{\Sigma}_j^{-1} \right)^{-\frac{n_j}{2}} \\
&= \sum_{j=1}^{k} w_j \left\{ \det \left(\boldsymbol{\Sigma}_j \right)^{\frac{n_j}{2}} \right\}^{-1} \det \left(\left[\mathbf{I}_p - 2 \boldsymbol{\Theta}_j \boldsymbol{\Sigma}_j \right] \right)^{-\frac{n_j}{2}} \det \left(\boldsymbol{\Sigma}_j \right)^{\frac{n_j}{2}} \\
&= \sum_{j=1}^{k} w_j \det \left(\mathbf{I}_p - 2 \boldsymbol{\Theta}_j \boldsymbol{\Sigma}_j \right)^{-\frac{n_j}{2}} \qquad \text{(C.47)}
\end{aligned}
$$

APPENDIX D

DERIVING THE MARGINAL DISTRIBUTION FROM A MIXTURE OF MULTIVARIATE GAUSSIAN DISTRIBUTIONS

Let the random variable $\mathbf{V}$ have the following distribution:

$$\mathbf{V} \sim \sum_{j=1}^{k} w_j \mathcal{N}_p \left(\boldsymbol{\mu}_j, \boldsymbol{\Sigma}_j\right) \tag{D.1}$$

Let us also assume that $\mathbf{V}, \boldsymbol{\mu}_j$, and $\boldsymbol{\Sigma}_j$ can be partitioned as follows:

$$\begin{aligned} \mathbf{V} &= \begin{pmatrix} \mathbf{v_1} \\ \mathbf{v_2} \end{pmatrix}, \\ \boldsymbol{\mu}_j &= \begin{pmatrix} \boldsymbol{\mu}_{1j} \\ \boldsymbol{\mu}_{2j} \end{pmatrix}, \\ \boldsymbol{\Sigma}_j &= \begin{pmatrix} (\boldsymbol{\Sigma}_j)_{\mathbf{11}} & (\boldsymbol{\Sigma}_j)_{\mathbf{12}} \\ (\boldsymbol{\Sigma}_j)_{\mathbf{21}} & (\boldsymbol{\Sigma}_j)_{\mathbf{22}} \end{pmatrix}, \end{aligned} \tag{D.2}$$

where $\mathbf{v_1}$ and $\mathbf{v_2}$ are two subvectors of dimension q_1 and q_2, respectively, with $q_1 + q_2 = p$. Similarly, $\boldsymbol{\mu}_{1j}$ and $\boldsymbol{\mu}_{2j}$ are also two subvectors of dimension q_1 and q_2, respectively. In (D.2), $\boldsymbol{\Sigma}_j$ is written as a partitioned matrix. For example, $(\boldsymbol{\Sigma}_j)_{\mathbf{11}}$ represents block 11 of $\boldsymbol{\Sigma}_j$. We also note that, due to symmetry, $\boldsymbol{\Sigma}_j = (\boldsymbol{\Sigma}_j)^T$, and $(\boldsymbol{\Sigma}_j)_{\mathbf{12}} = \left((\boldsymbol{\Sigma}_j)_{\mathbf{12}}\right)^T$. The joint density of $\mathbf{V}$ is:

$$\begin{aligned} f(\mathbf{V}) &= f(\mathbf{v_1}, \mathbf{v_2}) \\ &= \sum_{j=1}^{k} w_j \frac{1}{(2\pi)^{p/2} \left|\boldsymbol{\Sigma}_j\right|^{1/2}} \exp\left(-\frac{1}{2}\left(\mathbf{V} - \boldsymbol{\mu}_j\right)^T (\boldsymbol{\Sigma}_j)^{-1} \left(\mathbf{V} - \boldsymbol{\mu}_j\right)\right) \\ &= \sum_{j=1}^{k} w_j \frac{1}{(2\pi)^{p/2} \left|\boldsymbol{\Sigma}_j\right|^{1/2}} \exp\left(-\frac{1}{2}\mathbf{Q}_j(\mathbf{v_1}, \mathbf{v_2})\right), \end{aligned} \tag{D.3}$$

where $\mathbf{Q}_j\left(\mathbf{v_1}, \mathbf{v_2}\right)$ is defined as:

$$\begin{aligned} \mathbf{Q}_j(\mathbf{v_1}, \mathbf{v_2}) &= \left(\mathbf{V} - \boldsymbol{\mu}_j\right)^T (\boldsymbol{\Sigma}_j)^{-1} \left(\mathbf{V} - \boldsymbol{\mu}_j\right) \\ &= \left(\left(\mathbf{v_1} - \boldsymbol{\mu}_{1j}\right)^T \quad \left(\mathbf{v_2} - \boldsymbol{\mu}_{2j}\right)^T\right) \begin{pmatrix} (\boldsymbol{\Sigma}_j)^{\mathbf{11}} & (\boldsymbol{\Sigma}_j)^{\mathbf{12}} \\ (\boldsymbol{\Sigma}_j)^{\mathbf{21}} & (\boldsymbol{\Sigma}_j)^{\mathbf{22}} \end{pmatrix} \begin{pmatrix} \mathbf{v_1} - \boldsymbol{\mu}_{1j} \\ \mathbf{v_2} - \boldsymbol{\mu}_{2j} \end{pmatrix}, \end{aligned} \tag{D.4}$$

where $(\boldsymbol{\Sigma}_j)^{-1}$ can be written as the following partitioned matrix:

$$(\boldsymbol{\Sigma}_j)^{-1} = \begin{pmatrix} (\boldsymbol{\Sigma}_j)^{\mathbf{11}} & (\boldsymbol{\Sigma}_j)^{\mathbf{12}} \\ (\boldsymbol{\Sigma}_j)^{\mathbf{21}} & (\boldsymbol{\Sigma}_j)^{\mathbf{22}} \end{pmatrix} \tag{D.5}$$

Expanding the matrix multiplication in (D.4):

$$\begin{aligned} \mathbf{Q_j}\left(\mathbf{v_1}, \mathbf{v_2}\right) = {} & \left(\mathbf{v_1} - \boldsymbol{\mu}_{1j}\right)^T (\boldsymbol{\Sigma}_j)^{\mathbf{11}} \left(\mathbf{v_1} - \boldsymbol{\mu}_{1j}\right) + \\ & 2\left(\mathbf{v_1} - \boldsymbol{\mu}_{1j}\right)^T (\boldsymbol{\Sigma}_j)^{\mathbf{12}} \left(\mathbf{v_2} - \boldsymbol{\mu}_{2j}\right) + \\ & \left(\mathbf{v_2} - \boldsymbol{\mu}_{2j}\right)^T (\boldsymbol{\Sigma}_j)^{\mathbf{22}} \left(\mathbf{v_2} - \boldsymbol{\mu}_{2j}\right) \end{aligned} \tag{D.6}$$

We note that (D.6) follows from the symmetry of $(\boldsymbol{\Sigma}_j)^{-1}$. Before proceeding further, we will utilize the expression for the inverse of a partitioned symmetric matrix. Let us define the partitioned matrix, $\mathbf{M}$ as:

$$\mathbf{M}_{\mathbf{(n+m)\times(n+m)}} = \begin{pmatrix} \mathbf{M_{11}} & \mathbf{M_{12}} \\ \mathbf{M_{21}} & \mathbf{M_{22}} \end{pmatrix} \tag{D.7}$$

The following theorem is from Anderson (2003), although the proof presented here is different:

Theorem 11. (Anderson A.3.4 (2003)) Let the symmetric matrix, $\mathbf{M}$, be defined as in (D.7). Then, $\left(\mathbf{M_{11}} - \mathbf{M_{12}}\mathbf{M_{22}^{-1}}\mathbf{M_{21}}\right)^{-1} = \mathbf{M_{11}^{-1}} + \mathbf{M_{11}^{-1}}\mathbf{M_{12}}\left(\mathbf{M_{22}} - \mathbf{M_{21}}\mathbf{M_{11}^{-1}}\mathbf{M_{12}}\right)^{-1}\mathbf{M_{21}}\mathbf{M_{11}^{-1}}$.

