2016-01-01

A New Test for the Mean Vector in High Dimensional Setting

Behzad Aalipur Hafshejani
University of Texas at El Paso, baalipurhafshejani@miners.utep.edu

A NEW TEST FOR THE MEAN VECTOR IN HIGH DIMENSIONAL SETTING

Behzad Aalipur Hafshejani

Master's Program in Mathematical Sciences

APPROVED:

Panagis Moschopoulos, Ph.D., Chair

Naijun Sha, Ph.D.

Vladik Kreinovich, Ph.D.

Charles Ambler, Ph.D.
Dean of the Graduate School

to my

MOTHER and FATHER

with love

A NEW TEST FOR THE MEAN VECTOR IN HIGH DIMENSIONAL SETTING

by

Behzad Aalipur Hafshejani

THESIS

Presented to the Faculty of the Graduate School of

The University of Texas at El Paso

in Partial Fulfillment

of the Requirements

for the Degree of

MASTER OF SCIENCE

Master's Program in Mathematical Sciences

THE UNIVERSITY OF TEXAS AT EL PASO

May 2016

Acknowledgements

First and foremost, I would like to express my sincere gratitude to my advisor Prof. Panagis Moschopoulos. Words cannot express my gratitude and respect toward him. His patience, knowledge, philosophical point of view, and in one word his personality have influenced me in different ways. Working with a supervisor who considers both personal and professional growth of his students was a big fortune for me. He taught me having passions is one side of the story, but how we can have meaningful approaches toward them is another. He taught me that enough is never enough and there is no end to perfecting the quality of work we deliver. He taught me how to be a good researcher, and a better person. His lifetime lessons will stay with me forever.

I would also like to thank my thesis committee members, Prof. Naijun Sha and Prof. Vladik Kreinovich, for their time, insightful comments, and their support. I would like to thank Dr. Ori Rosen who believed in my abilities and gave me the opportunity to work as his research assistant. I would like to greatly acknowledge and thank the Mathematical Science department for creating a friendly environment to conduct research.

I would also like to thank my friends for all the great times we had together during my stay in this university, all laughs and cries, and ups and downs.

The last and not the least is my family. I would like to thank my family for all their love and encouragements from thousands of miles away. Their pure love and their memories have enabled me to go on. I have been extremely fortunate to have a thoughtful father who kindled the very first sparkles of curiosity in me, a person who taught me discovering truth needs dedication and nothing comes for free. I thank my mother for all her heartfelt dedications to my life. I thank my beloved brother who has always been supporting me as my closest friend. I thank my sister, an angel who is always making me feel blessed in my life.

NOTE: This thesis was submitted to my Supervising Committee on May 6, 2016.

Abstract

Traditional statistical data analysis mostly includes methods and techniques to deal with problems in which there are many observations but a few variables. Nonetheless, the current inclination is toward more observations but also, toward more variables. Today's observations gathered on individuals are images, curves, or even movies. Unfortunately many traditional methods do not work well in high dimensional settings. As an example Hotelling's T^2 test which is well known and widely used in the literature does not work when it comes to high dimensional problems. Consequently statisticians are making an effort to find remedies or new approaches to multivariate mean testing.

The issue with Hotelling's T^2 in high dimensions is that the sample covariance matrix is no longer invertible and hence the test statistic is unattainable. Alternative methods of estimating the covariance matrix have been fruitful. However, some approaches took another path. These approaches try to avoid estimating the covariance matrix.

Theses approaches mainly take advantage of density approximations. Approximation of the statistics is considered and the asymptotic behavior is studied.

We consider an approximation which is known to perform well in symmetrizing distributions, and we develop a new test statistic which has better power in simulations. Our results are compared with some well known tests in the literature.

Contents

Chapter 1

Preliminaries and Various Representations of Quadratic Forms

1.1 Matrix Algebra

Dealing with quadratic statistical models requires some mathematical techniques of matrix algebra. In this chapter we will have some preliminaries about matrix algebra. Then we move on to quadratic forms in random variables. Since we deal mostly with quadratic forms in normal random variables we will focus on them and investigate the moments of the quadratic form. Throughout this chapter we will have some theorems and examples. We will also prove some theorems if the proofs provide deeper insight into quadratic forms.

Definition 1.1.1. *A square matrix $\boldsymbol{A}$ is symmetric if $\boldsymbol{A}' = \boldsymbol{A}$, where $\boldsymbol{A}'$ is the transpose of the matrix $\boldsymbol{A}$.*

Definition 1.1.2. *A square matrix $\boldsymbol{A}$ is skew-symmetric if $\boldsymbol{A}' = -\boldsymbol{A}$.*

Definition 1.1.3. *The n-vectors $\boldsymbol{x}_1, ..., \boldsymbol{x}_n$ are called a collection of orthogonal vectors when $\boldsymbol{x}_i'\boldsymbol{x}_j = 0, i \neq j,\ i, j = 1, ..., n$. Moreover, they are called orthonormal if $\boldsymbol{x}_i'\boldsymbol{x}_i = 1, i = 1, ..., n$.*

Definition 1.1.4. *The trace of a square matrix $\boldsymbol{A}$ is defined as the sum of its diagonal elements and is denoted by $tr\boldsymbol{A}$.*

The trace of a matrix has the following easily shown properties:

(i) $tr(\boldsymbol{A} + \boldsymbol{B}) = tr(\boldsymbol{A}) + tr(\boldsymbol{B})$

(ii) $tr(\boldsymbol{AB}) = tr(\boldsymbol{BA})$

(iii) $tr(\boldsymbol{ABC}) = tr(\boldsymbol{BCA}) = tr(\boldsymbol{CAB})$

(iv) $tr(\boldsymbol{A}) = tr(\boldsymbol{A}')$

Definition 1.1.5. *The maximum number of linearly independent rows and columns of an $m \times n$ matrix $\boldsymbol{A}$ is called the rank of the matrix and is usually denoted by $\rho(\boldsymbol{A})$. A matrix $\boldsymbol{A}$ is called full rank when $\rho(\boldsymbol{A}) = min(m, n)$.*

Definition 1.1.6. *A squared matrix $\boldsymbol{A}$ is invertible (nonsingular) if there exists a square matrix $\boldsymbol{B}$ such that $\boldsymbol{AB} = \boldsymbol{I} = \boldsymbol{BA}$. The inverse matrix of $\boldsymbol{A}$ is denoted by $\boldsymbol{A}^{-1}$. If a square matrix is full rank, then it is invertible.*

Definition 1.1.7. *The determinant of an $n \times n$ matrix is a function f such that*

(i) $f(r_1, ..., cr_i, ..., r_n) = cf(r_1, ..., r_i, ..., r_n)$

(ii) $f(r_1, ..., r_i + r_j, ..., r_n) = f(r_1, ..., r_i, ..., r_n) + f(r_1, ..., r_j, ..., r_n)$

(iii) $f(e_1, ..., e_n) = 1$

where r_i denotes the ith row(column) of the matrix $\boldsymbol{A}$, and $\boldsymbol{e}_i$ is unit vector with the ith element equal to 1 and the rest zero.

Definition 1.1.8. *The leading $m \times m$ minor or principal minor of an $n \times n$ matrix $\boldsymbol{A}$, is determinant of the submatrix*

$$\begin{pmatrix} a_{11} & \dots & a_{1m} \\ \vdots & & \vdots \\ a_{m1} & \dots & a_{mm} \end{pmatrix} \quad , \quad m = 1, 2, ..., n. \tag{1.1}$$

Definition 1.1.9. *A matrix $\boldsymbol{A}$ is said to be idempotent if $\boldsymbol{AA} = \boldsymbol{A}$.*

An idempotent matrix $\boldsymbol{A}$ has some properties as follows:

(i) If $\boldsymbol{A}$ is invertible, then $\boldsymbol{A} = \boldsymbol{I}$

(ii) if $\rho(\boldsymbol{A}) = r$, then tr$\boldsymbol{A} = r$

Definition 1.1.10. *Let $\boldsymbol{P}$ be a $n \times n$ square matrix. $\boldsymbol{P}$ is called an orthogonal matrix if its column vectors are orthonormal vectors. In other words its columns are perpendicular to each other and the inner product of two columns is zero and inner product of one to itself is one. This definition can be summarized as follows*

1. The column vectors of $\boldsymbol{P}$ are orthonormal.

2. $\boldsymbol{P'P} = \boldsymbol{PP'} = \boldsymbol{I_n}$.

3. $\boldsymbol{P}' = \boldsymbol{P}^{-1}$

4. The row vectors of $\boldsymbol{P}$ are orthonormal.

Definition 1.1.11. *For an $n \times n$ matrix $\boldsymbol{A}$, a vector $\boldsymbol{x}$ in $\mathcal{R}^n$ is called characteristic vector or eigenvector iff*

$$\boldsymbol{Ax} = \lambda \boldsymbol{x} \quad or \quad (\boldsymbol{A} - \lambda \boldsymbol{I})\boldsymbol{x} = 0 \tag{1.2}$$

where λ is a scalar and is a root of the following equation

$$|\boldsymbol{A} - \lambda \boldsymbol{I}| = 0 \tag{1.3}$$

This equation is called characteristic equation of the matrix $\boldsymbol{A}$.

Theorem 1.1.1. *Consider $\boldsymbol{A}$ to be a $n \times n$ matrix with eigenvalues $\lambda_1, ..., \lambda_n$ then we have*

(i) $tr(\boldsymbol{A}^k) = \sum_{i=1}^{n} \lambda_i^k \quad , \quad k = 1, 2, ...$

(ii) $|\boldsymbol{A}| = \prod_{i=1}^{n} \lambda_i$

(iii) $|\boldsymbol{I}_n \pm \boldsymbol{A}| = \prod_{i=1}^{n} (1 \pm \lambda_i)$

Theorem 1.1.2. *All of the eigenvalues of a real $n \times n$ symmetric matrix are real.*

Theorem 1.1.3. *If $\boldsymbol{A}$ is idempotent matrix, then its eigenvalues are either zero or one. If all are one then $\boldsymbol{A} = \boldsymbol{I}_n$*

Now we have enough suitable matrix knowledge to start the discussion about the quadratic forms. Some properties such as moments of quadratic forms and different representations of quadratic forms are discussed in this chapter. To illustrate the properties and theorems about quadratic forms, some examples are also presented.

Definition 1.1.12. *A quadratic form is a polynomial of degree 2 in n variables $\boldsymbol{x}_1, \boldsymbol{x}_2, ..., \boldsymbol{x}_n$ of the form*

$$Q(\boldsymbol{x}) = \boldsymbol{x}'\boldsymbol{A}\boldsymbol{x} = \sum_{i=1}^{n}\sum_{j=1}^{n} a_{ij}\boldsymbol{x}_i\boldsymbol{x}_j$$

Where $\boldsymbol{A}$ is a real matrix and $\boldsymbol{x}$ is a real vector.

One can see easily that if $\boldsymbol{A}$ is a skew-symmetric matrix, a quadratic form associated to it would result in zero since for a skew-symmetric matrix $\boldsymbol{A}$, $Q(\boldsymbol{x}) = \boldsymbol{x}'\boldsymbol{A}\boldsymbol{x} = \boldsymbol{x}'\boldsymbol{A}'\boldsymbol{x} = -\boldsymbol{x}'\boldsymbol{A}\boldsymbol{x} = -Q(\boldsymbol{x})$. Hence, without any loss of generality the matrix associated with a quadratic form might be assumed to be symmetric. If it is not a symmetric matrix one can equivalently take the symmetric matrix $\frac{\boldsymbol{A}+\boldsymbol{A}'}{2}$ as the matrix associated with the quadratic form. The following relation shows how the matrix of a quadratic form can be assumed to be a symmetric matrix.

$$\begin{aligned} Q(\boldsymbol{x}) &= \boldsymbol{x}'\boldsymbol{A}\boldsymbol{x} = \boldsymbol{x}'(\frac{\boldsymbol{A}+\boldsymbol{A}'}{2} + \frac{\boldsymbol{A}-\boldsymbol{A}'}{2})\boldsymbol{x} \\ &= \boldsymbol{x}'(\frac{\boldsymbol{A}+\boldsymbol{A}'}{2})\boldsymbol{x} + \boldsymbol{x}'(\frac{\boldsymbol{A}-\boldsymbol{A}'}{2})\boldsymbol{x} \\ &= \boldsymbol{x}'\frac{\boldsymbol{A}+\boldsymbol{A}'}{2}\boldsymbol{x} \qquad (1.4) \end{aligned}$$

Definition 1.1.13. *A real quadratic form is said to be positive definite if $Q(\boldsymbol{x}) = \boldsymbol{x}'\boldsymbol{A}\boldsymbol{x} > 0$ for every $\boldsymbol{x} \neq \boldsymbol{0}$. A matrix $\boldsymbol{A}$ is similarly called positive definite and denoted by $\boldsymbol{A} > 0$ if the quadratic form $Q(\boldsymbol{x}) = \boldsymbol{x}'\boldsymbol{A}\boldsymbol{x}$ is positive definite. The followings are some properties of a positive definite matrix.*

(i) Its eigenvalues are positive

(ii) It can be written as a multiplication $\boldsymbol{P}'\boldsymbol{P}$ where $\boldsymbol{P}$ is a nonsingular matrix.

(iii) Its inverse is also positive definite.

(iv) It is nonsingular.

(v) Its square root exist and is denoted by $\boldsymbol{A}^{\frac{1}{2}}$.

Theorem 1.1.4. *The definiteness of a quadratic form in invariant under nonsingular linear transformation. In other words if $Q(\boldsymbol{x}) = \boldsymbol{x}'\boldsymbol{A}\boldsymbol{x}$ is a positive definite quadratic form and $\boldsymbol{B}$ is a nonsingular matrix, then $Q(\boldsymbol{y}) = \boldsymbol{y}'\boldsymbol{A}\boldsymbol{y}$ is also a positive definite quadratic form where $\boldsymbol{y} = \boldsymbol{B}\boldsymbol{x}$.*

The spectral decomposition is a decomposition where a matrix is represented in terms of its eigenvalues and eigenvectors. By the use of the spectral decomposition one can see any real symmetric can be written as $\boldsymbol{Q}'\boldsymbol{D}\boldsymbol{Q}$, where $\boldsymbol{Q}$ is an orthogonal matrix and D is a diagonal matrix. In other words it can be shown that a symmetric matrix is diagonalizable. Another decomposition which is frequently used is the following.