Proof. Assume the equality in Theorem 10 is true. Therefore, after pre-multiplying each side of the equation by $\left(\mathbf{M_{11}} - \mathbf{M_{12}}\mathbf{M_{22}^{-1}}\mathbf{M_{21}}\right)$:

$$
\begin{aligned}
\mathbf{I} &= \left(\mathbf{M_{11}} - \mathbf{M_{12}}\mathbf{M_{22}^{-1}}\mathbf{M_{21}}\right)\left(\mathbf{M_{11}^{-1}} + \mathbf{M_{11}^{-1}}\mathbf{M_{12}}\left(\mathbf{M_{22}} - \mathbf{M_{21}}\mathbf{M_{11}^{-1}}\mathbf{M_{12}}\right)^{-1}\mathbf{M_{21}}\mathbf{M_{11}^{-1}}\right)\\
\mathbf{I} &= \left(\mathbf{M_{11}} - \mathbf{M_{12}}\mathbf{M_{22}^{-1}}\mathbf{M_{21}}\right)\mathbf{M_{11}^{-1}} + \\
&\quad \left(\mathbf{M_{11}} - \mathbf{M_{12}}\mathbf{M_{22}^{-1}}\mathbf{M_{21}}\right)\mathbf{M_{11}^{-1}}\mathbf{M_{12}}\left(\mathbf{M_{22}} - \mathbf{M_{21}}\mathbf{M_{11}^{-1}}\mathbf{M_{12}}\right)^{-1}\mathbf{M_{21}}\mathbf{M_{11}^{-1}}\\
\mathbf{I} &= \mathbf{I} - \mathbf{M_{12}}\mathbf{M_{22}^{-1}}\mathbf{M_{21}}\mathbf{M_{11}^{-1}} \ + \ \mathbf{M_{12}}\left(\mathbf{M_{22}} - \mathbf{M_{21}}\mathbf{M_{11}^{-1}}\mathbf{M_{12}}\right)^{-1}\mathbf{M_{21}}\mathbf{M_{11}^{-1}} - \\
&\quad \mathbf{M_{12}}\mathbf{M_{22}^{-1}}\mathbf{M_{21}}\mathbf{M_{11}^{-1}}\mathbf{M_{12}}\mathbf{M_{12}}\left(\mathbf{M_{22}} - \mathbf{M_{21}}\mathbf{M_{11}^{-1}}\mathbf{M_{12}}\right)^{-1}\mathbf{M_{21}}\mathbf{M_{11}^{-1}}\\
\mathbf{I} &= \mathbf{I} - \mathbf{M_{12}}\mathbf{M_{22}^{-1}}\mathbf{M_{21}}\mathbf{M_{11}^{-1}} + \\
&\quad \left(\mathbf{M_{12}} - \mathbf{M_{12}}\mathbf{M_{22}^{-1}}\mathbf{M_{21}}\mathbf{M_{11}^{-1}}\mathbf{M_{12}}\right)\left(\mathbf{M_{22}} - \mathbf{M_{21}}\mathbf{M_{11}^{-1}}\mathbf{M_{12}}\right)^{-1}\mathbf{M_{21}}\mathbf{M_{11}^{-1}}\\
\mathbf{I} &= \mathbf{I} - \mathbf{M_{12}}\mathbf{M_{22}^{-1}}\mathbf{M_{21}}\mathbf{M_{11}^{-1}} + \\
&\quad \mathbf{M_{12}}\mathbf{M_{22}^{-1}}\left(\mathbf{M_{22}} - \mathbf{M_{21}}\mathbf{M_{11}^{-1}}\mathbf{M_{12}}\right)^{-1}\left(\mathbf{M_{22}} - \mathbf{M_{21}}\mathbf{M_{11}^{-1}}\mathbf{M_{12}}\right)\mathbf{M_{21}}\mathbf{M_{11}^{-1}}\\
\mathbf{I} &= \mathbf{I} - \mathbf{M_{12}}\mathbf{M_{22}^{-1}}\mathbf{M_{21}}\mathbf{M_{11}^{-1}} + \mathbf{M_{12}}\mathbf{M_{22}^{-1}}\mathbf{M_{21}}\mathbf{M_{11}^{-1}}\\
\mathbf{I} &= \mathbf{I}
\end{aligned}
$$

Therefore, the result has been proven assuming that $\mathbf{M_{11}}$, $\mathbf{M_{22}}$, and $\mathbf{M_{22}} - \mathbf{M_{21}}\mathbf{M_{11}^{-1}}\mathbf{M_{12}}$ are invertible matrices. □

Now, let the symmetric matrix $\mathbf{E_{n\times n}}$ be defined as:

$$
\mathbf{E} = \begin{pmatrix} \mathbf{E_{11}} & \mathbf{E_{12}} \\ \mathbf{E_{21}} & \mathbf{E_{22}} \end{pmatrix} = \begin{pmatrix} \mathbf{E_{11}} & \mathbf{E_{12}} \\ \left(\mathbf{E_{12}}\right)^T & \mathbf{E_{22}} \end{pmatrix} \tag{D.8}
$$

Similarly, let $\mathbf{F_{n\times n}} = \mathbf{E}^{-1}$ be defined as:

$$
\mathbf{F} = \mathbf{E}^{-1} = \begin{pmatrix} \mathbf{F_{11}} & \mathbf{F_{12}} \\ \mathbf{F_{21}} & \mathbf{F_{22}} \end{pmatrix} = \begin{pmatrix} \mathbf{F_{11}} & \mathbf{F_{12}} \\ \left(\mathbf{F_{12}}\right)^T & \mathbf{F_{22}} \end{pmatrix} \tag{D.9}
$$

Further, let us assume the dimensions of the respective blocks of the matrices in (D.8) and (D.9) are:

- $\mathbf{E_{11}}$ and $\mathbf{F_{11}}$ are $r \times r$
- $\mathbf{E_{22}}$ and $\mathbf{F_{22}}$ are $s \times s$
- $\mathbf{E_{12}} = (\mathbf{E_{21}})^T$ and $\mathbf{F_{12}} = (\mathbf{F_{21}})^T$ are both $r \times s$, with $r + s = n$. Then we have the following theorem:

Theorem 12. Let the symmetric matrices $\mathbf{E}$ and $\mathbf{F}$ be defined as in (D.8) and (D.9). Then, we have the following expressions:

$$\begin{aligned}
\mathbf{F_{11}} &= \left(\mathbf{E_{11}} - \mathbf{E_{12}}\left(\mathbf{E_{22}}\right)^{-1}\left(\mathbf{E_{12}}\right)^T\right)^{-1} \\
&= \left(\mathbf{E_{11}}\right)^{-1} + \left(\mathbf{E_{11}}\right)^{-1}\mathbf{E_{12}}\left(\mathbf{E_{22}} - \left(\mathbf{E_{12}}\right)^T\left(\mathbf{E_{11}}\right)^{-1}\mathbf{E_{12}}\right)^{-1}\left(\mathbf{E_{12}}\right)^T\left(\mathbf{E_{11}}\right)^{-1} &\text{(D.10)} \\
\mathbf{F_{22}} &= \left(\mathbf{E_{22}} - \left(\mathbf{E_{12}}\right)^T\left(\mathbf{E_{11}}\right)^{-1}\mathbf{E_{12}}\right)^{-1} \\
&= \left(\mathbf{E_{22}}\right)^{-1} + \left(\mathbf{E_{22}}\right)^{-1}\left(\mathbf{E_{12}}\right)^T\left(\mathbf{E_{11}} - \mathbf{E_{12}}\left(\mathbf{E_{22}}\right)^{-1}\left(\mathbf{E_{12}}\right)^T\right)^{-1}\mathbf{E_{12}}\left(\mathbf{E_{22}}\right)^{-1} &\text{(D.11)} \\
\left(\mathbf{F_{12}}\right)^T &= -\left(\mathbf{E_{22}}\right)^{-1}\left(\mathbf{E_{12}}\right)^T\left(\mathbf{E_{11}} - \mathbf{E_{12}}\left(\mathbf{E_{22}}\right)^{-1}\left(\mathbf{E_{12}}\right)^T\right)^{-1} &\text{(D.12)} \\
\mathbf{F_{12}} &= -\left(\mathbf{E_{11}}\right)^{-1}\mathbf{E_{12}}\left(\mathbf{E_{22}} - \left(\mathbf{E_{12}}\right)^T\left(\mathbf{E_{11}}\right)^{-1}\mathbf{E_{12}}\right)^{-1} &\text{(D.13)}
\end{aligned}$$