Definition 1.1.14. *A Cholesky decomposition of a symmetric $n \times n$ positive definite matrix $\boldsymbol{A}$ is $\boldsymbol{A} = \boldsymbol{T}\boldsymbol{T}'$, where $\boldsymbol{T}$ is a lower triangular matrix with positive elements on its diagonal.*

This decomposition is unique and one can easily see that the elements of $\boldsymbol{T}$ would be obtained from the elements of $\boldsymbol{A}$ by equating elements of $\boldsymbol{T}\boldsymbol{T}'$ to $\boldsymbol{A}$.

1.2 Statistical Preliminaries

We now state different representation of quadratic forms in random variables. Then we will derive the moments of the quadratic forms which are used frequently in the approximation of the distributions of quadratic forms. Before that we are going to have a brief overview of definitions and concepts of multivariate distribution.

Definition 1.2.1. *Let's consider n random variables $X_1, X_2, ..., X_n$ and outcomes are $x_1, x_2, ..., x_n$. The cumulative distribution then would be as follows*

$$P(X_1 \leq x_1, X_2 \leq x_2, ..., X_n \leq x_n) = F(x_1, x_2, ..., x_n) \tag{1.5}$$

Assuming that all partial derivatives exist, the density function is obtained from the cumulative distribution function:

$$f(x_1, x_2, ..., x_n) = \frac{\partial^n}{\partial x_1 \partial x_2 ... \partial x_n} F(x_1, x_2, ..., x_n) \tag{1.6}$$

Once we have a density function we can define marginal density function of a subset of variables by integrating the density function with respect to the remaining variables.

Definition 1.2.2. *The "last $n - k$ x's" or marginal density function is $f(x_1, x_2, ..., x_n)$ after integrating out the first k x's. Then the marginal of $x_{k+1}, ..., x_n$ is*

$$g(x_{k+1}, ..., x_n) = \int_{-\infty}^{\infty} ... \int_{-\infty}^{\infty} f(x_1, x_2, ...x_n) dx_1 dx_2 ... dx_k \tag{1.7}$$

Definition 1.2.3. *The conditional distribution for the "last $n - k$ x's" is the ratio of $f(x_1, x_2, ...x_n)$ to the marginal for the "last $n - k$ x's". It is as follows:*

$$f(x_1, ..., x_k | x_{k+1}, ..., x_n) = \frac{f(x_1, x_2, ..., x_n)}{g(x_{k+1}, ..., x_n)} \tag{1.8}$$

It is noteworthy that it is not necessary to have any order for x_is.

Definition 1.2.4. *The ith moment about zero of the pth variable is*

$$\boldsymbol{\mu}_{x_p}^{(i)} = E(x_p^i) = \int_{-\infty}^{\infty} x_p^i g(x_p) dx_p$$

which is computable through

$$\boldsymbol{\mu}_{x_p}^{(i)} = \int_{-\infty}^{\infty} ... \int_{-\infty}^{\infty} x_p^i f(x_1, x_2, ..., x_n) dx_1 dx_2 ... dx_n \tag{1.9}$$

Definition 1.2.5. *The covariance between two variables can be defined as*

$$\sigma_{k,l} = E(x_k - \mu_k)(x_l - \mu_l) = \int_{-\infty}^{\infty}\int_{-\infty}^{\infty}(x_k - \mu_k)(x_l - \mu_l)g(x_k, x_l)dx_k dx_l$$

where g is marginal density of the variables x_k and x_l.

Calculating and carrying covariances and variances into a matrix yields the variance-covariance matrix (σ_{ij}), $i, j = 1, ..., n$. Subsequently, the correlation matrix is then obtained as follows:

$$R = \{\frac{\sigma_{i,j}}{\sigma_i\sigma_j}\} = \boldsymbol{D}\{\frac{1}{\sigma_i}\}\boldsymbol{V}\boldsymbol{D}\{\frac{1}{\sigma_i}\} \quad \text{for} \quad i, j = 1, ..., n$$

were $\boldsymbol{D}$ is a diagonal matrix with diagonal elements $1/\sigma_i$, for $i = 1, 2, ..., n$.
When the variables $\boldsymbol{x}$ are transformed to variables $\boldsymbol{y}$ by a linear transformation $\boldsymbol{Y} = \boldsymbol{T}\boldsymbol{X}$, the mean and covariance matrix of Y are easily derived by

$$\boldsymbol{\mu_Y} = \boldsymbol{T}\boldsymbol{\mu_X} \quad \text{and} \quad \mathrm{Var}(\boldsymbol{Y}) = \boldsymbol{T}\boldsymbol{V}\boldsymbol{T}'. \tag{1.10}$$

Using a nonsingular transformation $\boldsymbol{T}$, an integral involving the differentials $dx_1, dx_2, .., dx_n$ will be substituted by the Jacobian of x's with respect to y's multiplied by $dy_1, dy_2, ..., dy_n$. For example

$$\int_{-\infty}^{\infty}...\int_{-\infty}^{\infty} f(x_1, x_2, ..., x_n)dx_1...dx_n = 1$$

and having transformation $\boldsymbol{y} = \boldsymbol{T}\boldsymbol{x}$ and hence $\boldsymbol{x} = \boldsymbol{T}^{-1}\boldsymbol{y}$ yields

$$\int_{-\infty}^{\infty}...\int_{-\infty}^{\infty} f(\boldsymbol{T}^{-1}\boldsymbol{y})(1/\|\boldsymbol{T}\|)dy_1...dy_n = 1$$

An alternative way to express a random variable X is characteristic function. It is defined by expectation of e^{itx}, in which i is imaginary unit and t is a real number. Since characteristic function always exists and it uniquely represent the random variable it is widely used. If a random variable has a density function, its characteristic function is

$$\varphi_{\boldsymbol{x}}(t) = E[e^{itx}] = \int_R e^{itx} f_X(\boldsymbol{x})dx. \tag{1.11}$$

A special case of characteristic function is moment generating function. Which is $\varphi_X(-it) = E[e^{i(-it)\boldsymbol{x}}] = E[e^t x] = M_{\boldsymbol{x}}(t)$. Analogous to characteristic function, the moment generating function determines the probability distribution of a random variable uniquely, but unlike characteristic function it does not always exist.

Definition 1.2.6. *for a random variable X with a univariate distribution the moment generating function(m.g.f) is defined by*

$$M_x(t) = E(e^{tx}) = \int_{-\infty}^{\infty} e^{tx} f(x)dx \tag{1.12}$$

Similarly, for a function of $\boldsymbol{x}$ as $h(\boldsymbol{x})$ one can have

$$M_{h(\boldsymbol{x})}(t) = E(e^{th(\boldsymbol{x})}) = \int_{-\infty}^{\infty} e^{th(\boldsymbol{x})} f(\boldsymbol{x})dx \tag{1.13}$$

It is noteworthy that this function is called moment generating function because it generates the moments.

$$M_n = E\left(X^n\right) = M_X^{(n)}(0) = \frac{d^n M_X}{dt^n}(0). \tag{1.14}$$

Here n must be a non-negative integer. We can extend the same idea to multivariate distributions. For $\boldsymbol{X} = (X_1, \ldots, X_n)$ which is an n-dimensional random vector, and $\boldsymbol{t}\cdot\boldsymbol{X} = \boldsymbol{t}'\boldsymbol{X}$ we have

$$M_{\boldsymbol{X}}(\boldsymbol{t}) = \mathbb{E}\left(e^{\boldsymbol{t}'\boldsymbol{X}}\right). \tag{1.15}$$

Imitating the same idea for the moment generating function of the function h of $\boldsymbol{x}$ and considering $\boldsymbol{x}'\boldsymbol{A}\boldsymbol{x}$ as a function of $\boldsymbol{x}$ we have.

$$M_{\boldsymbol{x}'\boldsymbol{A}\boldsymbol{x}}(t) = E(e^{t\boldsymbol{x}'\boldsymbol{A}\boldsymbol{x}}) = \int_{-\infty}^{\infty} ... \int_{-\infty}^{\infty} e^{t\boldsymbol{x}'\boldsymbol{A}\boldsymbol{x}} f(x_1, x_2, ..., x_n)dx_1...dx_n \tag{1.16}$$

An important use of m.g.f is due to the fact that it is unique and one can characterize the density function given m.g.f. A nice property of m.g.f is that if a joint m.g.f is of two random variable is can be factorized to two separate m.g.f's they are independent.

$$M_{(\boldsymbol{x}_1,\boldsymbol{x}_2)}(t_1, t_2) = M_{\boldsymbol{x}_1}(t_1)M_{\boldsymbol{x}_2}(t_2)$$

Definition 1.2.7. *A multivariate normal density is defined as follows:*

$$f_{\boldsymbol{x}}(x_1, \dots, x_k) = \frac{1}{\sqrt{(2\pi)^k |\boldsymbol{V}|}} \exp\left(-\frac{1}{2}(\boldsymbol{x}-\boldsymbol{\mu})'\boldsymbol{V}^{-1}(\boldsymbol{x}-\boldsymbol{\mu})\right) \tag{1.17}$$

where $\boldsymbol{\mu}$ and $\boldsymbol{V}$ denotes the mean and the matrix of variance-covariance for the random vector $\boldsymbol{x}$.

the moment generating function for a multivariate normal random variable $\mathbf{X} \sim N_p(\boldsymbol{\mu}, \boldsymbol{\Sigma})$ and $\mathbf{t}' = (t_1, t_2, \dots, t_p)$ is shown to be

$$M_{\mathbf{X}}(\mathbf{t}) = e^{\mathbf{t}'\boldsymbol{\mu} + \frac{1}{2}\mathbf{t}'\boldsymbol{V}\mathbf{t}} . \tag{1.18}$$

The moment generating function for a subset of random vector $\boldsymbol{x}$ can be readily obtained. Considering $\boldsymbol{x}' = [\boldsymbol{x}_1', \boldsymbol{x}_2']$ in which $\boldsymbol{x}_1$ and $\boldsymbol{x}_2$ are two subset of random vector $\boldsymbol{x}$, we have $t = [t_1', t_2']$ as corresponding ts to those subsets and $\boldsymbol{V} = \begin{bmatrix} V_{11} & V_{12} \\ V_{21} & V_{22} \end{bmatrix}$ is block variance covariance matrix. Then we have

$$M_{\boldsymbol{x}_{1,},\dots,\boldsymbol{x}_k}(t) = \int_{-\infty}^{\infty} \dots \int_{-\infty}^{\infty} e^{t_1 x_1 + \dots + t_k x_k} f(x1, \dots \boldsymbol{x}_n) dx_1 \dots dx_n = e^{t_1'\boldsymbol{\mu}_1 + \frac{1}{2}t_1' V_{11} t_1} \tag{1.19}$$

For the marginal distributions of a multivariate normal random variable we imitate the univariate one.

$$g(\boldsymbol{x}_1, \dots, \boldsymbol{x}_k) = \int_{-\infty}^{\infty} \dots \int_{-\infty}^{\infty} f(\boldsymbol{x}_1, \boldsymbol{x}_2, \dots, \boldsymbol{x}_n) dx_{k+1} \dots dx_n$$

So we write $\boldsymbol{x}' = [\boldsymbol{x}_1', \boldsymbol{x}_2']$ and similarly $\boldsymbol{\mu}' = [\boldsymbol{\mu}_1', \boldsymbol{\mu}_2']$. Then $t' = [t_1', t_2']$ would be corresponding ts and $\boldsymbol{V} = \begin{bmatrix} V_{11} & V_{12} \\ V_{21} & V_{22} \end{bmatrix}$ is splitted variance covariance matrix. We know

$$M_{\boldsymbol{x}_1, \dots \boldsymbol{x}_l}(t_1) = \int_{-\infty}^{\infty} \dots \int_{-\infty}^{\infty} g(\boldsymbol{x}_1, \dots, \boldsymbol{x}_k) dx_1 \dots dx_k$$

And the result is $e^{t_1'\boldsymbol{\mu}_1 + \frac{1}{2}t_1' V_{11} t_1}$ which is a moment generating function for a multivariate normal distribution with $\boldsymbol{\mu} = \boldsymbol{\mu}_1$ and matrix of variance covariance as V_{11}. So it means

that

$$g(\boldsymbol{x_1}) = g(\boldsymbol{x}_1, ..., \boldsymbol{x}_k) = \frac{\exp[-\frac{1}{2}(\boldsymbol{x}_1 - \boldsymbol{\mu}_1)' V_{11}^{-1}(\boldsymbol{x}_1 - \boldsymbol{\mu}_1)]}{(2\pi)^{\frac{1}{2}k} \mid V_{11} \mid^{\frac{1}{2}}}$$

Definition 1.2.8. *Consider*

$$Q(\boldsymbol{X}) = Q(X_1, ..., X_n) = \boldsymbol{X'AX} = \boldsymbol{\Sigma}_{i=1}^{n}\boldsymbol{\Sigma}_{i=1}^{n} a_{ij} X_i X_j, \tag{1.20}$$

where $\boldsymbol{X}$ *denotes a random vector with mean* $\boldsymbol{\mu} = (\boldsymbol{\mu}_1, ..., \boldsymbol{\mu}_n)'$ *and covariance matrix* $\boldsymbol{\Sigma}$*.* $\boldsymbol{A} = (a_{ij})$ *is a* $n \times n$ *symmetric matrix.* $Q(\boldsymbol{X}) = \boldsymbol{X}'A\boldsymbol{X}$ *is called quadratic form in the random variable* X*.*

It is of the interest to have some properties of a quadratic forms with the random vector $\boldsymbol{x}$. As we have shown $\boldsymbol{x'Ax}$ is a quadratic form and without loss of generality we can assume that the matrix $\boldsymbol{A}$ is symmetric. Let's try to find out what is the expected value of a quadratic form $\boldsymbol{x'AX}$. Let's assume that $\boldsymbol{X} = [\boldsymbol{x}_1, ...\boldsymbol{x}_n]$ is an $n \times 1$ vector of random variables and $\boldsymbol{A}$ be a $n \times n$ symmetric matrix. $E(\boldsymbol{X}) = \boldsymbol{\mu}$ and $Var[\boldsymbol{x}] = \boldsymbol{\Sigma}$ then

$$E[\boldsymbol{X'AX}] = E[(\boldsymbol{X} - \boldsymbol{\mu})'\boldsymbol{A}(\boldsymbol{X} - \boldsymbol{\mu}) + \boldsymbol{\mu' AX} + \boldsymbol{X'A\mu} - \boldsymbol{\mu'\mu}]$$

$$and$$

$$\boldsymbol{X'A\mu} = (\boldsymbol{X'A\mu})' = \boldsymbol{\mu'A'X} = \boldsymbol{\mu'AX}$$

Hence

$$\begin{aligned}
&E[\boldsymbol{X'AX}] = E[(\boldsymbol{X} - \boldsymbol{\mu})'\boldsymbol{A}(\boldsymbol{X} - \boldsymbol{\mu})] + \boldsymbol{\mu'A\mu} = \\
&\sum_i \sum_j aijE[(X_i - \mu_i)(X_j - \mu_j)] + \boldsymbol{\mu'A\mu} = \\
&\qquad \sum_i \sum_j a_{ij}\sigma_{ij}\boldsymbol{\mu'A\mu} = \\
&\qquad tr[\boldsymbol{A\Sigma}] + \boldsymbol{\mu'A\mu}
\end{aligned}$$

Corollary 1. Having $\boldsymbol{Y} = \boldsymbol{X} - \boldsymbol{b}$

$$E[\boldsymbol{Y'AY}] = E[(\boldsymbol{X} - b)'\boldsymbol{A}(\boldsymbol{X} - b)] = tr[\boldsymbol{A\Sigma}] + (\boldsymbol{\mu} - b)'\boldsymbol{A}(\boldsymbol{\mu} - b)$$

Example 1. *As an example suppose* $X_1, X_2, ..., X_n$ *are independently and identically distributed with* $\boldsymbol{\mu}_i = 0$ *and we are looking for the expected value of the random variable* $H = \sum_{i=1}^{n-1}(X_i - X_{i+1})^2$

We can expand the sum and try to write it as a quadratic form. It yield $Q = 2\sum_{i=1}^{n} X_i^2 - X_1^2 - X_n^2 - 2\sum_{i=1}^{n-1} X_i X_{i=1}$ *Noticing coefficients of* X_i^2 *we get to know that diagonal of* $\boldsymbol{A}$ *should have all entries equal to 2 except first and last one. Then* $tr(\boldsymbol{A}) = 2n - 2$*. Since* $\boldsymbol{\Sigma} = \sigma^2 \boldsymbol{I}_n$*,* $Cov(X_i, X_j) = 0$ *for* $i \neq j$ *we have* $E[Q] = tr(\boldsymbol{A\Sigma}) = \sigma^2(2n-2)$ *since the second term of* $E[Q]$ *is zero in this case.*

Theorem 1.2.1. *If* $X_1, X_2, ..., X_n$ *are independent random variables with means* $\mu_1, \mu_2, ..., \mu_n$ *and the same variance of* μ_2 *and their third and forth moment about their mean are the same and equal to* μ_3 *and* μ_4 *If* $\boldsymbol{A}$ *is a symmetric matrix and a is the column vector of diagonal elements of* $\boldsymbol{A}$ *then*

$$Var[\boldsymbol{X'AX}] = (\mu_4 - 3\mu_2^2)\boldsymbol{a'a} + 2\mu_2^2 tr(\boldsymbol{A}^2) + 4\mu_2 \boldsymbol{\mu' A^2 \mu} + 4\mu_3 \boldsymbol{\mu' A} a,$$

in which $\boldsymbol{\mu}$ *denotes the vector of means of the variables.*

Proof. We know that $\mathrm{Var}(\boldsymbol{X}) = \mu_2 \boldsymbol{I}_n$

$$Var[\boldsymbol{X'AX}] = E[(\boldsymbol{X'AX})^2] - (E[\boldsymbol{X'AX}])^2$$

Now we express it using $\boldsymbol{\mu}$.

$$\begin{aligned}
\boldsymbol{X'AX} &= (\boldsymbol{X} - \boldsymbol{\mu})'\boldsymbol{A}(\boldsymbol{X} - \boldsymbol{\mu}) + 2\boldsymbol{\mu}'\boldsymbol{A}(\boldsymbol{X} - \boldsymbol{\mu}) + \boldsymbol{\mu}'\boldsymbol{A\mu} \\
(\boldsymbol{X'AX})^2 &= [(\boldsymbol{X} - \boldsymbol{\mu})'\boldsymbol{A}(\boldsymbol{X} - \boldsymbol{\mu})]^2 + 4[\boldsymbol{\mu}'\boldsymbol{A}(\boldsymbol{X} - \boldsymbol{\mu})]^2 + (\boldsymbol{\mu}'\boldsymbol{A\mu})^2 \\
&\quad + 2\boldsymbol{\mu}'\boldsymbol{A\mu}[(\boldsymbol{X} - \boldsymbol{\mu})'\boldsymbol{A}(\boldsymbol{X} - \boldsymbol{\mu}) + 2\boldsymbol{\mu}'\boldsymbol{A}(\boldsymbol{X} - \boldsymbol{\mu})] \\
&\quad + 4\boldsymbol{\mu}'\boldsymbol{A}(\boldsymbol{X} - \boldsymbol{\mu})(\boldsymbol{X} - \boldsymbol{\mu})'\boldsymbol{A}(\boldsymbol{X} - \boldsymbol{\mu})
\end{aligned}$$

Letting $\boldsymbol{Y} = \boldsymbol{X} - \boldsymbol{\mu}$ we have

$$\begin{aligned}
E[(\boldsymbol{X'AX})^2] &= E[(\boldsymbol{Y'AY})^2] + 4E[(\boldsymbol{\mu'AY})^2] + (\boldsymbol{\mu'A\mu})^2 + 2\boldsymbol{\mu'A\mu}(\mu_2 tr\boldsymbol{A}) \\
&\quad + 4E[\boldsymbol{\mu'A}YY'\boldsymbol{AY}]
\end{aligned}$$

To obtain the variance, first we have

$$E(\boldsymbol{Y}'\boldsymbol{A}\boldsymbol{Y})^2 = E[\sum_i \sum_j \sum_k \sum_l a_{ij}a_{kl}Y_iY_jY_kY_l].$$

Hence, for different indexes we have

$$E[Y_iY_jY_kY_l] = \begin{cases} \mu_4, & \text{if } (i=j=k=l) \\ \mu_2^2, & \text{if } (i=j, k=l; i=k, j=l; i=l, j=k) \\ 0, & \text{if}(otherwise) \end{cases}$$

So one can have

$$E[(\boldsymbol{Y}'\boldsymbol{A}\boldsymbol{Y})^2] = \mu_4 \sum_i a_{ii}^2 + \mu_2^2(\sum_{i\neq j}\sum a_{ii}a_{kk} + \sum_{i\neq j}\sum a_{ij}^2 + \sum_{i\neq j}\sum a_{ij}a_{ji})$$
$$(\mu_4 - 3\mu_2^2)a'a + \mu_2^2[(tr\boldsymbol{A})^2 + 2tr\boldsymbol{A}^2].$$

On the other hand we have

$$(\boldsymbol{\mu}'\boldsymbol{A}\boldsymbol{Y})^2 = (b'\boldsymbol{Y})^2 = \sum_i \sum_j b_ib_jY_iY_j$$
$$(\boldsymbol{\mu}'\boldsymbol{A}\boldsymbol{Y})(\boldsymbol{Y}'\boldsymbol{A}\boldsymbol{Y}) = \sum_i \sum_j \sum_k b_ia_{jk}Y_iY_jY_k$$
$$E[(\boldsymbol{\mu}'\boldsymbol{A}\boldsymbol{Y})^2] = \mu_2 \sum_i b_i^2 = \mu_2 b'b = \mu_2 \boldsymbol{\mu}'\boldsymbol{A}^2\boldsymbol{\mu}$$
$$E[(\boldsymbol{\mu}'\boldsymbol{A}\boldsymbol{Y})(\boldsymbol{Y}'\boldsymbol{A}\boldsymbol{Y})] = \mu_3 \sum_i b_ia_{ii} = \mu_3 b'a = \mu_3 \boldsymbol{\mu}'\boldsymbol{A}a$$

and finally by collecting all the terms and using

$$E[\boldsymbol{X}'\boldsymbol{A}\boldsymbol{X}] = \mu_2 tr\boldsymbol{A} + \boldsymbol{\mu}'\boldsymbol{A}\boldsymbol{\mu}$$

we come to the desired relation. □

As a very practical result for normal distribution we have an example

Example 2. *If $X_1, X_2, ..., X_n$ are normally distributed then $\mu_3 = 0$ and $\mu_4 = 3\mu_2^2$ we have*

$$Var[\boldsymbol{X}'\boldsymbol{A}\boldsymbol{X}] = 2\mu_2^2 tr\boldsymbol{A}^2 + 4\mu_2\boldsymbol{\mu}'\boldsymbol{A}\boldsymbol{\mu}$$

and if $\boldsymbol{\mu} = 0$ and $\mu_2 = \sigma^2$ then we have

$$Var[\boldsymbol{X}'\boldsymbol{A}\boldsymbol{X}] = 2\sigma^4 tr\boldsymbol{A}^2$$

It is quite usual in analysis of variance to partition and square the observed data. Therefore quadratic forms play an important part in many statistical applications.

1.3 Reviews on Quadratic Forms in Normal Random Variables

Let $\boldsymbol{Y}$ be a normal random vector with $\boldsymbol{\mu} = \mathbf{0}$ and $\boldsymbol{\Sigma} = \boldsymbol{I}_n$. Then $\boldsymbol{Y}'\boldsymbol{Y} = \sum_{i=1}^{n} y_i^2$ is distributed as $\chi^2(n)$ which is call a chi-square distribution with n degree of freedom. It is now of the interest to find the distribution of $\boldsymbol{Y}'\boldsymbol{Y}$ when $\boldsymbol{Y}$ is a normal random vector with distribution $N(\boldsymbol{\mu}, \boldsymbol{I})$.

Theorem 1.3.1. *Consider $y_1, y_2, ..., y_n$ as independent normal random variables with means $\mu_1, \mu_2, ..., \mu_n$ and $\boldsymbol{\Sigma} = \boldsymbol{I}_n$, then $\sum_{i=1}^{n} y_i^2 = W$ is distributed as a noncentral chi-square with noncentrality parameter $\lambda = \frac{1}{2}\boldsymbol{\Sigma}\mu_i^2$ and the distribution function would be as follows*

$$f(w) = e^{-\lambda} \sum_{i=0}^{\infty} \frac{\lambda^i w^{\frac{1}{2}(n+2i)-1} e^{-w/2}}{i! 2^{\frac{1}{2}(n+2i)} \Gamma(\frac{n+2i}{2})} \quad 0 \le w < \infty$$

Theorem 1.3.2. *If $W_1, W_2, ..., W_k$ are independent noncentral chi-squares such $W_i, i = 1, ..., k$ has p_i degree of freedom, then $\sum_{i=1}^{k} W_i$ has a noncentral chi-square distribution with degree of freedom $p = \sum_{i=1}^{k} p_i$*

Theorem 1.3.3. *If a $n \times 1$ vector $\boldsymbol{Y}$ has a density function of $N(\boldsymbol{\mu}, \boldsymbol{D})$ where $\boldsymbol{D}$ is diagonal, then $\boldsymbol{Y}'\boldsymbol{D}^{-1}\boldsymbol{Y}$ has the noncentral chi-square distribution with n degree of freedom and parameter $\lambda = \frac{1}{2}\boldsymbol{\mu}'\boldsymbol{D}^{-1}\boldsymbol{\mu}$*

Proof. This proof is to some extent dependent on what we reviewed about matrix algebra.

Since $\boldsymbol{D}$ is positive definite matrix, $\boldsymbol{D}$ is orthogonally diagonalizable to the identity matrix. Then there exist a nonsingular $n \times n$ matrix $\boldsymbol{B}$ such that $\boldsymbol{B'DB} = \boldsymbol{I}$ Let $\boldsymbol{Z} = \boldsymbol{B'DB} = \boldsymbol{I}$ By changing variable to $\boldsymbol{Z} = \boldsymbol{B'Z}$, it is not hard to confirm that $\boldsymbol{Z}$ is distributed as $N(\boldsymbol{B'\mu}, \boldsymbol{I})$. So $\boldsymbol{Z'Z}$ has the density function of noncentral chi-square with n degrees of freedom and parameter $\lambda = \frac{1}{2}\boldsymbol{\mu' BB' \mu}$. From $\boldsymbol{B'DB} = \boldsymbol{I}$ we can infer $\boldsymbol{BB'} = \boldsymbol{D}^{-1}$, then $\boldsymbol{ZZ'} = \boldsymbol{Y'BB'Y} = \boldsymbol{Y'D}^{-1}\boldsymbol{Y}$ and $\lambda = \frac{1}{2}\boldsymbol{\mu' BB' \mu} = \frac{1}{2}\boldsymbol{\mu' D}^{-1}\boldsymbol{\mu}$ □

Now we are ready to go through different representation of quadratic forms. We only consider quadratic forms in random variables in the nonsingular case.