Proof. Based on the above definitions, we have:

$$\begin{aligned}
\mathbf{I_n} = \mathbf{E}\mathbf{E}^{-1} = \mathbf{E}\mathbf{F} &= \begin{pmatrix} \mathbf{E_{11}} & \mathbf{E_{12}} \\ \left(\mathbf{E_{12}}\right)^T & \mathbf{E_{22}} \end{pmatrix}\begin{pmatrix} \mathbf{F_{11}} & \mathbf{F_{12}} \\ \left(\mathbf{F_{12}}\right)^T & \mathbf{F_{22}} \end{pmatrix} \\
&= \begin{pmatrix} \mathbf{E_{11}}\mathbf{F_{11}} + \mathbf{E_{12}}\left(\mathbf{F_{12}}\right)^T & \mathbf{E_{11}}\mathbf{F_{12}} + \mathbf{E_{12}}\mathbf{F_{22}} \\ \left(\mathbf{E_{12}}\right)^T\mathbf{F_{11}} + \mathbf{E_{22}}\left(\mathbf{F_{12}}\right)^T & \left(\mathbf{E_{12}}\right)^T\mathbf{F_{12}} + \mathbf{E_{22}}\mathbf{F_{22}} \end{pmatrix} &\text{(D.14)} \\
&= \begin{pmatrix} \mathbf{I_r} & \mathbf{0_r} \\ \mathbf{0_s} & \mathbf{I_s} \end{pmatrix} &\text{(D.15)}
\end{aligned}$$

Thus, using (D.14) - (D.15), we now have the following for $\mathbf{F_{11}}$:

$$\begin{aligned}
\mathbf{E_{11}}\mathbf{F_{11}} + \mathbf{E_{12}}\left(\mathbf{F_{12}}\right)^T &= \mathbf{I_r} \\
\mathbf{E_{11}}\mathbf{F_{11}} &= \mathbf{I_r} - \mathbf{E_{12}}\left(\mathbf{F_{12}}\right)^T \\
\mathbf{F_{11}} &= \left(\mathbf{E_{11}}\right)^{-1} - \left(\mathbf{E_{11}}\right)^{-1}\mathbf{E_{12}}\left(\mathbf{F_{12}}\right)^T &\text{(D.16)}
\end{aligned}$$

For $\mathbf{F_{12}}$:

$$\begin{aligned}
\mathbf{E_{11}F_{12}} + \mathbf{E_{12}F_{22}} &= \mathbf{0_r} \\
\mathbf{E_{11}F_{12}} &= -\mathbf{E_{12}F_{22}} \\
\mathbf{F_{12}} &= -\left(\mathbf{E_{11}}\right)^{-1}\mathbf{E_{12}F_{22}}
\end{aligned} \tag{D.17}$$

For $\left(\mathbf{F_{12}}\right)^T$:

$$\begin{aligned}
\left(\mathbf{E_{12}}\right)^T\mathbf{F_{11}} + \mathbf{E_{22}}\left(\mathbf{F_{12}}\right)^T &= \mathbf{0_s} \\
\mathbf{E}_{22}\left(\mathbf{F}_{12}\right)^T &= -\left(\mathbf{E}_{12}\right)^T\mathbf{F}_{11} \\
\left(\mathbf{F_{12}}\right)^T &= -\left(\mathbf{E_{22}}\right)^{-1}\left(\mathbf{E_{12}}\right)^T\mathbf{F_{11}}
\end{aligned} \tag{D.18}$$

for $\mathbf{F_{22}}$:

$$\begin{aligned}
\left(\mathbf{E_{12}}\right)^T\mathbf{F_{12}} + \mathbf{E_{22}F_{22}} &= \mathbf{I_s} \\
\mathbf{E_{22}F_{22}} &= \mathbf{I_s} - \left(\mathbf{E_{12}}\right)^T\mathbf{F_{12}} \\
\mathbf{F_{22}} &= \left(\mathbf{E_{22}}\right)^{-1} - \left(\mathbf{E_{22}}\right)^{-1}\left(\mathbf{E_{12}}\right)^T\mathbf{F_{12}}
\end{aligned} \tag{D.19}$$

Now, substitute (D.18) into (D.16):

$$\begin{aligned}
\mathbf{F_{11}} &= \left(\mathbf{E_{11}}\right)^{-1} - \left(\mathbf{E_{11}}\right)^{-1}\mathbf{E_{12}}\left(\mathbf{F_{12}}\right)^T \\
\mathbf{F_{11}} &= \left(\mathbf{E_{11}}\right)^{-1} - \left(\mathbf{E_{11}}\right)^{-1}\mathbf{E_{12}}\left(-\left(\mathbf{E_{22}}\right)^{-1}\left(\mathbf{E_{12}}\right)^T\mathbf{F_{11}}\right) \\
\mathbf{F_{11}} &= \left(\mathbf{E_{11}}\right)^{-1} + \left(\mathbf{E_{11}}\right)^{-1}\mathbf{E_{12}}\left(\mathbf{E_{22}}\right)^{-1}\left(\mathbf{E_{12}}\right)^T\mathbf{F_{11}} \\
\left(\mathbf{I_r} - \left(\mathbf{E_{11}}\right)^{-1}\mathbf{E_{12}}\left(\mathbf{E_{22}}\right)^{-1}\left(\mathbf{E_{12}}\right)^T\right)\mathbf{F_{11}} &= \left(\mathbf{E_{11}}\right)^{-1} \\
\left(\mathbf{E_{11}} - \mathbf{E_{12}}\left(\mathbf{E_{22}}\right)^{-1}\left(\mathbf{E_{12}}\right)^T\right)\mathbf{F_{11}} &= \mathbf{I_r} \\
\mathbf{F_{11}} &= \left(\mathbf{E_{11}} - \mathbf{E_{12}}\left(\mathbf{E_{22}}\right)^{-1}\left(\mathbf{E_{12}}\right)^T\right)^{-1}
\end{aligned} \tag{D.20}$$

Applying Theorem 10 to (D.20), we obtain:

$$\begin{aligned}
\mathbf{F_{11}} &= \left(\mathbf{E_{11}} - \mathbf{E_{12}}\left(\mathbf{E_{22}}\right)^{-1}\left(\mathbf{E_{12}}\right)^T\right)^{-1} \\
&= \left(\mathbf{E_{11}}\right)^{-1} + \left(\mathbf{E_{11}}\right)^{-1}\mathbf{E_{12}}\left(\mathbf{E_{22}} - \left(\mathbf{E_{12}}\right)^T\left(\mathbf{E_{11}}\right)^{-1}\mathbf{E_{12}}\right)^{-1}\left(\mathbf{E_{12}}\right)^T\left(\mathbf{E_{11}}\right)^{-1}
\end{aligned} \tag{D.21}$$

Now, substitute (D.20) into (D.18):

$$\begin{aligned}(\mathbf{F_{12}})^T &= -(\mathbf{E_{22}})^{-1}(\mathbf{E_{12}})^T\mathbf{F_{11}} \\ &= -(\mathbf{E_{22}})^{-1}(\mathbf{E_{12}})^T\left(\mathbf{E_{11}} - \mathbf{E_{12}}(\mathbf{E_{22}})^{-1}(\mathbf{E_{12}})^T\right)^{-1}\end{aligned} \tag{D.22}$$