Consider the random variable $\boldsymbol{X} = (X_1, ...X_p)'$ with $\boldsymbol{\mu} = (\boldsymbol{\mu}_1, ..., \boldsymbol{\mu}_p)'$ and covariance matrix $\boldsymbol{\Sigma} > 0$ then

$$\boldsymbol{Y} = \boldsymbol{\Sigma}^{-\frac{1}{2}}\boldsymbol{X} \rightarrow E(\boldsymbol{Y}) = \boldsymbol{\Sigma}^{-\frac{1}{2}}\boldsymbol{\mu} \quad \text{and} \quad Cov(\boldsymbol{Y}) = \boldsymbol{\Sigma}^{-\frac{1}{2}}Cov(\boldsymbol{X})\boldsymbol{\Sigma}^{-\frac{1}{2}} = \boldsymbol{I} \tag{1.21}$$

No considering the random variable $\boldsymbol{Z} = (\boldsymbol{Y} - \boldsymbol{\Sigma}^{-\frac{1}{2}}\boldsymbol{\mu})$ we have

$$E(\boldsymbol{Z}) = 0 \quad \text{and} \quad Cov(\boldsymbol{Z}) = \boldsymbol{I} \tag{1.22}$$

Writing $Q(\boldsymbol{X})$ in terms of $\boldsymbol{Z}$ would yield

$$Q(\boldsymbol{X}) = \boldsymbol{X'AX} = \boldsymbol{Y'}\boldsymbol{\Sigma}^{\frac{1}{2}}\boldsymbol{A}\boldsymbol{\Sigma}^{\frac{1}{2}}\boldsymbol{Y} = (\boldsymbol{Z} + \boldsymbol{\Sigma}^{-\frac{1}{2}}\boldsymbol{\mu})'\boldsymbol{\Sigma}^{\frac{1}{2}}\boldsymbol{A}\boldsymbol{\Sigma}^{\frac{1}{2}}(\boldsymbol{Z} + \boldsymbol{\Sigma}^{-\frac{1}{2}}\boldsymbol{\mu}) \tag{1.23}$$

It is noteworthy that any decomposition of the form $\boldsymbol{\Sigma} = \boldsymbol{BB'}$ where $\boldsymbol{B}$ is a $n \times n$ symmetric matrix and $|\boldsymbol{B}| \neq 0$ could be used instead of $\boldsymbol{\Sigma} = \boldsymbol{\Sigma}^{\frac{1}{2}}\boldsymbol{\Sigma}^{\frac{1}{2}}$. For the practical purposes a lower triangular matrix $\boldsymbol{B}$ would work well.

Now consider an orthogonal matrix $\boldsymbol{P}$ that diagonalizes $\boldsymbol{\Sigma}^{\frac{1}{2}}\boldsymbol{A}\boldsymbol{\Sigma}^{\frac{1}{2}}$ that is

$$\boldsymbol{P'}\boldsymbol{\Sigma}^{\frac{1}{2}}\boldsymbol{A}\boldsymbol{\Sigma}^{\frac{1}{2}}\boldsymbol{P} = \mathbf{diag}(\lambda_1, ..., \lambda_p), \boldsymbol{P'P} = \boldsymbol{I} \tag{1.24}$$

where $\lambda_1, ..., \lambda_p$ are the eigenvalues of $\boldsymbol{\Sigma}^{\frac{1}{2}}\boldsymbol{A}\boldsymbol{\Sigma}^{\frac{1}{2}}$ or $\boldsymbol{\Sigma A}$.

Now consider the random variable $\boldsymbol{U} = \boldsymbol{P'Z}$. One can have

$$E(\boldsymbol{U}) = E(\boldsymbol{P'Z}) = \boldsymbol{P'}E(\boldsymbol{Z}) = 0 \quad \text{and} \quad Cov(\boldsymbol{U}) = Cov(\boldsymbol{P'Z}) = \boldsymbol{P'\Sigma P} = \boldsymbol{I} \tag{1.25}$$

Hence writing $Q(\boldsymbol{X})$ in terms of $\boldsymbol{U}$ would yield

$$Q(\boldsymbol{X}) = (\boldsymbol{Z} + \boldsymbol{\Sigma}^{-\frac{1}{2}}\boldsymbol{\mu})'\boldsymbol{\Sigma}^{\frac{1}{2}}\boldsymbol{A}\boldsymbol{\Sigma}^{\frac{1}{2}}(\boldsymbol{Z} + \boldsymbol{\Sigma}^{-\frac{1}{2}}\boldsymbol{\mu}) \tag{1.26}$$

$$= (\boldsymbol{U} + \boldsymbol{b})'\boldsymbol{P}'\boldsymbol{\Sigma}^{\frac{1}{2}}\boldsymbol{A}\boldsymbol{\Sigma}^{\frac{1}{2}}\boldsymbol{P}(\boldsymbol{U} + \boldsymbol{b})$$

$$= (\boldsymbol{U} + \boldsymbol{b})'\mathrm{diag}(\lambda_1, ..., \lambda_p)(\boldsymbol{U} + \boldsymbol{b}) \tag{1.27}$$

where $\boldsymbol{U}' = (U_1, ..., U_p)$ and $\boldsymbol{b}' = (p'\boldsymbol{\Sigma}^{-\frac{1}{2}}\boldsymbol{\mu})' = (b_1, ..., b_p)$

Another representation of a quadratic form is as follows:

$$Q(\boldsymbol{X}) = \boldsymbol{X}'\boldsymbol{A}\boldsymbol{X} = \sum_{i=1}^{p} \lambda_j (U_j + b_j)^2, \boldsymbol{A} = \boldsymbol{A}' \tag{1.28}$$

and when $E(\boldsymbol{X}) = \boldsymbol{\mu} = 0$ all b_j's are zero and the quadratic form $Q(\boldsymbol{x})$ turns to $\sum_{j=1}^{p} \lambda_j U_j^2$. It will be shown in the next chapter that when $\boldsymbol{X} \sim N_p(\boldsymbol{\mu}, \boldsymbol{\Sigma})$,$\boldsymbol{\Sigma} > 0$, the quadratic form $Q(\boldsymbol{X})$ turns to a linear combination of central chi-square when $\boldsymbol{\mu} = 0$ and a linear combination of noncentral chi-square when $\boldsymbol{\mu} \neq 0$.

Example 3. *Consider the quadratic form* $Q(\boldsymbol{X}) = 4X_1^2 + 3X_2^2 + 2X_1X_2$, $\boldsymbol{X} \sim N_2(\boldsymbol{\mu}, \boldsymbol{\Sigma})$, $\boldsymbol{\mu}' = (-1, 1)$ *and*

$$\boldsymbol{\Sigma} = \begin{pmatrix} 1 & 1 \\ 1 & 2 \end{pmatrix}$$

As a decomposition for $\boldsymbol{\Sigma}$ *consider*

$$\boldsymbol{\Sigma} = \begin{pmatrix} 1 & 1 \\ 1 & 2 \end{pmatrix} = \boldsymbol{B}\boldsymbol{B}' \quad , \quad \boldsymbol{B} = \begin{pmatrix} 1 & 0 \\ 1 & 1 \end{pmatrix} \quad , \quad \boldsymbol{B}' = \begin{pmatrix} 1 & 1 \\ 0 & 1 \end{pmatrix}$$

$$\boldsymbol{B}^{-1} = \begin{pmatrix} 1 & 0 \\ -1 & 1 \end{pmatrix} \quad , \quad (\boldsymbol{B}^{-1})' = \begin{pmatrix} 1 & -1 \\ 0 & 1 \end{pmatrix}$$

As mentioned before the matrix $\boldsymbol{B}$ *can be used instead of* $\boldsymbol{\Sigma}^{\frac{1}{2}}$. *Writing* $Q(\boldsymbol{X})$ *in the matrix*

form we have

$$Q(\boldsymbol{X}) = \boldsymbol{X}'\boldsymbol{A}\boldsymbol{X} \quad , \quad \boldsymbol{A} = \boldsymbol{A}' = \begin{pmatrix} 4 & 1 \\ 1 & 3 \end{pmatrix}$$

$$\boldsymbol{B}'\boldsymbol{A}\boldsymbol{B} = \begin{pmatrix} 1 & 1 \\ 0 & 1 \end{pmatrix} \begin{pmatrix} 4 & 1 \\ 1 & 3 \end{pmatrix} \begin{pmatrix} 1 & 0 \\ 1 & 1 \end{pmatrix} = \begin{pmatrix} 9 & 4 \\ 4 & 3 \end{pmatrix}$$

$|\boldsymbol{B}'\boldsymbol{A}\boldsymbol{B} - \lambda \boldsymbol{I}| = 0 \rightarrow \lambda_1 = 1, \lambda_2 = 11$. *Thus the eigenvectors are*

$$p_1 = \begin{pmatrix} \frac{1}{\sqrt{5}} \\ -\frac{2}{\sqrt{5}} \end{pmatrix} \quad , \quad p_2 = \begin{pmatrix} \frac{2}{\sqrt{5}} \\ \frac{1}{\sqrt{5}} \end{pmatrix} \quad \textit{and let} \quad \boldsymbol{P} = \begin{pmatrix} \frac{1}{\sqrt{5}} & \frac{2}{\sqrt{5}} \\ -\frac{2}{\sqrt{5}} & \frac{1}{\sqrt{5}} \end{pmatrix}$$

$$\boldsymbol{B}^{-1}\boldsymbol{\mu} = \begin{pmatrix} 1 & 0 \\ -1 & 1 \end{pmatrix} \begin{pmatrix} -1 \\ -1 \end{pmatrix} = \begin{pmatrix} -1 \\ 0 \end{pmatrix}$$

$$\boldsymbol{P}'\boldsymbol{B}^{-1}\boldsymbol{\mu} = \begin{pmatrix} \frac{1}{\sqrt{5}} & -\frac{2}{\sqrt{5}} \\ \frac{2}{5} & \frac{1}{\sqrt{5}} \end{pmatrix} \begin{pmatrix} 1 & 0 \\ -1 & 1 \end{pmatrix} \begin{pmatrix} -1 \\ -1 \end{pmatrix} = \begin{pmatrix} -\frac{1}{\sqrt{5}} \\ -\frac{2}{\sqrt{5}} \end{pmatrix} = \begin{pmatrix} b_1 \\ b_2 \end{pmatrix} = b$$

Then

$$\boldsymbol{Y} = \boldsymbol{B}^{-1}\boldsymbol{X} = \begin{pmatrix} 1 & 0 \\ -1 & 1 \end{pmatrix} \begin{pmatrix} X_1 \\ X_2 \end{pmatrix} = \begin{pmatrix} X_1 \\ -X_1 + X_2 \end{pmatrix}$$

$$\boldsymbol{Z} = \begin{pmatrix} \boldsymbol{Z}_1 \\ \boldsymbol{Z}_2 \end{pmatrix} = \boldsymbol{Y} - \boldsymbol{B}^{-1}\boldsymbol{\mu} = \begin{pmatrix} Y_1 \\ Y_2 \end{pmatrix} - \begin{pmatrix} -1 \\ 0 \end{pmatrix} = \begin{pmatrix} Y_1 + 1 \\ Y_2 \end{pmatrix} = \begin{pmatrix} X_1 + 1 \\ -X_1 + X_2 \end{pmatrix}$$

$$\boldsymbol{U} = \begin{pmatrix} \boldsymbol{U}_1 \\ \boldsymbol{U}_2 \end{pmatrix} = \boldsymbol{P}'\boldsymbol{Z} = \begin{pmatrix} \frac{1}{\sqrt{5}} & -\frac{2}{\sqrt{5}} \\ \frac{2}{\sqrt{5}} & \frac{1}{\sqrt{5}} \end{pmatrix} \begin{pmatrix} X_1 + 1 \\ -X_1 + X_2 \end{pmatrix} = \begin{pmatrix} \frac{1}{\sqrt{5}}(3X_1 - 2X_2 + 1) \\ \frac{1}{\sqrt{5}}(X_1 + X_2 + 2) \end{pmatrix}$$

and now it is easy to check that $Q(\boldsymbol{X}) = \lambda_1(U_1 + b_1)^2 + \lambda_2(U_2 + b_2)^2$ *is the same as* $Q(\boldsymbol{X}) = 4X_1^2 + 3X_2^2 + 2X_1X_2$

Now we are ready to examine moment generating function of Quadratic forms in Normal variables. We will investigate nonsingular Normal case as we only need this case.

Definition 1.3.1. *Let* $\boldsymbol{X} \sim N_p(\boldsymbol{\mu}, \boldsymbol{\Sigma}), \boldsymbol{\Sigma} > 0$, $Q = \boldsymbol{X}'\boldsymbol{A}\boldsymbol{X}$, $\boldsymbol{A} = \boldsymbol{A}'$. *The moment generating function of* Q *is given by*

$$M_Q(t) = E(e^{tQ}) = \int_{\boldsymbol{x}} \frac{exp\{tx'\boldsymbol{A}\boldsymbol{x} - \frac{1}{2}(\boldsymbol{x}-\boldsymbol{\mu})'\boldsymbol{\Sigma}^{-1}(\boldsymbol{x}-\boldsymbol{\mu})\}}{(2\pi)^{\frac{p}{2}}|\boldsymbol{\Sigma}^{\frac{1}{2}}|}dx \tag{1.29}$$

Using the expansion of $(\boldsymbol{x}-\boldsymbol{\mu})'\boldsymbol{\Sigma}^{-1}(\boldsymbol{x}-\boldsymbol{\mu})$ one can have

$$\begin{aligned} t\boldsymbol{x}'\boldsymbol{A}\boldsymbol{x} - \frac{1}{2}(\boldsymbol{x}-\boldsymbol{\mu})'\boldsymbol{\Sigma}^{-1}(\boldsymbol{x}-\boldsymbol{\mu}) & \\ &= -\frac{1}{2}\boldsymbol{\mu}'\boldsymbol{\Sigma}^{-1}\boldsymbol{\mu} + \frac{1}{2}\boldsymbol{\mu}'(\boldsymbol{I} - 2t\boldsymbol{A}\boldsymbol{\Sigma})^{-1}\boldsymbol{\Sigma}'\boldsymbol{\mu} \\ &\quad -\frac{1}{2}(\boldsymbol{x}-c)'(\boldsymbol{\Sigma}^{-1} - 2t\boldsymbol{A})(\boldsymbol{x}-c) \end{aligned}$$

where

$$c = (\boldsymbol{\Sigma}^{-1} - 2ta)^{-1}\boldsymbol{\Sigma}^{-1}\boldsymbol{\mu}$$

Since $|t|$ can be in an arbitrarily small neighborhood of 0, $(\boldsymbol{\Sigma}^{-1} - 2t\boldsymbol{A})$ is assumed symmetric positive definite without any loss of generality. Letting $\boldsymbol{y} = (\boldsymbol{\Sigma}^{-1} - 2t\boldsymbol{A})^{\frac{1}{2}}(\boldsymbol{x} - c)$, one has $|\boldsymbol{\Sigma}^{-1} - 2t\boldsymbol{A}|^{\frac{1}{2}}dx = dy$. By integrating out of $\boldsymbol{y}$ and noting that $\int_{\boldsymbol{y}} \exp(-\frac{1}{2}\boldsymbol{y}'\boldsymbol{y})d\boldsymbol{y} = (2\pi)^{\frac{p}{2}}$ Mathai and Provost (1992b) Then one can write

$$\begin{aligned} M_Q(t) &= |\boldsymbol{I} - 2t\boldsymbol{A}\boldsymbol{\Sigma}|^{-\frac{1}{2}}\exp\{-\frac{1}{2}\boldsymbol{\mu}'[\boldsymbol{I} - (\boldsymbol{I} - 2t\boldsymbol{A}\boldsymbol{\Sigma})^{-1}]\boldsymbol{\Sigma}^{-1}\boldsymbol{\mu}\} \\ &= |\boldsymbol{I} - 2t\boldsymbol{A}\boldsymbol{\Sigma}|^{-\frac{1}{2}}\exp\{-\frac{1}{2}\boldsymbol{\mu}'\boldsymbol{\Sigma}^{-1}\boldsymbol{\mu} + \frac{1}{2}\boldsymbol{\mu}'(\boldsymbol{I} - 2t\boldsymbol{A}\boldsymbol{\Sigma})^{-1}\boldsymbol{\Sigma}^{-1}\boldsymbol{\mu}\} \\ &= |\boldsymbol{I} - 2t\boldsymbol{A}\boldsymbol{\Sigma}|^{-\frac{1}{2}}\exp\{t\boldsymbol{\mu}'\boldsymbol{A}\boldsymbol{\Sigma}(\boldsymbol{I} - 2t\boldsymbol{A}\boldsymbol{\Sigma})^{-1}\boldsymbol{\Sigma}^{-1}\boldsymbol{\mu}\} \\ &= |\boldsymbol{I} - wt\boldsymbol{\Sigma}^{\frac{1}{2}}\boldsymbol{A}\boldsymbol{\Sigma}^{\frac{1}{2}}|^{\frac{1}{2}}\exp\{-\frac{1}{2}\boldsymbol{\mu}'\boldsymbol{\Sigma}^{-1}\boldsymbol{\mu} \\ &\qquad + \frac{1}{2}\boldsymbol{\mu}'\boldsymbol{\Sigma}^{-\frac{1}{2}}(\boldsymbol{I} - 2t\boldsymbol{\Sigma}^{\frac{1}{2}}\boldsymbol{A}\boldsymbol{\Sigma}^{\frac{1}{2}})^{-1}\boldsymbol{\Sigma}^{-\frac{1}{2}}\boldsymbol{\mu}\} \\ &= |\boldsymbol{I} - 2t\boldsymbol{\Sigma}^{\frac{1}{2}}\boldsymbol{A}\boldsymbol{\Sigma}^{\frac{1}{2}}|^{-\frac{1}{2}}\exp\{t\boldsymbol{\mu}'\boldsymbol{\Sigma}^{-\frac{1}{2}}(\boldsymbol{\Sigma}^{\frac{1}{2}}\boldsymbol{A}\boldsymbol{\Sigma}^{\frac{1}{2}}) \\ &\qquad (\boldsymbol{I} - 2t\boldsymbol{\Sigma}^{\frac{1}{2}}\boldsymbol{A}\boldsymbol{\Sigma}^{\frac{1}{2}})^{-1}\boldsymbol{\Sigma}^{-\frac{1}{2}}\boldsymbol{\mu}\} \end{aligned}$$

It is worthy of note to extend the moment generating function to $Q^\star(\boldsymbol{X}) = \boldsymbol{X}'\boldsymbol{A}\boldsymbol{X} + \boldsymbol{a}'\boldsymbol{x} + d$, $\boldsymbol{A} = \boldsymbol{A}'$ following the same steps. Then we have

$$\begin{aligned}
M_{Q^\star}(t) &= |\boldsymbol{I} - 2ta\boldsymbol{\Sigma}|^{-\frac{1}{2}}\exp\{-\frac{1}{2}(\boldsymbol{\mu}'\boldsymbol{\Sigma}^{-1}\boldsymbol{\mu} - 2td) \\
&\quad + \frac{1}{2}(\boldsymbol{\mu} + t\boldsymbol{\Sigma}a)'(\boldsymbol{I} - 2t\boldsymbol{A}\boldsymbol{\Sigma})^{-1}\boldsymbol{\Sigma}^{-1}(\boldsymbol{\mu} + t\boldsymbol{\Sigma})\} \\
&= |\boldsymbol{I} - 2t\boldsymbol{\Sigma}^{\frac{1}{2}}\boldsymbol{A}\boldsymbol{\Sigma}^{\frac{1}{2}}|^{-\frac{1}{2}}\exp\{t(d + \boldsymbol{\mu}'\boldsymbol{A}\boldsymbol{\mu} + a'\boldsymbol{\mu}) \\
&\quad + \frac{t^2}{2}(\boldsymbol{\Sigma}^{\frac{1}{2}}a + 2\boldsymbol{\Sigma}^{\frac{1}{2}}\boldsymbol{A}\boldsymbol{\mu})'(\boldsymbol{I} - 2t\boldsymbol{\Sigma}^{\frac{1}{2}}\boldsymbol{A}\boldsymbol{\Sigma}^{\frac{1}{2}})^{-1} \\
&\qquad (\boldsymbol{\Sigma}^{\frac{1}{2}}a + 2\boldsymbol{\Sigma}^{\frac{1}{2}}\boldsymbol{A}\boldsymbol{\mu})\}
\end{aligned}$$

As a corollary for $\boldsymbol{X} \sim N_p(0, \boldsymbol{\Sigma})$ we have

$$M_{Q^\star} = |\boldsymbol{I} - 2t\boldsymbol{A}\boldsymbol{\Sigma}|^{-\frac{1}{2}}\exp\{td + \frac{t^2}{2}a'\boldsymbol{\Sigma}^{\frac{1}{2}}(\boldsymbol{I} - 2t\boldsymbol{\Sigma}^{\frac{1}{2}}\boldsymbol{A}\boldsymbol{\Sigma}^{\frac{1}{2}})^{-1}\boldsymbol{\Sigma}^{\frac{1}{2}}a\}$$

$$M_Q(t) = |\boldsymbol{I} - 2t\boldsymbol{A}\boldsymbol{\Sigma}|^{-\frac{1}{2}}$$

By some calculations one can write

$$|\boldsymbol{I} - 2t\boldsymbol{A}\boldsymbol{\Sigma}|^{-\frac{1}{2}} = |\boldsymbol{I} - 2t\boldsymbol{\Sigma}^{\frac{1}{2}}\boldsymbol{A}\boldsymbol{\Sigma}^{-\frac{1}{2}}| = \prod_{i=1}^{p}(1 - 2t\lambda_j)^{-\frac{1}{2}} \tag{1.30}$$

where $\lambda_1, ..., \lambda_p$ are the eigenvalues of $\boldsymbol{\Sigma}^{\frac{1}{2}}\boldsymbol{A}\boldsymbol{\Sigma}^{\frac{1}{2}}$.

Now we illustrate the procedure to obtain the m.g.f of a quadratic form through the use of an example.

Example 4. *Consider the quadratic form* $Q(\boldsymbol{X}) = 8X_1^2 - 8X_1X_2 + 3X_2^2$ *when* $\boldsymbol{X} \sim N_2(\boldsymbol{\mu}, \boldsymbol{\Sigma})$, $\boldsymbol{\mu}'(-1, 1)$, *and*

$$\boldsymbol{\Sigma} = \begin{pmatrix} 1 & 1 \\ 1 & 2 \end{pmatrix}$$

This example is adopted from Mathai and Provost (1992b).

Obtaining the eigenvalues $\lambda_1 = 2$, $\lambda_2 = 4$

$$\boldsymbol{b} = \begin{pmatrix} \frac{1}{\sqrt{2}} \\ -\frac{3}{\sqrt{2}} \end{pmatrix} \quad , \quad \boldsymbol{P} = \begin{pmatrix} \frac{1}{\sqrt{2}} & \frac{1}{\sqrt{2}} \\ \frac{1}{\sqrt{2}} & -\frac{1}{\sqrt{2}} \end{pmatrix} \quad , \quad \boldsymbol{\Sigma} = \boldsymbol{B}\boldsymbol{B}'$$

Thus one can see $\sum_{j=1}^{2} = (\frac{1}{2}) + (\frac{9}{2}) = 5$, *and*

$$M_Q(t) = (1-4t)^{-\frac{1}{2}}(1-8t)^{-\frac{1}{2}} exp\{-\frac{5}{2} + \frac{1}{4}(1-4t)^{-1} + \frac{3}{4}(1-8t)^{-1}\}$$

As other functions of the random variables it is possible to computer moments of quadratic forms. We investigate obtaining moments through moments generating function and through direct computation via particular representations of quadratic forms.

We have seen in equation (1.28) that a quadratic form can be represented as $\sum_{i=1}^{n} \lambda_j(U_j + b_j)$, where $\lambda_1, ..., \lambda_p$ are the eigenvalues of $\boldsymbol{\Sigma}^{\frac{1}{2}}\boldsymbol{A}\boldsymbol{\Sigma}^{\frac{1}{2}}$ or $\boldsymbol{\Sigma}\boldsymbol{A}$, $\boldsymbol{U}' = (U_1, ..., U_p) = \boldsymbol{P}'\boldsymbol{Z}$, and $\boldsymbol{b}' = (p'\boldsymbol{\Sigma}^{-\frac{1}{2}}\boldsymbol{\mu})' = (b_1, ..., b_p)$ such that $\boldsymbol{Z} = \boldsymbol{\Sigma}^{-\frac{1}{2}}(\boldsymbol{X} - \boldsymbol{\mu})$.
This representation is easier to deal with for obtaining moments of a quadratic form. By applying expectation function on the quadratic form Q we have

$$E[Q(\boldsymbol{X})]^r = E[\sum_{j=1}^{p}(U_j + b_j)2]^r \tag{1.31}$$

$$= \sum_{r_1+r+2+...+r_p=r} ... \sum \frac{r!\lambda_1^{r_1}...\lambda_p^{r_p}}{r_1!...r_p!} E(V_1^{r_1}...V_p^{r_p}) \tag{1.32}$$

where $V_j = (U_j + b_j)2$. When V_1,...,V_p are independently distributed then $E(V_1^{r_1}...V_p^{r_p}) = \boldsymbol{\mu}_{r_1}^{(1)}...\boldsymbol{\mu}_{r_p}^{(p)}$, $\boldsymbol{\mu}_{r_j}^{(j)} = E(V_j^{r_j})$, $j = 1, ..., p$ Mathai and Provost (1992b).

Theorem 1.3.4. *Consider the quadratic form* $Q(\boldsymbol{X}) = \boldsymbol{X}'\boldsymbol{A}\boldsymbol{X}$, $\boldsymbol{A} = \boldsymbol{A}'$, $E(\boldsymbol{X}) = \boldsymbol{\mu}$, $Cov(\boldsymbol{X}) = \boldsymbol{\Sigma} > 0$. *Then for* $r = 1, 2, ..$ *then*

$$E(Q(\boldsymbol{X})]^r = \sum_{r_1+...+r_p} ... \sum \frac{r!\lambda_1^{r_1}...\lambda_p^{r_p}}{r_1!...r_p!} E(V_1^{r_1}...V_p^{r_p)} \tag{1.33}$$

where $\lambda_1, ..., \lambda_p$ are the eigenvalues of $\boldsymbol{\Sigma}^{\frac{1}{2}} \boldsymbol{A} \boldsymbol{\Sigma}^{\frac{1}{2}}$, and $V_j = (U_j + b_j)^2$. U_j's are noncorrelated with $E(U_j) = 0$ and $Var(U_j) = 1$, $\boldsymbol{B}' = (b_1, ..., b_p) = \boldsymbol{\mu}' \boldsymbol{\Sigma}^{-\frac{1}{2}} \boldsymbol{P}$ such that $\boldsymbol{P}\boldsymbol{P}' = \boldsymbol{I}$

Note that in (1.33) there is no assumed distribution for the random variable $\boldsymbol{X}$.
The expression for lower moments of the quadratic form is simple. For example for $r = 1$ one can have:

$$E[Q(\boldsymbol{X})] = E\left\{\sum_{j=1}^{p} \lambda_j (U_j + b_j)^2\} = E\{\sum_{j=1}^{p} \lambda_j (U_j^2 + 2b_j U_j + b_j^2)\right\} \tag{1.34}$$

$$= \sum_{j=1}^{p} \lambda_j + \sum_{j=1}^{p} \lambda_j b_j^2, \tag{1.35}$$

If the $U_j, j = 1, ...p$ have mean 0 and variance 1, then

$$E[Q(\boldsymbol{X})] = tr \boldsymbol{\Sigma}^{\frac{1}{2}} \boldsymbol{A} \boldsymbol{\Sigma}^{\frac{1}{2}} + \boldsymbol{b}' \mathrm{diag}(\lambda_1, ...\lambda_p) \boldsymbol{b} \quad , \quad \boldsymbol{b} = \boldsymbol{\mu} \boldsymbol{\Sigma}^{-\frac{1}{2}} \boldsymbol{P} \tag{1.36}$$

By considering that $\mathrm{diag}(\lambda_1, ..., \lambda_p) = \boldsymbol{P}' \boldsymbol{\Sigma}^{\frac{1}{2}} \boldsymbol{A} \boldsymbol{\Sigma}^{\frac{1}{2}} \boldsymbol{P}$ and $\boldsymbol{P}\boldsymbol{P}' = \boldsymbol{I}$, we have

$$E[Q(\boldsymbol{X})] = tr \boldsymbol{A} \boldsymbol{\Sigma} + \boldsymbol{\mu}' \boldsymbol{A} \boldsymbol{\mu} \tag{1.37}$$

One can obtain the mean of a quadratic form directly. Since $\boldsymbol{X}' \boldsymbol{A} \boldsymbol{X}$ is a scalar its trace is itself. Thus

$$\begin{aligned} E(\boldsymbol{X}' \boldsymbol{A} \boldsymbol{X}) &= tr E(\boldsymbol{X}' \boldsymbol{A} \boldsymbol{X}) = E tr(\boldsymbol{A} \boldsymbol{X} \boldsymbol{X}') = E tr(\boldsymbol{A} \boldsymbol{X} \boldsymbol{X}') = tr \boldsymbol{A} [\boldsymbol{\Sigma} + \boldsymbol{\mu} \boldsymbol{\mu}'] \\ &= tr \boldsymbol{A} \boldsymbol{\Sigma} + tr(\boldsymbol{A} \boldsymbol{\mu} \boldsymbol{\mu}') \\ &= tr \boldsymbol{A} \boldsymbol{\Sigma} + tr \boldsymbol{\mu}' \boldsymbol{A} \boldsymbol{\mu} = tr \boldsymbol{A} \boldsymbol{\Sigma} + \boldsymbol{\mu}' \boldsymbol{A} \boldsymbol{\mu} \end{aligned}$$

Clearly $\boldsymbol{A}$ does not have to be symmetric, and $\boldsymbol{\Sigma}$ can be singular or nonsingular.