Similarly, substitute (D.17) into (D.19) to obtain:

$$\begin{aligned}\mathbf{F_{22}} &= (\mathbf{E_{22}})^{-1} - (\mathbf{E_{22}})^{-1}(\mathbf{E_{12}})^T\mathbf{F_{12}} \\ \mathbf{F_{22}} &= (\mathbf{E_{22}})^{-1} - (\mathbf{E_{22}})^{-1}(\mathbf{E_{12}})^T\left(-(\mathbf{E_{11}})^{-1}\mathbf{E_{12}}\mathbf{F_{22}}\right) \\ \mathbf{F_{22}} &= (\mathbf{E_{22}})^{-1} + (\mathbf{E_{22}})^{-1}(\mathbf{E_{12}})^T(\mathbf{E_{11}})^{-1}\mathbf{E_{12}}\mathbf{F_{22}} \\ \left(\mathbf{I_s} - (\mathbf{E_{22}})^{-1}(\mathbf{E_{12}})^T(\mathbf{E_{11}})^{-1}\mathbf{E_{12}}\right)\mathbf{F_{22}} &= (\mathbf{E_{22}})^{-1} \\ \left(\mathbf{E_{22}} - (\mathbf{E_{12}})^T(\mathbf{E_{11}})^{-1}\mathbf{E_{12}}\right)\mathbf{F_{22}} &= \mathbf{I_s} \\ \mathbf{F_{22}} &= \left(\mathbf{E_{22}} - (\mathbf{E_{12}})^T(\mathbf{E_{11}})^{-1}\mathbf{E_{12}}\right)^{-1}\end{aligned} \tag{D.23}$$

Applying Theorem 10 to (D.23), we now have:

$$\mathbf{F_{22}} = (\mathbf{E_{22}})^{-1} + (\mathbf{E_{22}})^{-1}(\mathbf{E_{12}})^T\left(\mathbf{E_{11}} - \mathbf{E_{12}}(\mathbf{E_{22}})^{-1}(\mathbf{E_{12}})^T\right)^{-1}\mathbf{E_{12}}(\mathbf{E_{22}})^{-1} \tag{D.24}$$

Finally, substitute (D.23) into (D.17):

$$\begin{aligned}\mathbf{F_{12}} &= -(\mathbf{E_{11}})^{-1}\mathbf{E_{12}}\mathbf{F_{22}} \\ &= -(\mathbf{E_{11}})^{-1}\mathbf{E_{12}}\left(\mathbf{E_{22}} - (\mathbf{E_{12}})^T(\mathbf{E_{11}})^{-1}\mathbf{E_{12}}\right)^{-1}\end{aligned} \tag{D.25}$$

Thus, based on (D.20) - (D.25), the expressions in (D.10) - (D.13) have been verified and the proof is complete. □

Returning to (D.6) and using Theorem 4, we now have the following expressions:

$$\begin{aligned}(\boldsymbol{\Sigma}_j)^{\mathbf{11}} &= \left((\boldsymbol{\Sigma}_j)_{\mathbf{11}} - (\boldsymbol{\Sigma}_j)_{\mathbf{12}}\left((\boldsymbol{\Sigma}_j)_{\mathbf{22}}\right)^{-1}\left((\boldsymbol{\Sigma}_j)_{\mathbf{12}}\right)^{T}\right)^{-1} \\ &= \left((\boldsymbol{\Sigma}_j)_{\mathbf{11}}\right)^{-1} + \left((\boldsymbol{\Sigma}_j)_{\mathbf{11}}\right)^{-1}(\boldsymbol{\Sigma}_j)_{\mathbf{12}} \times \\ &\quad \left((\boldsymbol{\Sigma}_j)_{\mathbf{22}} - \left((\boldsymbol{\Sigma}_j)_{\mathbf{12}}\right)^{T}\left((\boldsymbol{\Sigma}_j)_{\mathbf{11}}\right)^{-1}(\boldsymbol{\Sigma}_j)_{\mathbf{12}}\right)^{-1}\left((\boldsymbol{\Sigma}_j)_{\mathbf{12}}\right)^{T}\left((\boldsymbol{\Sigma}_j)_{\mathbf{11}}\right)^{-1}\end{aligned} \tag{D.26}$$

$$\begin{aligned}(\boldsymbol{\Sigma}_j)^{\mathbf{22}} &= \left((\boldsymbol{\Sigma}_j)_{\mathbf{22}} - \left((\boldsymbol{\Sigma}_j)_{\mathbf{12}}\right)^{T}\left((\boldsymbol{\Sigma}_j)_{\mathbf{11}}\right)^{-1}(\boldsymbol{\Sigma}_j)_{\mathbf{12}}\right)^{-1} \\ &= \left((\boldsymbol{\Sigma}_j)_{\mathbf{22}}\right)^{-1} + \left((\boldsymbol{\Sigma}_j)_{\mathbf{22}}\right)^{-1}\left((\boldsymbol{\Sigma}_j)_{\mathbf{12}}\right)^{T} \times \\ &\quad \left((\boldsymbol{\Sigma}_j)_{\mathbf{11}} - (\boldsymbol{\Sigma}_j)_{\mathbf{12}}\left((\boldsymbol{\Sigma}_j)_{\mathbf{22}}\right)^{-1}\left((\boldsymbol{\Sigma}_j)_{\mathbf{12}}\right)^{T}\right)^{-1}(\boldsymbol{\Sigma}_j)_{\mathbf{12}}\left((\boldsymbol{\Sigma}_j)_{\mathbf{22}}\right)^{-1}\end{aligned} \tag{D.27}$$

$$(\boldsymbol{\Sigma}_j)^{\mathbf{12}} = \left((\boldsymbol{\Sigma}_j)^{\mathbf{21}}\right)^{T} = -\left((\boldsymbol{\Sigma}_j)_{\mathbf{11}}\right)^{-1}(\boldsymbol{\Sigma}_j)_{\mathbf{12}}\left((\boldsymbol{\Sigma}_j)_{\mathbf{22}} - \left((\boldsymbol{\Sigma}_j)_{\mathbf{12}}\right)^{T}\left((\boldsymbol{\Sigma}_j)_{\mathbf{11}}\right)^{-1}(\boldsymbol{\Sigma}_j)_{\mathbf{12}}\right)^{-1} \tag{D.28}$$

Substituting (D.26) - (D.28) into (D.6), we have:

$$\begin{aligned}\mathbf{Q}_j(\mathbf{v_1}, \mathbf{v_2}) = {} & \left(\mathbf{v_1} - \boldsymbol{\mu}_{1j}\right)^{T}\left((\boldsymbol{\Sigma}_j)_{\mathbf{11}}\right)^{-1}\left(\mathbf{v_1} - \boldsymbol{\mu}_{1j}\right) + \left(\mathbf{v_1} - \boldsymbol{\mu}_{1j}\right)^{T} \times \\ & \left[\left((\boldsymbol{\Sigma}_j)_{\mathbf{11}}\right)^{-1}(\boldsymbol{\Sigma}_j)_{\mathbf{12}}\left((\boldsymbol{\Sigma}_j)_{\mathbf{22}} - \left((\boldsymbol{\Sigma}_j)_{\mathbf{12}}\right)^{T}\left((\boldsymbol{\Sigma}_j)_{\mathbf{11}}\right)^{-1}(\boldsymbol{\Sigma}_j)_{\mathbf{12}}\right)^{-1}\left((\boldsymbol{\Sigma}_j)_{\mathbf{12}}\right)^{T}\left((\boldsymbol{\Sigma}_j)_{\mathbf{11}}\right)^{-1}\right] \times \\ & \left(\mathbf{v_1} - \boldsymbol{\mu}_{1j}\right) \\ & - 2\left(\mathbf{v_1} - \boldsymbol{\mu}_{1j}\right)^{T} \times \\ & \left[\left((\boldsymbol{\Sigma}_j)_{\mathbf{11}}\right)^{-1}(\boldsymbol{\Sigma}_j)_{\mathbf{12}}\left((\boldsymbol{\Sigma}_j)_{\mathbf{22}} - \left((\boldsymbol{\Sigma}_j)_{\mathbf{12}}\right)^{T}\left((\boldsymbol{\Sigma}_j)_{\mathbf{11}}\right)^{-1}(\boldsymbol{\Sigma}_j)_{\mathbf{12}}\right)^{-1}\right]\left(\mathbf{v_2} - \boldsymbol{\mu}_{2j}\right) \\ & + \left(\mathbf{v_2} - \boldsymbol{\mu}_{2j}\right)^{T}\left[(\boldsymbol{\Sigma}_j)_{\mathbf{22}} - \left((\boldsymbol{\Sigma}_j)_{\mathbf{12}}\right)^{T}\left((\boldsymbol{\Sigma}_j)_{\mathbf{11}}\right)^{-1}(\boldsymbol{\Sigma}_j)_{\mathbf{12}}\right]^{-1}\left(\mathbf{v_2} - \boldsymbol{\mu}_{2j}\right)\end{aligned} \tag{D.29}$$