One can obtain the moments of a random variable by taking derivatives of its moment generating function. We now apply this rule on the moment generating function of a quadratic form.

Lemma 1. Let $f(t)$ be a differentiable scalar function with the kth derivative function $g^{(k)}(t)$. Then by use of Leibnitz's rule for differentiation the rth derivative can be written as

$$\frac{d^r}{dt^r}f(t) = \frac{d^{r-1}}{dt^{r-1}}[g(t)f(t)] \quad , \quad g(t) = \frac{d}{dt}\ln f(t) \tag{1.38}$$
$$= \{\sum_{r_1=0}^{r-1}\binom{r-1}{r_1}g^{(r-1-r_1)}(t)\sum_{r_2=0}^{r_1-1}\binom{r_1-1}{r_2}g^{(r_1-1-r_2)}(t)...\}f(t)$$

Now considering $g_Q(t) = \frac{d}{dt}\ln M_Q(t)$ and using 4 one may note that

$$g_Q(t) = \frac{d}{dt}lnM_Q(t) = \sum_{j=1}^{p} b_j^2\lambda_j(1-2t\lambda_j)^{-2} + \sum_{j=1}^{p}\lambda_j(1-2t\lambda_j)^{-1}$$

where b_j's and λ_j's are the same as stated before. Thus,

$$\frac{d^{r-1}}{dt^{r-1}}g_Q(t) = \frac{d^r}{dt^r}\ln M_Q(t)$$
$$= 2^{r-1}r!\sum_{j=1}^{p} b_j^2\lambda_j^r(1-2t\lambda_j\boldsymbol{I})^{-(r+1)} \tag{1.39}$$
$$+ 2^{r-1}(r-1)!\sum_{j=1}^{p}\lambda_j^r(1-2t\lambda_j)^{-r}$$

Lemma 2. Let $\boldsymbol{X}$ be a p-variate random variable such that $\boldsymbol{X} \sim N_p(\boldsymbol{\mu}, \boldsymbol{\Sigma})$, $\boldsymbol{\Sigma} > 0$, and $Q(\boldsymbol{X}) = \boldsymbol{X'AX}$, where $\boldsymbol{A}$ is a symmetric matrix. Then the r-th logarithmic derivative of the m.g.f of Q evaluated at $t = 0$ is given by

$$\frac{d^r}{dt^r}\ln M_Q(t)|_{t=0} = \frac{d^{r-1}}{dt^{r-1}}g_Q(t)|_{t=0} = g_Q^{(r-1)}(t)|_{t=0}$$
$$= 2^{r-1}(r-1)!\sum_{j=1}^{p}\lambda_j^r[1+rb_j^2] \tag{1.40}$$
$$= 2^{r-1}(r-1)!\{tr(\boldsymbol{\Sigma}^{\frac{1}{2}}\boldsymbol{A}\boldsymbol{\Sigma}^{\frac{1}{2}})^r + r\boldsymbol{\mu}'\boldsymbol{\Sigma}^{-\frac{1}{2}}(\boldsymbol{\Sigma}^{\frac{1}{2}}\boldsymbol{A}\boldsymbol{\Sigma}^{\frac{1}{2}})^r\boldsymbol{\Sigma}^{-\frac{1}{2}}$$
$$= 2^{r-1}(r-1)!\{tr(\boldsymbol{A\Sigma})^r + r\boldsymbol{\mu}'(\boldsymbol{A\Sigma})^{r-1}\boldsymbol{A\mu}\} \tag{1.41}$$

The following statement will help to have a better insight into the above relation.

$$
\begin{aligned}
\operatorname{diag}(\lambda_1^r, ..., \lambda_p^r) &= \boldsymbol{P}'(\boldsymbol{\Sigma}^{\frac{1}{2}}\boldsymbol{A}\boldsymbol{\Sigma}^{\frac{1}{2}})\boldsymbol{P} \quad , \quad \boldsymbol{P}\boldsymbol{P}' = \boldsymbol{I} \\
\sum_{j=1}^{p} b_j^2 \lambda_j^r &= \boldsymbol{b}'\operatorname{diag}(\lambda_1^r, ..., \lambda_p^r)\boldsymbol{b} \quad , \quad \boldsymbol{b} = \boldsymbol{P}'\boldsymbol{\Sigma}^{-\frac{1}{2}}\boldsymbol{\mu} \\
&= (\boldsymbol{\mu}\boldsymbol{\Sigma}^{-\frac{1}{2}}\boldsymbol{P})(\boldsymbol{P}'[\boldsymbol{\Sigma}^{\frac{1}{2}}\boldsymbol{A}\boldsymbol{\Sigma}^{\frac{1}{2}}]^r\boldsymbol{P})\boldsymbol{P}'\boldsymbol{\Sigma}^{-\frac{1}{2}}\boldsymbol{\mu} \\
&= \boldsymbol{\mu}\boldsymbol{\Sigma}^{-\frac{1}{2}}(\boldsymbol{\Sigma}^{\frac{1}{2}}\boldsymbol{A}\boldsymbol{\Sigma}^{\frac{1}{2}})^r\boldsymbol{\Sigma}^{-\frac{1}{2}}\boldsymbol{\mu} = \boldsymbol{\mu}'(\boldsymbol{A}\boldsymbol{\Sigma})^{r-1}\boldsymbol{A}\boldsymbol{\mu} \\
\sum_{j=1}^{p} \lambda_j^r &= tr(\boldsymbol{\Sigma}^{\frac{1}{2}}\boldsymbol{A}\boldsymbol{\Sigma}^{\frac{1}{2}})^r = tr(\boldsymbol{A}\boldsymbol{\Sigma})^r
\end{aligned}
$$

Equation (1.38) leads us to following theorem:

Theorem 1.3.5. *Let $Q(\boldsymbol{X}) = \boldsymbol{X}'\boldsymbol{A}\boldsymbol{X}$ for $\boldsymbol{X} \sim N_p(\boldsymbol{\mu}, \boldsymbol{\Sigma})$ and $\boldsymbol{A} = \boldsymbol{A}'$. Then the rth moment of $Q(\boldsymbol{X})$ is given by*

$$
E[Q(\boldsymbol{X})]^r = \left\{ \sum_{r_1=0}^{r-1} \binom{r-1}{r_1} g^{(r-1-r_1)} \sum_{r_2=0}^{r_1-1} \binom{r_1-1}{r_2} g^{(r_1-1-r_2)} \dots \right\} \tag{1.42}
$$

where $g^{(k)} = 2^k k! \{tr(\boldsymbol{A}\boldsymbol{\Sigma})^{k+1} + (k+1)\boldsymbol{\mu}'(\boldsymbol{A}\boldsymbol{\Sigma})^k\boldsymbol{A}\boldsymbol{\mu}\}$, for $k = 0, 1, 2, ...$

Now we are ready to explore some useful moments of a quadratic form in normal random variable. To illustrate the above theorem the moments of the quadratic form are presented below.

$$
E[Q(\boldsymbol{X})] = g^{(0)} = tr(\boldsymbol{A}\boldsymbol{\Sigma}) + \boldsymbol{\mu}'\boldsymbol{A}\boldsymbol{\mu} \tag{1.43}
$$

$$
\begin{aligned}
E[Q(\boldsymbol{X})]^2 &= g^{(1)} + [g^{(0)}]^2 \\
&= 2[tr(\boldsymbol{A}\boldsymbol{\Sigma})^2 + 2\boldsymbol{\mu}'(\boldsymbol{A}\boldsymbol{\Sigma})\boldsymbol{A}\boldsymbol{\mu}] + [tr(\boldsymbol{A}\boldsymbol{\Sigma}) + \boldsymbol{\mu}'\boldsymbol{A}\boldsymbol{\mu}]^2
\end{aligned} \tag{1.44}
$$

$$
\begin{aligned}
E[Q(\boldsymbol{X})]^3 &= g^{(2)} + 3g^{(1)}g^{(0)} + [g^{(0)}]^3 \\
&= 8[tr(\boldsymbol{A}\boldsymbol{\Sigma})^3 + 3\boldsymbol{\mu}'(\boldsymbol{A}\boldsymbol{\Sigma})^2\boldsymbol{A}\boldsymbol{\mu}] \\
&+ 6[tr(\boldsymbol{A}\boldsymbol{\Sigma})^2 + 2\boldsymbol{\mu}'\boldsymbol{A}\boldsymbol{\Sigma}\boldsymbol{\mu}][tr\boldsymbol{A}\boldsymbol{\Sigma} + \boldsymbol{\mu}'\boldsymbol{A}\boldsymbol{\mu}] \\
&+ [tr(\boldsymbol{A}\boldsymbol{\Sigma}) + \boldsymbol{\mu}'\boldsymbol{A}\boldsymbol{\mu}]^3
\end{aligned} \tag{1.45}
$$

Having the first and second moment of the quadratic form, it is of the interest to obtain its variance.

$$\text{Var}(Q(\boldsymbol{X})) = E[Q(\boldsymbol{X})]^2 - E^2[Q(\boldsymbol{X})] = g^{(1)} = 2tr(\boldsymbol{A\Sigma})^2 + 4\boldsymbol{\mu}'\boldsymbol{A\Sigma A\mu} \tag{1.46}$$

Now we can move on to cumulants and cumulant generating function of a quadratic form. We will also consider some special cases of interest.

Definition 1.3.2. *Consider the scalar variable X with its moment generating function (m.g.f) $M_X(t)$. The logarithm ln $M_X(t)$ is defined as the cumulant generating function of X.*

Just like the moment generating function generates moments of random variables the cumulant generating function can produce cumulants of random variables.

Definition 1.3.3. *Let $lnM_X(t)$ be the cumulant generating function of random variable X. If $lnM_X(t)$ admits a power series expansion, then the coefficients of $t^s/s!$ is called s'th cumulant of X and it is denoted by K_s. Hence, one can write*

$$lnM_{\boldsymbol{X}}(t) = \sum_{s=1}^{\infty} K_s \frac{t^s}{s!} \tag{1.47}$$

Theorem 1.3.6. *Suppose $\boldsymbol{X} \sim N_p(\boldsymbol{\mu}, \boldsymbol{\Sigma})$ and $\boldsymbol{\Sigma}$ is a positive definite matrix. The s'th cumulant of $Q = \boldsymbol{X}'\boldsymbol{AX} + a'\boldsymbol{X} + d$ is as follows:*

$$\begin{aligned} K_s &= 2^{s-1}s!\Big\{\frac{tr(\boldsymbol{A\Sigma})^s}{s} + \frac{1}{4}\boldsymbol{a}'(\boldsymbol{\Sigma A})^{s-2}\boldsymbol{\Sigma a} + \boldsymbol{\mu}'(\boldsymbol{A\Sigma})^{s-1}\boldsymbol{A\mu} \\ &\quad + \boldsymbol{a}'(\boldsymbol{\Sigma A})^{s-1}\boldsymbol{A\mu}\Big\} \quad , \quad \textit{for } s \geq 2; \qquad (1.48) \\ &= tr(\boldsymbol{A\Sigma}) + \boldsymbol{\mu}'\boldsymbol{A\mu} + \boldsymbol{a}'\boldsymbol{\mu} + \boldsymbol{d} \quad , \quad \textit{for } s = 1 \qquad (1.49) \end{aligned}$$

By letting a and d to be zero, one can obtain cumulant generating function for $Q(\boldsymbol{X}) = \boldsymbol{X}'\boldsymbol{AX}$.

$$\begin{aligned} K_s &= 2^{s-1}s!\Big\{\frac{tr(\boldsymbol{A\Sigma})^s}{s} + \boldsymbol{\mu}'(\boldsymbol{A\Sigma})^{s-1}\boldsymbol{A\mu}\Big\} \quad , \quad s \geq 2 \qquad (1.50) \\ &= tr(\boldsymbol{A\Sigma}) + \boldsymbol{\mu}'\boldsymbol{A\mu} \quad \text{for } s = 1 \qquad (1.51) \end{aligned}$$

We finish this chapter with an interesting point that $K_1 = E(\boldsymbol{X})$ and $K_2 = \mathrm{Var}(\boldsymbol{X})$. One may also note that an alternative representation of K_s for $Q = \boldsymbol{X}'\boldsymbol{A}\boldsymbol{X}$ is

$$2^{s-1}s!\sum_{j=1}^{s}\lambda_j^s(b_j^2+1/s) = 2^{s-1}(s-1)!\sum_{j=1}^{s}\lambda_j^s(sb_j^2+1) \quad , \quad s \geq 1 \tag{1.52}$$

where λ_j's are eigenvalues of $\boldsymbol{\Sigma}^{\frac{1}{2}}\boldsymbol{A}\boldsymbol{\Sigma}^{\frac{1}{2}}$ and $\boldsymbol{b} = \boldsymbol{P}'\boldsymbol{\Sigma}^{-\frac{1}{2}}\boldsymbol{\mu}$. Mathai and Provost (1992b)

Chapter 2

The Distribution of Quadratic Forms

The distribution of quadratic forms has been studied by many authors, and various representations of the distribution function are provided. Most of the representations concern the case of multivariate normality. The theorems and examples are adopted from Mathai and Provost (1992b).

Consider the random variable $\boldsymbol{X} \sim N_p(\boldsymbol{\mu}, \boldsymbol{\Sigma}), \boldsymbol{\Sigma} > 0$, then the quadratic form $Q = \boldsymbol{X'AX}$, with the symmetric matrix $\boldsymbol{A}$, has the following representation.

$$Q = \sum_{j=1}^{p} \lambda_j (U_j + b_j)^2$$

where

$$\boldsymbol{P'\Sigma^{\frac{1}{2}}A\Sigma^{\frac{1}{2}}P} = diag(\lambda_1, ..., \lambda_p), \boldsymbol{PP'} = \boldsymbol{I}$$

$\boldsymbol{U'} = (U_1, ..., U_P), \boldsymbol{U} = \boldsymbol{P'\Sigma^{-\frac{1}{2}}}(\boldsymbol{X} - \boldsymbol{\mu})$, and $\boldsymbol{b'} = (b_2, ..., b_p) = (\boldsymbol{P'\Sigma^{-\frac{1}{2}}\mu})'$
As we discussed before the U_j's are mutually independent standard normal variables.

From now on, for a symmetric matrix $\boldsymbol{A}$, the density of $Y = Q(\boldsymbol{X}) = \boldsymbol{X'AX}$ will be denoted by $f_p(\lambda; b; y)$, and the distribution function $P\{Y \leq y\}$, will be denoted by $F_p(\lambda; b; y)$. The Laplace transform of the densities will be denoted by $L(\lambda; b; s)$ where

$$L(\lambda; b; s) = \int_0^\infty \exp(-sy) f_p(\lambda; b; s) dy$$

and

$$F_p(\lambda; b; y) = \int_0^y f_p(\lambda; b; t) dt$$

assuming the integrals exist.

2.1 Series Expansion

There are several methods used to derive expansions of the distribution of central and non-central quadratic forms in normal random variables, see Shah and Khatri (1961), Pachares (1955), Ruben (1960),Ruben (1962). The representations in Maclaurin (power) series, Laguerre polynomials, and χ^2 densities are discussed in the following. We will briefly outline some of the general methods here.

The series expansion sought for $f_p(\lambda; b; y)$ is of the form

$$f_p(\lambda; b; y) = \sum_{k=0}^{\infty} c_k h_k(y)$$

where $\{h_k\}$ is a sequence of known function. Denoting the Laplace transform of the function $h_k(y)$ in $[0, \infty]$ by $\hat{h}_k(s)$ we have

$$\hat{h}_k(s) = \xi(s)\eta^k(s)$$

where $\xi(s)$ in a non-vanishing analytic function for $Re(s) > \beta$, $\eta(s)$ is analytic for $Re(s) > \beta$, and it has an inverse function ζ. Hence, if

$$\sum_{k=0}^{\infty} c_k h_k(y) \leq \sum_{k=0}^{\infty} |c_k||h_k^{(y)}| \leq \beta e^{ay}, y \in [0, \infty]$$

holds almost everywhere, then

$$\int_0^{\infty} e^{-sy} \beta e^{ay} dy = \beta \int_0^{\infty} e^{-(s-a)y} dy < \infty.$$

Therefore, by the use of the Lebesgue's dominated convergence theorem we have the following lemma.

Lemma 3. Consider a sequence of measurable complex valued function on $[0, \infty]$ of $h_0, h_1, ...$ and let $c_0, c_1, ...$ be a sequence of complex numbers such that

$$\sum_{k=0}^{\infty} |c_k||h_k(y)| \leq \beta e^{ay}, y \in [0, \infty]$$

where β and a are real constant, then when

$$f_p(\lambda; b; y) = \sum_{k=0}^{\infty} c_k h_k(y)$$

we have

$$L(\lambda; b; s) = \sum_{k=0}^{\infty} c_k \hat{h}_k(s), \quad Re(s) > a$$

Therefor, one can say when the series in the equation (3) is dominated by an exponential function as in (3), it is possible to take the Laplace transform for each term of the series to get the Laplace transform of the series representation for $f_p(\lambda; b; y)$, see Mathai and Provost (1992b).

2.1.1 Power Series Expansion: Nonsingular Normal Case

Let

$$M(\theta) = \sum_{k=0}^{\infty} c_k \theta^k, \quad M(0) = c_0 \tag{2.1}$$

be a power series in θ which is uniformly convergent is a region. Also consider $M(\theta) > 0$ for all θ in that region. Now let

$$\ln M(\theta) = d_0 + \sum_{k=1}^{\infty} d_k \frac{\theta^k}{k}. \tag{2.2}$$

If $M(\theta)$ is differentiable then

$$M(\theta) \left\{ \frac{d}{d\theta} \ln M(\theta) \right\} = M^{'}(\theta) = \sum_{k=1}^{\infty} k c_k \theta^{k-1}$$

and hence

$$\sum_{k=1}^{\infty} k c_k \theta^{k-1} \equiv \left(\sum_{k=0}^{\infty} c_k \theta^k \right).$$

Now by comparing the coefficients of θ^{k-1} one can have the following lemma, Mathai and Provost (1992b).

Lemma 4. The coefficients c_k's in (2.1) are given by

$$c_k = \frac{1}{k}\sum_{r=0}^{k-1} d_{k-r}c_r \ , \ \geq 1$$

where c_k and d_k are defined in (2.1) and (2.2).

Now, we are ready to look at the representation of the Laplace transform of $f_p(\lambda; b; y)$ when Y is a quadratic form in normal random variables. Then,

$$L(\lambda; b; s) = \exp\left(-\frac{1}{2}\sum_{j=1}^{p} b_j^2\right)\exp\left\{\frac{1}{2}\sum_{j=1}^{p} b_j^2(1+2s\lambda_i)^{-1}\right\}\prod_{j=1}^{p}(1+2s\lambda_j)^{-\frac{1}{2}} \qquad (2.3)$$

where $\boldsymbol{b} = \boldsymbol{P}'\boldsymbol{\Sigma}^{-\frac{1}{2}}\boldsymbol{\mu}$, $\boldsymbol{P}'\boldsymbol{\Sigma}^{\frac{1}{2}}\boldsymbol{A}\boldsymbol{\Sigma}^{\frac{1}{2}}\boldsymbol{P} = diag(\lambda_1, ..., \lambda_p), \boldsymbol{P}\boldsymbol{P}' = \boldsymbol{I}$. Then for $|2s\lambda_j| < 1, j = 1, ...p$, $L(\lambda; b; s)$ exists. Assuming that $\boldsymbol{A} > 0$ which means that each λ_j is positive, one can write

$$L(\lambda; b; s) = s^{-\frac{p}{2}}M(\theta) \ , \ \theta = \frac{1}{s}$$

and hence

$$\begin{aligned} M(0) &= c_o = \exp\left(-\frac{1}{2}\sum_{j=1}^{p} b_j^2\right)\prod_{j=1}^{p}(2\lambda_j)^{-\frac{1}{2}} && (2.4)\\ &= \prod_{j=1}^{p}(2\lambda_i)^{-\frac{1}{2}} \quad \text{when} \quad \boldsymbol{\mu} = 0. && (2.5)\end{aligned}$$

Therefore, for $|s| > \frac{1}{2\lambda_j}, j = 1, ..., p$, $\ln M(\theta)$ admits the following expansion

$$\begin{aligned} \ln M(\theta) = &- \frac{1}{2}\sum_{j=1}^{p} b_j^2 + \frac{1}{2}\sum_{j=1}^{p} b_j^2\left\{\left(\frac{\theta}{2\lambda_j}\right) - \left(\frac{\theta}{2\lambda_j}\right)^2 + \cdots\right\}\\ &+ \ln\prod_{j=1}^{p}(2\lambda_j)^{-\frac{1}{2}} - \frac{1}{2}\sum_{j=1}^{p}\left\{\left(\frac{\theta}{2\lambda_j}\right) - \left(\frac{\theta}{2\lambda_j}\right)^2 + \cdots\right\}\\ &= d_o + \sum_{k=1}^{\infty}(-1)^k d_k\theta^k/k \end{aligned}$$

where

$$\begin{aligned} d_k &= \frac{1}{2}\sum_{j=1}^{p}(1-kb_j^2)(2\lambda_j)^{-k} \ , \ k \geq 1\\ &= \frac{1}{2}\sum_{j=1}^{p}(2\lambda_j)^{-k} \ , \ \text{when } \boldsymbol{\mu} = 0. \end{aligned}$$

Now we can have the following equation as follows

$$L(\lambda; b; s) = \sum_{k=0}^{\infty} c_k (-1)^k s^{-(\frac{p}{2}+k)},$$

where c_k is available from lemma (4). It is noteworthy that $s^{-(p/2+k)}$ is the Laplace transform of $y^{p/2+k-1}/\Gamma(P/2+k)$. Hence, we can have the following theorem.

Theorem 2.1.1. *Consider the quadratic form $Q(\boldsymbol{X}) = \boldsymbol{X}'\boldsymbol{A}\boldsymbol{X}$ for a symmetric positive definite matrix $\boldsymbol{A}$ and its density function $f_p(\lambda; b; y)$ with the distribution function $F_p(\lambda; b; y)$, then*

$$f_p(\lambda; b; y) = \sum_{k=0}^{\infty} (-1)^k c_k \frac{y^{\frac{p}{2}+k-1}}{\Gamma(\frac{p}{2}+k)}\ ,\ 0 < y < \infty$$

and

$$F_p(\lambda; b; y) = \sum_{k=0}^{\infty} (-1)^k c_k \frac{y^{\frac{p}{2}+k}}{\Gamma(\frac{p}{2}+k+1)}\ ,\ 0 < y < \infty$$

where the coefficients of c_0 and c_k are given by (2.4) and (4).

Example 5. *Obtain a power series expansion for the density $Q(\boldsymbol{X}) = 8X_1^2 - 8X_1X_2 + 3X_2^2$ when $\boldsymbol{X} \sim N_2(\boldsymbol{\mu}, \boldsymbol{\Sigma}), \boldsymbol{\mu}' = (-1, 1), \boldsymbol{\Sigma} = \begin{pmatrix} 1 & 1 \\ 1 & 2 \end{pmatrix}$.*

SOLUTION. Write $\boldsymbol{\Sigma} = \boldsymbol{B}\boldsymbol{B}'$ where, for example take $\boldsymbol{B} = \begin{pmatrix} 1 & 0 \\ 1 & 1 \end{pmatrix}$. Write $Q(\boldsymbol{X}) = \boldsymbol{X}'\boldsymbol{A}\boldsymbol{X}$, $\boldsymbol{A} = \boldsymbol{A}'$ then the eigenvalues of $\boldsymbol{B}'\boldsymbol{A}\boldsymbol{B}$ are computed in Example 3.1a.1. That is, $\lambda_1 = 2, \lambda_2 = 4, \boldsymbol{P}'\boldsymbol{B}'\boldsymbol{A}\boldsymbol{B}\boldsymbol{P} = \mathrm{diag}(2, 4)$,

$$\boldsymbol{A} = \begin{pmatrix} 8 & -4 \\ -4 & 3 \end{pmatrix}, \boldsymbol{P} = \begin{pmatrix} \frac{1}{\sqrt{2}} & \frac{1}{\sqrt{2}} \\ \frac{1}{\sqrt{2}} & -\frac{1}{\sqrt{2}} \end{pmatrix}$$

$$\boldsymbol{b} = \boldsymbol{P}'\boldsymbol{B}^{-1}\boldsymbol{\mu} = \begin{pmatrix} \frac{1}{\sqrt{2}} \\ -\frac{3}{\sqrt{2}} \end{pmatrix} = \begin{pmatrix} b_1 \\ b_2 \end{pmatrix}; p = 2, b_1^2 = \frac{1}{2}, b_2^2 = \frac{9}{2}.$$

The Laplace transform of the density of $Q(\boldsymbol{X})$, denoted by $L_Q(s)$, is given by

$$L_Q(s) = (1+4s)^{-\frac{1}{2}}(1+8s)^{-\frac{1}{2}} \exp\left\{-\frac{5}{2} + \frac{1}{4}(1+4s)^{-1} + \frac{9}{4}(1+8s)^{-1}\right\}.$$

Using the same notations as in equations (4.2b.5) and (4.2b.13) we have

$$d_k = \frac{(-1)^k}{2}\left\{\frac{1}{4^k}\left(1-\frac{k}{2}\right)+\frac{1}{8^k}\left(1-\frac{9k}{2}\right)\right\}$$

$$d_1 = \frac{5}{32}, d_2 = -\frac{1}{16}, d_3 = \frac{33}{2048}, \cdots$$

Therefore,

$$\begin{aligned} c_0 &= \frac{e^{-\frac{5}{2}}}{4\sqrt{2}}, c_1 = d_1 c_o = \frac{5}{32}c_o, c_2 = \frac{1}{2}(d_2 c_o + d_1 c_1) \\ &= -\frac{39}{1024}c_o, c_3 = \frac{1}{3}(d_3 c_o + d_2 c_1 + d_1 c_2) = \frac{1}{6144}c_o. \end{aligned}$$

Then if $Y = Q(\boldsymbol{X})$ and if the density of Y is denoted by $f(y)$ then

$$f(y) = c_o\left(1+\frac{5}{32}y-\frac{39}{2048}y^2+\frac{y^3}{36864}-\cdots\right), 0 < y < \infty$$

and $f(y) = 0$ elsewhere. Note that since $p = 2, y^{\frac{p}{2}+k-1}/\Gamma(\frac{p}{2}+k)$ becomes $y^k/k!$. Mathai and Provost (1992b)

2.1.2 Laguerre Series Expansion: Nonsingular Normal case

The distribution function $F_p(\lambda; b; y)$ and density function $f_p(\lambda; b; y)$ can be derived in series expansion in terms of generalized Laguerre polynomials $L_k^{(\alpha)}(x)$, defined by:

$$L_k^{(a)}(x) = \frac{1}{k!}e^x x^{-a}\left[\frac{d^k}{dx^k}(e^{-x}x^{k+\alpha})\right], \alpha > -1, k = 0, 1, \cdots . \tag{2.6}$$

It is interesting to know that it can also be represented as the coefficient of t^k in the power series expansion of $(1-t)^{-\alpha-1}\exp\{-\frac{xt}{1-t}\}$ for $|t| < 1$ that is

$$\sum_{k=0}^{\infty} L_k^{(\alpha)}(x)t^k = (1-t)^{-(\alpha+1)}e^{-xt/(1-t)}, |t| < 1.$$