Before proceeding further, the following theorem will be useful for a given symmetric matrix $\mathbf{A}$ and any two vectors $\mathbf{c}$ and $\mathbf{d}$:

Theorem 13. Let $\mathbf{A}$ be a symmetric matrix of dimension $m \times m$ and let $\mathbf{c}$ and $\mathbf{d}$ be two vectors, each of dimension $m \times 1$. Assuming conformability, we have the following:

$$(\mathbf{c} - \mathbf{d})^{T}\mathbf{A}(\mathbf{c} - \mathbf{d}) = (\mathbf{d} - \mathbf{c})^{T}\mathbf{A}(\mathbf{d} - \mathbf{c}) \tag{D.30}$$

Proof.

$$\mathbf{c}^T\mathbf{A}\mathbf{c} - 2\mathbf{c}^T\mathbf{A}\mathbf{d} + \mathbf{d}^T\mathbf{A}\mathbf{d} = \mathbf{c}^T\mathbf{A}\mathbf{c} - \mathbf{c}^T\mathbf{A}\mathbf{d} - \mathbf{c}^T\mathbf{A}\mathbf{d} + \mathbf{d}^T\mathbf{A}\mathbf{d}$$

$$\mathbf{c}^T\mathbf{A}\mathbf{c} - 2\mathbf{c}^T\mathbf{A}\mathbf{d} + \mathbf{d}^T\mathbf{A}\mathbf{d} = \mathbf{c}^T\mathbf{A}\mathbf{c} - \mathbf{c}^T\mathbf{A}\mathbf{d} - \mathbf{d}^T\mathbf{A}\mathbf{c} + \mathbf{d}^T\mathbf{A}\mathbf{d} \tag{D.31}$$

The r.h.s. of (D.31) follows from the following:

$$\langle \mathbf{c}, \mathbf{A}\mathbf{d} \rangle \; = \mathbf{c}^T\mathbf{A}\mathbf{d} = \mathbf{c}^T\mathbf{A}^T\mathbf{d} = \left(\underbrace{\mathbf{d}^T\mathbf{A}\mathbf{c}}_{\text{scalar}} \right)^T = \mathbf{d}^T\mathbf{A}\mathbf{c} \tag{D.32}$$

We note that from (D.32) that $< . >$ is defined as the inner product. Further, (D.32) utilizes the fact that $\mathbf{A}$ is a symmetric matrix. Continuing the development of (D.31), we now have:

$$\mathbf{c}^T\mathbf{A}\left(\mathbf{c} - \mathbf{d}\right) - \left(\mathbf{c} - \mathbf{d}\right)^T\mathbf{A}\mathbf{d} = \mathbf{c}^T\mathbf{A}\left(\mathbf{c} - \mathbf{d}\right) - \mathbf{d}^T\mathbf{A}\left(\mathbf{c} - \mathbf{d}\right) \tag{D.33}$$

Applying the method in (D.32) to (D.33), we now have:

$$\left(\mathbf{c} - \mathbf{d}\right)^T\mathbf{A}\mathbf{c} - \left(\mathbf{c} - \mathbf{d}\right)^T\mathbf{A}\mathbf{d} = -\mathbf{c}^T\mathbf{A}\left(\mathbf{d} - \mathbf{c}\right) + \mathbf{d}^T\mathbf{A}\left(\mathbf{d} - \mathbf{c}\right)$$

$$\left(\mathbf{c} - \mathbf{d}\right)^T\mathbf{A}\left(\mathbf{c} - \mathbf{d}\right) = \left(\mathbf{d} - \mathbf{c}\right)^T\mathbf{A}\left(\mathbf{d} - \mathbf{c}\right) \tag{D.34}$$

□

Applying Theorem 12 to (D.29), we now have:

$$\begin{aligned}
\mathbf{Q}_j\left(\mathbf{v_1}, \mathbf{v_2}\right) = {} & \left(\mathbf{v_1} - \boldsymbol{\mu}_{1j}\right)^T \left(\left(\boldsymbol{\Sigma}_j\right)_{\mathbf{11}}\right)^{-1} \left(\mathbf{v_1} - \boldsymbol{\mu}_{1j}\right) + \\
& \left(\left(\mathbf{v_2} - \boldsymbol{\mu}_{2j}\right) - \left(\left(\boldsymbol{\Sigma}_j\right)_{\mathbf{12}}\right)^T \left(\left(\boldsymbol{\Sigma}_j\right)_{\mathbf{11}}\right)^{-1} \left(\mathbf{v_1} - \boldsymbol{\mu}_{1j}\right)\right)^T \times \\
& \left(\left(\boldsymbol{\Sigma}_j\right)_{\mathbf{22}} - \left(\left(\boldsymbol{\Sigma}_j\right)_{\mathbf{12}}\right)^T \left(\left(\boldsymbol{\Sigma}_j\right)_{\mathbf{11}}\right)^{-1} \left(\boldsymbol{\Sigma}_j\right)_{\mathbf{12}}\right)^{-1} \times \\
& \left(\left(\mathbf{v_2} - \boldsymbol{\mu}_{2j}\right) - \left(\left(\boldsymbol{\Sigma}_j\right)_{\mathbf{12}}\right)^T \left(\left(\boldsymbol{\Sigma}_j\right)_{\mathbf{11}}\right)^{-1} \left(\mathbf{v_1} - \boldsymbol{\mu}_{1j}\right)\right) && \text{(D.35)} \\
= {} & \mathbf{Q}_{1j}\left(\mathbf{v_1}\right) + \mathbf{Q}_{2j}\left(\mathbf{v_1}, \mathbf{v_2}\right) && \text{(D.36)}
\end{aligned}$$

where

$$
\begin{aligned}
\mathbf{Q}_{1j}\left(\mathbf{v_1}\right) &= \left(\mathbf{v_1} - \boldsymbol{\mu}_{1j}\right)^T \left(\left(\boldsymbol{\Sigma}_j\right)_{\mathbf{11}}\right)^{-1} \left(\mathbf{v_1} - \boldsymbol{\mu}_{1j}\right) \\
\mathbf{Q}_{2j}\left(\mathbf{v_1}, \mathbf{v_2}\right) &= \left(\left(\mathbf{v_2} - \boldsymbol{\mu}_{2j}\right) - \left(\left(\boldsymbol{\Sigma}_j\right)_{\mathbf{12}}\right)^T \left(\left(\boldsymbol{\Sigma}_j\right)_{\mathbf{11}}\right)^{-1} \left(\mathbf{v_1} - \boldsymbol{\mu}_{1j}\right)\right)^T \times \\
&\quad \left(\left(\boldsymbol{\Sigma}_j\right)_{\mathbf{22}} - \left(\left(\boldsymbol{\Sigma}_j\right)_{\mathbf{12}}\right)^T \left(\left(\boldsymbol{\Sigma}_j\right)_{\mathbf{11}}\right)^{-1} \left(\boldsymbol{\Sigma}_j\right)_{\mathbf{12}}\right)^{-1} \times \\
&\quad \left(\left(\mathbf{v_2} - \boldsymbol{\mu}_{2j}\right) - \left(\left(\boldsymbol{\Sigma}_j\right)_{\mathbf{12}}\right)^T \left(\left(\boldsymbol{\Sigma}_j\right)_{\mathbf{11}}\right)^{-1} \left(\mathbf{v_1} - \boldsymbol{\mu}_{1j}\right)\right)
\end{aligned}
$$

Now, returning to (D.3) and using (D.35) - (D.36), we can write the distribution of $\mathbf{V}$ as:

$$
\begin{aligned}
f(\mathbf{V}) &= f(\mathbf{v_1}, \mathbf{v_2}) \\
&= \sum_{j=1}^{k} w_j \frac{1}{(2\pi)^{p/2} \left|\boldsymbol{\Sigma}_j\right|^{1/2}} \exp\left(-\frac{1}{2}\left(\mathbf{V} - \boldsymbol{\mu}_j\right)^T \left(\boldsymbol{\Sigma}_j\right)^{-1} \left(\mathbf{V} - \boldsymbol{\mu}_j\right)\right) \\
&= \sum_{j=1}^{k} w_j \frac{1}{(2\pi)^{p/2} \left|\boldsymbol{\Sigma}_j\right|^{1/2}} \exp\left(-\frac{1}{2}\mathbf{Q}_j(\mathbf{v_1}, \mathbf{v_2})\right) \\
&= \sum_{j=1}^{k} w_j \frac{1}{(2\pi)^{p/2} \left|\boldsymbol{\Sigma}_j\right|^{1/2}} \exp\left(-\frac{1}{2}\left[\mathbf{Q}_{1j}(\mathbf{v_1}) + \mathbf{Q}_{2j}(\mathbf{v_{1j}}, \mathbf{v_{2j}})\right]\right) \qquad \text{(D.37)}
\end{aligned}
$$

For the next step in the development, the following theorem adapted from Anderson (2003) will be useful. This theorem from Anderson (2003) is stated slightly differently and presents a slightly different proof than what follows.

Theorem 14. (Anderson A.3.2 (2003)). Let the symmetric matrix $\mathbf{E}$ be defined as in (D.8). Then:

$\det\left(\mathbf{E}\right) = \det\left(\mathbf{E_{11}}\right) \det\left(\mathbf{E_{22}} - \left(\mathbf{E_{12}}\right)^T \left(\mathbf{E_{11}}\right)^{-1} \mathbf{E_{12}}\right)$.

Proof. As in (D.8):

$$
\begin{aligned}
\mathbf{E} &= \begin{pmatrix} \mathbf{E_{11}} & \mathbf{E_{12}} \\ \mathbf{E_{21}} & \mathbf{E_{22}} \end{pmatrix} \\
&= \begin{pmatrix} \mathbf{E_{11}} & \mathbf{0_r} \\ \left(\mathbf{E_{12}}\right)^T & \mathbf{I_s} \end{pmatrix} \begin{pmatrix} \mathbf{I_r} & \left(\mathbf{E}_{11}\right)^{-1} \mathbf{E_{12}} \\ \mathbf{0_r} & \mathbf{E_{22}} - \left(\mathbf{E_{12}}\right)^T \left(\mathbf{E_{11}}\right)^{-1} \mathbf{E_{12}} \end{pmatrix} \qquad \text{(D.38)}
\end{aligned}
$$

Applying determinants to (D.38), we have:

$$\det(\mathbf{E}) = \det(\mathbf{E_{11}})\det\left(\mathbf{E_{22}} - (\mathbf{E_{12}})^T(\mathbf{E_{11}})^{-1}\mathbf{E_{12}}\right)$$

$\square$

Using Theorem 14 and (D.2), we can now express the determinant of the symmetric matrix $\mathbf{\Sigma}_j$ as:

$$\det(\mathbf{\Sigma}_j) = \det\left((\mathbf{\Sigma}_j)_{\mathbf{11}}\right)\det\left((\mathbf{\Sigma}_j)_{\mathbf{22}} - \left((\mathbf{\Sigma}_j)_{\mathbf{12}}\right)^T\left((\mathbf{\Sigma}_j)_{\mathbf{11}}\right)^{-1}(\mathbf{\Sigma}_j)_{\mathbf{12}}\right) \tag{D.39}$$

Using (D.39), we can now write (D.37) as:

$$\begin{aligned}
f(\mathbf{V}) = &\sum_{j=1}^{k} w_j \frac{1}{(2\pi)^{q_1/2}\left(\det(\mathbf{\Sigma}_j)_{\mathbf{11}}\right)^{1/2}}\exp\left(-\frac{1}{2}\left(\mathbf{Q}_{1j}(\mathbf{v_1})\right)\right)\times \\
&\frac{1}{(2\pi)^{q_2/2}\left(\det\left((\mathbf{\Sigma}_j)_{\mathbf{22}} - \left((\mathbf{\Sigma}_j)_{\mathbf{12}}\right)^T\left((\mathbf{\Sigma}_j)_{\mathbf{11}}\right)^{-1}(\mathbf{\Sigma}_j)_{\mathbf{12}}\right)\right)^{1/2}}\exp\left(-\frac{1}{2}(\mathbf{Q}_{2j}(\mathbf{v_1},\mathbf{v_2}))\right) \\
= &\sum_{j=1}^{k} w_j \mathcal{N}_{q_1}\left(\mathbf{v_1};\boldsymbol{\mu}_{1j},(\mathbf{\Sigma}_j)_{\mathbf{11}}\right)\mathcal{N}_{q_2}\left(\mathbf{v_2};\mathbf{b}^*_{2j},\mathbf{\Sigma}^*_j\right), \qquad \text{(D.40)}
\end{aligned}$$

where

$$\begin{aligned}
\mathbf{b}^*_{2j} &= \boldsymbol{\mu}_{2j} + \left((\mathbf{\Sigma}_j)_{\mathbf{12}}\right)^T\left((\mathbf{\Sigma}_j)_{\mathbf{11}}\right)^{-1}\left(\mathbf{v_1} - \boldsymbol{\mu}_{1j}\right) \\
\mathbf{\Sigma}^*_j &= (\mathbf{\Sigma}_j)_{\mathbf{22}} - \left((\mathbf{\Sigma}_j)_{\mathbf{12}}\right)^T\left((\mathbf{\Sigma}_j)_{\mathbf{11}}\right)^{-1}(\mathbf{\Sigma}_j)_{\mathbf{12}}
\end{aligned}$$

Therefore, using (D.40) the marginal distribution of $\mathbf{v_1}$ can be derived as:

$$f_1(\mathbf{v_1}) = \int_{-\infty}^{\infty}\sum_{j=1}^{k} w_j \mathcal{N}_{q_1}\left(\mathbf{v_1};\boldsymbol{\mu}_{1j},(\mathbf{\Sigma}_j)_{\mathbf{11}}\right)\mathcal{N}_{q_2}\left(\mathbf{v_2};\mathbf{b}^*_{2j},\mathbf{\Sigma}^*_j\right)d\mathbf{v_2} \tag{D.41}$$

Because k in (D.41) is finite, we can write the marginal distribution of $\mathbf{v_1}$ as:

$$\begin{aligned}
f_1(\mathbf{v_1}) &= \sum_{j=1}^{k} w_j \mathcal{N}_{q_1}\left(\mathbf{v_1};\boldsymbol{\mu}_{1j},(\mathbf{\Sigma}_j)_{\mathbf{11}}\right)\underbrace{\int_{-\infty}^{\infty}\mathcal{N}_{q_2}\left(\mathbf{v_2};\mathbf{b}^*_{2j},\mathbf{\Sigma}^*_j\right)d\mathbf{v_2}}_{\text{Integrates to 1}} \\
&= \sum_{j=1}^{k} w_j \mathcal{N}_{q_1}\left(\mathbf{v_1};\boldsymbol{\mu}_{1j},(\mathbf{\Sigma}_j)_{\mathbf{11}}\right) \qquad \text{(D.42)}
\end{aligned}$$

Based on (D.42), it would appear that the marginal distribution of $\mathbf{v_1}$ is a k-component mixture of multivariate Gaussian distributions. However, if at least 2 of the component distributions have identical values for both $\boldsymbol{\mu}_{1j}$ and $(\boldsymbol{\Sigma}_j)_{\mathbf{11}}$, the marginal distribution of $\mathbf{v_1}$ will have less components than k. These results are summarized in the following theorem:

Theorem 15. Let the random variable $\mathbf{V} = \begin{pmatrix} \mathbf{v_1} & \mathbf{v_2} \end{pmatrix}^{\mathbf{T}}$ follow a k-component mixture of multivariate Gaussian distributions as defined in (D.1) and (D.2). The marginal distribution of $\mathbf{v_1}$ can be categorized into one of the following scenarios:

- $\mathbf{v_1} \sim \sum_{j=1}^{k} w_j \mathcal{N}_{q_1}\left(\mathbf{v_1}; \boldsymbol{\mu}_{1j}, (\boldsymbol{\Sigma}_j)_{\mathbf{11}}\right)$ as long as all the k-component distributions have distinct values for both $\boldsymbol{\mu}_{1j}$ and $(\boldsymbol{\Sigma}_j)_{\mathbf{11}}$

- Let us assume that the number of component distributions with distinct values for both $\boldsymbol{\mu}_{1j}$ and $(\boldsymbol{\Sigma}_j)_{\mathbf{11}}$, k^*, is such that $1 < k^* < k$. Further, let us also assume that the order of the distributions in (D.42) is such that $j = 1, 2, \ldots, k^*, k^* + 1, \ldots, k$. That is, the first k^* distributions all have distinct parameter values. Thus, we now have:

$$\mathbf{v_1} \sim \sum_{j=1}^{k^*+1} w_j^* \mathcal{N}_{q_1}\left(\mathbf{v_1}; \boldsymbol{\mu}_{1j}, (\boldsymbol{\Sigma}_j)_{\mathbf{11}}\right),$$

  where $w_j^* = w_j$ if $j \leq k^*$ and $w_{(k^*+1)} = \sum_{j=k^*+1}^{k} w_j$

- $\mathbf{v_1} \sim \mathcal{N}_{q_1}\left(\mathbf{v_1}; \boldsymbol{\mu}_{1j}, (\boldsymbol{\Sigma}_j)_{\mathbf{11}}\right)$ if none of the k-component distributions have distinct values for both $\boldsymbol{\mu}_{1j}$ and $(\boldsymbol{\Sigma}_j)_{\mathbf{11}}$

APPENDIX E

SOURCE CODE FOR THE R FUNCTION "RWISHART" (BASE R)

```c
/*
 *  R : A Computer Language for Statistical Data Analysis
 *  Copyright (C) 2012-2016  The R Core Team
 *
 *  This program is free software; you can redistribute it and/or modify
 *  it under the terms of the GNU General Public License as published by
 *  the Free Software Foundation; either version 2 of the License, or
 *  (at your option) any later version.
 *
 *  This program is distributed in the hope that it will be useful,
 *  but WITHOUT ANY WARRANTY; without even the implied warranty of
 *  MERCHANTABILITY or FITNESS FOR A PARTICULAR PURPOSE.  See the
 *  GNU General Public License for more details.
 *
 *  You should have received a copy of the GNU General Public License
 *  along with this program; if not, a copy is available at
 *  https://www.R-project.org/Licenses/
 */

#ifdef HAVE_CONFIG_H
# include <config.h>
#endif

#include <math.h>
#include <string.h>  // memset, memcpy
#include <R.h>
#include <Rinternals.h>
#include <Rmath.h>
#include <R_ext/Lapack.h>        /* for Lapack (dpotrf, etc.) and BLAS */

#include "stats.h" // for _()
#include "statsR.h"
```

Figure 18: Syntax for R(base) Function "rWishart"

```c
/**
 * Simulate the Cholesky factor of a standardized Wishart variate with
 * dimension p and nu degrees of freedom.
 *
 * @param nu degrees of freedom
 * @param p dimension of the Wishart distribution
 * @param upper if 0 the result is lower triangular, otherwise upper
                triangular
 * @param ans array of size p * p to hold the result
 *
 * @return ans
 */
static double
*std_rWishart_factor(double nu, int p, int upper, double ans[])
{
    int pp1 = p + 1;

    if (nu < (double) p || p <= 0)
        error(_("inconsistent degrees of freedom and dimension"));

    memset(ans, 0, p * p * sizeof(double));
    for (int j = 0; j < p; j++) { /* jth column */
        ans[j * pp1] = sqrt(rchisq(nu - (double) j));
        for (int i = 0; i < j; i++) {
            int uind = i + j * p, /* upper triangle index */
                lind = j + i * p; /* lower triangle index */
            ans[(upper ? uind : lind)] = norm_rand();
            ans[(upper ? lind : uind)] = 0;
        }
    }
    return ans;
}
```

Figure 19: Syntax for R(base) Function "rWishart" (cont.)

```
/**
 * Simulate a sample of random matrices from a Wishart distribution
 *
 * @param ns Number of samples to generate
 * @param nuP Degrees of freedom
 * @param scal Positive-definite scale matrix
 *
 * @return
 */
SEXP
rWishart(SEXP ns, SEXP nuP, SEXP scal)
{
    SEXP ans;
    int *dims = INTEGER(getAttrib(scal, R_DimSymbol)), info,
        n = asInteger(ns), psqr;
    double *scCp, *ansp, *tmp, nu = asReal(nuP), one = 1, zero = 0;

    if (!isMatrix(scal) || !isReal(scal) || dims[0] != dims[1])
        error(_("'scal' must be a square, real matrix"));
    if (n <= 0) n = 1;
    // allocate early to avoid memory leaks in Callocs below.
    PROTECT(ans = alloc3DArray(REALSXP, dims[0], dims[0], n));
    psqr = dims[0] * dims[0];
    tmp = Calloc(psqr, double);
    scCp = Calloc(psqr, double);
```

Figure 20: Syntax for R(base) Function "rWishart" (cont.)

```
Memcpy(scCp, REAL(scal), psqr);
    memset(tmp, 0, psqr * sizeof(double));
    F77_CALL(dpotrf)("U", &(dims[0]), scCp, &(dims[0]), &info);
    if (info)
        error(_("'scal' matrix is not positive-definite"));
    ansp = REAL(ans);
    GetRNGstate();
    for (int j = 0; j < n; j++) {
        double *ansj = ansp + j * psqr;
        std_rWishart_factor(nu, dims[0], 1, tmp);
        F77_CALL(dtrmm)("R", "U", "N", "N", dims, dims,
                        &one, scCp, dims, tmp, dims);
        F77_CALL(dsyrk)("U", "T", &(dims[1]), &(dims[1]),
                        &one, tmp, &(dims[1]),
                        &zero, ansj, &(dims[1]));

        for (int i = 1; i < dims[0]; i++)
            for (int k = 0; k < i; k++)
                ansj[i + k * dims[0]] = ansj[k + i * dims[0]];
    }

    PutRNGstate();
    Free(scCp); Free(tmp);
    UNPROTECT(1);
    return ans;
}
```

Figure 21: Syntax for R(base) Function "rWishart" (cont.)

BIBLIOGRAPHY

[1] Krzanowski, W.J. (1993). The location model for mixtures of categorical and continuous variables. *Journal of Classification* **10**(1): 25-49.

[2] Tate, R.F. (1954). Correlation between a discrete and a continuous variable. *Annals of Mathematical Statistics* **25**: 603-607.

[3] Tate, R.F. (1955). Applications of correlation models for biserial data. *Journal of the American Statistical Association* **50**: 1078-1095.

[4] Olkin, I. and Tate, R.F. (1961). Multivariate correlation models with mixed discrete and continuous variables. *Annals of Mathematical Statistics* **32**(2): 448-465.

[5] Little, R.J. and Rubin, D.B. (2002). *Statistical Analysis with Missing Data*. 2nd edition. Hoboken, John Wiley and Sons.

[6] Little, R.J. and Schlucher, M.D. (1985). Maximum likelihood estimation for mixed continuous and categorical data with missing values. *Biometrika***72**(3): 497-512.

[7] Schafer, J.L. (1997). *Analysis of Incomplete Multivariate Data*. London, Chapman and Hall/CRC Press.

[8] Afifii, A.A. and Elashoff, R.M. (1969). Multivariate two-sample tests with dichotomous and continuous variables. I. The location model. *Annals of Mathematical Statistics* **40**(1): 290-298.

[9] Krzanowski, W.J. (1975). Discrimination and classification using both binary and continuous variables. *Journal of the American Statistical Association* **70**(352): 782-790.

[10] Krzanowski, W.J. (1976). Canonical representation of the location model for discrimination or classification. *Journal of the American Statistical Association* **71**(356): 845-848.

[11] Krzanowski, W.J. (1980). Mixtures of continuous and categorical variables in discriminant analysis. *Biometrics* **36**(3): 493-499.

[12] Krzanowski, W.J. (1982). Mixtures of continuous and categorical variables in discriminant analysis: A hypothesis testing approach. *Biometrics* **38**(4): 991-1002.

[13] Krzanowski, W.J. (1983). Distance between populations using mixed continuous and categorical variables. *Biometrika* **70**(1): 235-243.

[14] Chang, P.C. and Afifi, A.A. (1974). Classification based on dichotomous and continuous variables. *Journal of the American Statistical Association* **69**(346): 336-339.

[15] de Leon, A.R., Soo, A., and Williamson, T. (2011). Classification with discrete and continuous variables via general mixed-data models. *Journal of Applied Statistics* **38**(5): 1021-1032.

[16] Bar-Hen, A. and Daudin, J.-J. (1995). Generalization of the Mahalanobis distance in the mixed case. *Journal of Multivariate Analysis* **53**: 333-342.

[17] Morales, D., Pardo, L., and Zografos, K. (1998). Informational distances and related statistics in mixed continous and categorical variables. *Journal of Statistical Planning and Inference* **75**: 47-63.

[18] Lauritzen, S.L. and Wermuth, N. (1989). Graphical models for association between variables, some of which are qualitative and some quantitative. *Annals of Statistics* **17**(1): 31-54.

[19] de Leon, A.R. and Carriére, K.C. (2000). On the one-sample location hypothesis for mixed bivariate data. *Communications in Statistics-Theory and Methods* **29**(11): 2573-2581.

[20] de Leon, A.R. (2007). One-sample likelihood ratio tests for mixed data. *Communications in Statistics-Theory and Methods* **36**(1): 129-141.

[21] Cox, D.R. (1972). The analysis of multivariate binary data. *Journal of the Royal Statistical Society. Series C (Applied Statistics)* **21**(2): 113-120.

[22] Cox, D.R. and Wermuth N. (1992). Response models for mixed binary and quantitative variables. *Biometrika* **79**(3): 441-461.

[23] Nelsen, R.B. (2006). *An Introduction to Copulas*. 2nd Edition. New York, Springer.

[24] Casella, G. and Berger, R.L. (2002). *Statistical Inference*. 2nd Edition. Pacific Grove, Duxbury.

[25] Sklar, A. (1959). Fonctions de répartition à n-dimensions et leurs marges. *Publications de l'Institut de statistique de l'Université de Paris* **8**: 229-231.

[26] Genest, C. and Nešlehovà, J. (2007). A primer on copulas for count data. *ASTIN Bulletin* **37**(2): 475-515.

[27] Song, P.X.-K., Li, M., and Yuan, Y (2009). Joint regression analysis of correlated data using gaussian copulas. *Biometrics* **65**: 60-68.

[28] de Leon, A.R. and Wu, B. (2011). Copula-based regression models for a bivariate mixed discrete and continuous outcome. *Statistics in Medicine* **30**: 175-185.

[29] Spissu, E., Pinjari, A.R., Pendyala, R.M., and Bhat, C.R. (2009). A copula-based joint multinomial discrete-continuous model of vehicle type choice and miles of travel. *Transportation.* **36**: 403-422.

[30] Wu, B. and de Leon, A.R. (2014). Gaussian copula mixed models for clustered mixed outcomes, with application in developmental toxicology. *Journal of Agricultural, Biological, and Environmental Statistics* **19**(1): 39-56.

[31] de Leon, A.R. and Chough, K.C. (2013). *Analysis of Mixed Data: Methods and Applications.* Boca Raton, CRC Press.

[32] Cox, N.R. (1974). Estimation of the correlation between a continuous and a discrete variable. *Biometrics* **30**: 171-178.

[33] Bedrick, E.J., Lapidus, J., and Powell, J.F. (2000). Estimating the Mahalanobis distance from mixed continuous and discrete data. *Biometrics* **56**: 394-401.

[34] de leon, A.R. and Carrière, K.C. (2005). A generalized Mahalanobis distance for mixed data. *Journal of Multivariate Analysis* **92**(1): 174-185.

[35] Satterthwaite, F.E. (1946). An approximation of estimates of variance components. *Biometrics Bulletin* **2**(6): 110-114.

[36] Satterthwaite, F.E. (1941). Synthesis of variance. *Psychometrika* **6**: 309-316.

[37] Welch, B.L. (1947). The generalization of 'student's' problem when several different population variances are involved. *Biometrika* **34**(1/2): 28-35.

[38] Davis, A.W. (1977). A differential equation approach to linear combinations of independent chi-squares. *Journal of the American Statistical Association* **72**(357): 212-218.

[39] Solomon, H. and Stephens, M.A. (1977). Distribution of a sum of weighted chi-square variables. *Journal of the American Statistical Association* **72**(360): 881-885.

[40] Oman, S.D. and Zacks, S. (1981). A mixture approximation to the distribution of a weighted sum of chi-squared variables. *Journal of Statistical Computation and Simulation* **13**(3-4): 215-224.

[41] Mathai, A.M. (1982). Storage capacity of a dam with gamma type inputs. *Annals of the Institute of Statistical Mathematics, Part A* **34**: 591-597.

[42] Moschopoulos, P.G. and Canada, W.B. (1984). The distribution function of a linear combination of chi-squares. *Computers & Mathematics with Applications* **10**(4/5): 383-386.

[43] Lindsay, B.G. et al. (2000). Moment-based approximations of a distributions using mixtures: theory and applications. *Annals of the Institute of Statistical Mathematics* **52**(2): 215-230.

[44] Di Salvo, F. (2006). Exact distribution of the weighted convolution of two gamma distributions. *Atti della XLIII Riumione Scientifica SIS (Acts of the 43 Annual Meeting of the Italian Scientific Society)* **Università di Torino: Società Italiana di Scienctifica, 2006**: 511-514.

[45] Anderson, T.W. (2003). *An introduction to multivariate statistical analysis.* 3rd edition. Hoboken, John Wiley and Sons.

[46] Gupta, A.K. and Nagar, D.K. (1999). *Matrix Variate Distributions.* Boca Raton, Chapman & Hall/CRC.

[47] Deemer, W.L. and Olkin, I. (1951). The jacobians of certain matrix transformations useful in multivariate analysis: based on lectures of P.L. Hsu at the University of North Carolina, 1947. *Biometrika* **38**(3/4): 345-367.

[48] Searle, S.R. (1982). *Matrix Algebra Useful for Statistics.* New York, John Wiley & Sons.

[49] Tan, W.Y. and Gupta, R.P. (1983). On approximating a linear combination of central Wishart matrices with positive coefficients. *Communications in Statistics - Theory & Methods* **12**(22): 2589-2600.

[50] Gindikin, S.G. (1975). Invariant generaized functions in homogeneous domans. *Funktsional'nyi Analiz i Ego Prilozheniya.* **9**(1): 56-58. Translated from Russian.

[51] Graczyk, P., Letac, G., and Massam, H. (2003). The complex Wishart distribution and the symmetric group. *The Annals of Statistics* **31**(1): 287-309.

[52] Xiao, Y., Ku, Y.-C., Bloomfield, P., and Ghosh, S.K. (2015). On the degrees of freedom in MCMC-based Wishart models for time series data. *Statistics and Probability Letters* **98**(2015): 59-64.

[53] R Core Team (2017). R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria.

[54] Wijsman, R.A. (1957). Random orthogonal transformatins and their use in some classical distribution problems in multivariate analysis. *Annals of Mathematical Statistics* **28**(2): 415-423.

[55] Kshirsagar, A.M. (1959). Barlett decomposition and Wishart distribution. *Annals of Mathematical Statistics* **30**(1): 239-241.

[56] CodeForge (2015). Simulation of MATLAB source code for WISHRND - random matrices frm Wishart distribution. Accessed 7 May 2018. http://www.codeforge.com/read/286358/wishrnd.m_html.

[57] Mardia, K.V., Kent, J.T., and Bibby J.M. (1979). *Multivariate analysis*. London, Academic Press.

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

[59] Kubokawa, T. and Srivastava, M.S. (2008). Estimation of the precision matrix of a singular Wishart distribution and its application in high-dimensional data. *Journal of Multivariate Analysis* **99**(2008): 1906-1928.

[60] Doetsch, G. (1961). *Guide to the application of Laplace transforms*. London, D. Van Nostrand Company, Ltd.

[61] Spiegel, M.R., Lipschutz, S., Schiller, J.J., and Spellman, D. (2009). *Complex variables: with an introduction to conformal mapping and its applications*. Second edition. New York, McGraw Hill.

[62] Arfken, G. (1985). *Mathematical Methods for Physicists*. Third edition. Orlando, Academic Press.

[63] Bartle, R.G. and Sherbert, D.R. (1992). Introduction to Real Analysis. Second edition. New York, John Wiley and Sons, Inc.