Therefore, one can have explicit forms of the above equations by noting the following

$$\begin{aligned}(1-t)^{-(\alpha+1)e^{-\frac{xt}{1-t}}} &= \sum_{r=0}^{\infty}\frac{(-x)^r}{r!}t^r(1-t)^{-(\alpha+1+r)} \\ &= \sum_{r=0}^{\infty}\sum_{m=0}^{\infty}\frac{(-x)^r t^{m+r}}{r!m!}(\alpha+1+r)_m, |t|<1\end{aligned}$$

where $(a)_m$ denotes $a(a+1)...(a+m-1)$ and $(a)_0 = 1$. On expanding $(a+1+r)_m$, we get

$$\begin{aligned}(1-t)^{-(\alpha+1)e^{-\frac{xt}{1-t}}} &= \sum_{r=0}^{\infty}\sum_{m=0}^{\infty}\frac{(-x)^r t^{m+r}}{r!m!}(\alpha+1+r)(\alpha+2+r)\cdots(\alpha+r+m) \\ &= \sum_{n=0}^{\infty}t^n\left(\sum_{r=0}^{n}\frac{(-x)^r(\alpha+r+1)(\alpha+r+2)\cdots(\alpha+r+m)}{r!(n-r)!}\right).\end{aligned}$$

Hence from the above relation one can have

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

On differentiation (2.6) we have:

$$\tfrac{d^k}{dx^k}(e^{-x}x^{k+\alpha}) = \textstyle\sum_{r=0}^{k}\left(\begin{array}{c}k\\ r\end{array}\right)\left(\tfrac{d^r}{dx^r}e^{-x}\right)\left(\tfrac{d^{k-r}}{dx^{k-r}}x^{k+\alpha}\right)$$

$$= \textstyle\sum_{r=0}^{k}\left(\begin{array}{c}k\\ r\end{array}\right)(-1)^r e^{-x}(k+\alpha)(k+\alpha-1)\cdots(\alpha+r+1)x^{\alpha+r}.$$

and therefore

$$\tfrac{1}{k!}e^x x^{-\alpha}\tfrac{d^k}{dx^k}(e^{-x}x^{k+\alpha})$$

$$= \textstyle\sum_{r=0}^{k}\tfrac{(-x)^r}{r!(k-r)!}(\alpha+k)(\alpha+k-1)\cdots(\alpha+r+1)$$

Summarizing, we have the following representation:

Lemma 5.

$$\begin{aligned}L_k^{(a)}(x) &= \sum_{r=0}^{k}\frac{(-x)^r}{r!(k-r)!}(\alpha+r+1)(\alpha+r+2)\cdots(\alpha+k). \\ &= \frac{(\alpha+1)}{k!}{}_1F_1(-k;\alpha+1;x) \\ &= \frac{(-x)^k}{k!}{}_2F_0(-k,-\alpha-k;-\frac{1}{x}).\end{aligned}$$

There is one more result to be proved for obtaining $f_p(\lambda; b; y)$ and $F_p(\lambda; b; y)$ in terms of Laguerre polynomials. Note that if Laguerre transform of $g(y)$ is

Lemma 6.

$$\int_0^\infty e^{-sy} g(y)dy = (2s\beta)^k (1+2s\beta)^{-\frac{n}{2}+k} \tag{2.7}$$

where, $\beta > 0$ and $|2s\beta| < 1$, then $g(y)$ is given by:

$$g(y) = \frac{k!}{2\beta\Gamma(\frac{n}{2}+k)} \left(\frac{y}{2\beta}\right)^{\frac{n}{2}-1} e^{-y}/(2\beta) L_k^{(\frac{n}{2}-1)}(y/(2\beta)), 0 < y < \infty. \tag{2.8}$$

From the above equations one can see that the Laplace transform of $g(y)$ in (2.8) and (2.7) can be obtained from each other. Then one can rewrite the Laplace transform of $f_p(\lambda; b; y)$ through the use of the following relation.

$$\begin{aligned}
\prod_{j=1}^{p}(1+2s\lambda_j)^{-\frac{1}{2}} &= \prod_{j=1}^{p}(1+2s\beta-2s\beta+2s\lambda_j)^{-\frac{1}{2}} \\
&= (1+2s\beta)^{-\frac{p}{2}} \prod_{j=1}^{p}(1-\gamma_j\theta)^{-\frac{1}{2}}
\end{aligned}$$

$\lambda_j > 0$, $j = 1, ..., p$, and β a positive constant.

$$\gamma_i = 1 - \frac{\lambda_j}{\beta}, \theta = \frac{2s\beta}{1+2s\beta}$$

and hence

$$\begin{aligned}
(1+2s\lambda_j)^{-1} &= (1+2s\beta)^{-1}(1-\gamma_j\theta)^{-1} \\
&= (1-\theta)(1-\gamma_j\theta)^{-1}
\end{aligned}$$

and

$$L(\lambda; b; s) = (1+2s\beta)^{-\frac{p}{2}} M(\theta)$$

where

$$\ln M(\theta) = -\frac{1}{2}\sum_{j=1}^{p} b_j^2 \frac{1}{2}(1-\theta)\sum_{j=1}^{p} b_j^2 (1-\gamma_j\theta)^{-1} - \frac{1}{2}\sum_{j=1}^{p}\ln(1-\gamma_j\theta).$$

hence, under the condition that $|\gamma_j \theta| < 1$ for all j and $|\theta| < (1, 1/(max_j|\gamma_j|))$ we have

$$\ln M(\theta) = d_0 + \sum_{k=1}^{\infty} d_k \frac{\theta^k}{k}$$

where

$$d_k = \frac{1}{2}\left\{ -\frac{k}{\beta}\sum_{j=1}^{p} \lambda_j b_j^2 \left(1 - \frac{\lambda_j}{\beta}\right)^{k-1} + \sum_{j=1}^{p}\left(1 - \frac{\lambda_j}{\beta}\right)^{k} \right\}, k \geq 1.$$

Therefore, noting that

$$(1 + 2s\beta)^{-p/2}\theta^k = (2s\beta)^k (1 + 2s\beta)^{-\frac{p}{2}+k} \tag{2.9}$$

The final representation is given by:

Theorem 2.1.2. *For a symmetric matrix $\boldsymbol{A}$, consider the quadratic form $Y = \boldsymbol{X'AX}$, where $\boldsymbol{X} \sim N(\boldsymbol{\mu}, \boldsymbol{\Sigma})$, and $\boldsymbol{\Sigma} > 0$. Then, we have*

$$\begin{aligned} f_p(\lambda; b; y) &= \sum_{k=0}^{\infty} c_k \frac{k!}{2\beta\Gamma(\frac{p}{2}+k)} \left(\frac{y}{2\beta}\right)^{\frac{p}{2}-1} e^{-y/(2\beta)} L_k^{(\frac{p}{2}-1)}\left(\frac{y}{2\beta}\right), c_0 = 1 \quad (2.10) \\ &= \frac{(y/(2\beta))^{\frac{p}{2}-1} e^{-y/(2\beta)}}{2\beta\Gamma(p/2)} + \sum_{k=1}^{\infty} c_k \frac{k!}{2\beta\Gamma(\frac{p}{2}+k)} (y/(2\beta))^{\frac{p}{2}-1} \quad (2.11) \\ & \quad e^{-y/(2\beta)} L_k^{(\frac{p}{2}-1)}\left(\frac{y}{2\beta}\right), 0 < y < \infty. \end{aligned}$$

The expression for the cdf is $F_p(\lambda; b; y) = \int_0^y f_p(\lambda; b; t)dt$ and it is obtained by integrating the terms of (2.10), and using Rodrigues formula for Laguerre polynomials,

$$\int_0^y \frac{1}{2\beta}\left(\frac{x}{2\beta}\right)^{\frac{p}{2}-1} e^{-x/(2\beta)}(x/(2\beta))dx \tag{2.12}$$

$$= \frac{1}{k} e^{\frac{y}{2\beta}} \left(\frac{y}{2\beta}\right)^{\frac{p}{2}} L_{k-1}^{(p/2)}(y/(2\beta)), k \geq 1. \tag{2.13}$$

The final expression for the cdf is

$$F_p(\lambda; b; y) = F(p; \frac{y}{\beta}) + \sum_{k=1}^{\infty} c_k \frac{(k-1)!}{\Gamma(\frac{p}{2}+k)} \left(\frac{y}{2\beta}\right)^{\frac{p}{2}} e^{-y/(2\beta)} L_k^{(p/2)}(y/(2\beta)),$$

where

$$F(p; \frac{y}{\beta}) = \int_0^y \frac{(x/(2\beta))^{\frac{p}{2}-1}}{(2\beta)\Gamma(\frac{p}{2})} e^{-x/(2\beta)} dx, 0 < y < \infty$$

2.1.3 Expansion in Central Chi-square Densities: Nonsingular Normal Case

It is of the interest to have a series representation in terms of chi-square densities for the non-singular normal case. Consider

$$\begin{aligned}\prod_{j=1}^{p}(1+2s\lambda_j)^{-\frac{1}{2}} &= \prod_{j=1}^{p}\left(1+2s\beta\frac{\lambda_j}{\beta}+\frac{\lambda_j}{\beta}-\frac{\lambda_j}{\beta}\right)^{-\frac{1}{2}}\\ &= (1+2s\beta)^{-p/2}\left(\prod_{j=1}^{p}\left(\frac{\lambda_j}{\beta}\right)^{-\frac{1}{2}}\right)\prod_{j=1}^{p}(1-\delta_j\theta)^{-\frac{1}{2}}\end{aligned}$$

where β is an arbitrary positive constant, and

$$\delta_j = 1-\beta/\lambda_j, \theta = (1+2s\beta)^{-1}, \lambda_j > 0, j = 1, \cdots, p.$$

Therefore, one can rewrite (2.3) as

$$L(\lambda; b; s) = (1+2s\beta)^{-\frac{p}{2}} M(\theta)$$

where

$$\begin{aligned}\ln M(\theta) = &- \frac{1}{2}\sum_{j=1}^{p} b_j^2 + \frac{1}{2}\sum_{j=1}^{p} b_j^2\theta(\beta/\lambda_j)(1-\delta_j\theta)^{-1}\\ &- \frac{1}{2}\sum_{j=1}^{p}\ln(1-\delta_j\theta) + \ln\left(\prod_{j=1}^{p}(\beta/\lambda_j)^{\frac{1}{2}}\right).\end{aligned}$$

After expanding it into a power series in θ for

$$|\delta_j\theta| < 1 \rightarrow |\theta| < \frac{1}{\max_j|\delta_j|}.$$

just like the previous section we can write

$$\ln M(\theta) = d_0 + \sum_{k=1}^{\infty} d_k\frac{\theta^k}{k} \text{ and } M(\theta) = \sum_{k=0}^{\infty} c_k\theta^k,$$

where $\theta^{p/2}$ is the Laplace transform and

$$\frac{1}{\beta}f(p; \frac{y}{\beta}) = \frac{y^{p/2-1}e^{-y/(2\beta)}}{(2\beta)^{p/2}\Gamma(p/2)}, 0 < y < \infty \tag{2.14}$$

After going through the same steps as in previous sections one can have the distribution and density function as in the following

Theorem 2.1.3. *For $Y = \boldsymbol{X}'\boldsymbol{A}\boldsymbol{X}, \boldsymbol{A} = \boldsymbol{A}' > 0$, $\boldsymbol{X} \sim N_p(\boldsymbol{\mu}, \boldsymbol{\Sigma}), \boldsymbol{\Sigma} > 0$, the density and distribution functions are given by*

$$f_p(\lambda; b; y) = \sum_{k=0}^{\infty} c_k \beta^{-1} f(p + 2k; y/\beta), 0 < y < \infty$$

and

$$F_p(\lambda; b; y) = \sum_{k=0}^{\infty} c_k F(p + 2k; y/\beta), 0 < y < \infty$$

where $F(\nu; t) = \int_0^t f(\nu; y)dy$, $f(\nu; y)$ is given in (2.14), , for more details see Mathai and Provost (1992b).

To compute coefficients c_k one can use the following formulas

$$c_0 = \exp\left(-\frac{1}{2}\sum_{j=1}^{p} b_j^2\right)\prod_{j=1}^{p}(\beta/\lambda_j)^{\frac{1}{2}}$$

$$c_k = (2k)^{-1}\sum_{r=0}^{k-1} d_{k-r}c_r, k \geq 1$$

with

$$d_k = \sum_{j=1}^{p}(1 - \beta/\lambda_j)^k + k\beta\sum_{j=1}^{p}(b_i^2/\lambda_j)(1 - \beta/\lambda_j)^{k-10}.$$

This expansion is derived by Ruben (1962). There are other expansions like expansion in non-central χ^2 where they are not convenient for computational purposes. It is worthy of note that it has been shown that from the computational perspective Laguerre series representation is the most effective and convenient method.

Example 6. *Obtain a representation of the densities of $Q(\boldsymbol{X}) = 8X_1^2 - 8X_1X_2 + 3X_2^2$ in terms of chi-square densities , where $\boldsymbol{X} \sim N_2(\boldsymbol{\mu}, \boldsymbol{\Sigma})$, $\boldsymbol{\mu}' = (-1, 1)$, $\boldsymbol{\Sigma} = \begin{pmatrix} 1 & 1 \\ 1 & 2 \end{pmatrix}$.*

The Laplace transform of Q is denoted by $L_Q(s)$ and is as follows

$$L_Q(s) = (1 + 4s)^{-\frac{1}{2}}(1 + 8s)^{-\frac{1}{2}}\exp\left\{-\frac{5}{2} + \frac{1}{4}(1 + 4s)^{-1} + \frac{9}{4}(1 + 8s)^{-1}\right\}$$

Then one can write

$$\prod_{j=1}^{p} = (1+2s\lambda_j)^{\frac{1}{2}} = (1+2s\beta)^{p/2}\left(\prod_{j=1}^{p}\left(\frac{\lambda_j}{\beta}\right)^{-\frac{1}{2}}\right)\prod_{j=1}^{p}(1-\delta_j\theta)^{-\frac{1}{2}}$$

where $\beta > 0$ is an arbitrary parameter, $\delta_j = 1 - \frac{\beta}{\lambda_j}$, $\theta = (1+2s\beta)^{-1}$ and

$$\begin{aligned}(1+2s\lambda_j)^{-1} &= (1+2\beta s)^{-1}\left(\frac{\lambda_j}{\beta}\right)^{-1}(1-\delta_j\theta)^{-1} \\ &= \theta(\beta/\lambda_j)(1-\delta_j\theta)^{-1}.\end{aligned}$$

In our example $\lambda_1 = 2$, $\lambda_2 = 4$ and $p = 2$. Taking $\beta = 2$, $\delta_1 = 0$, $\delta_2 = \frac{1}{2}$, and $\theta = (1+4s)^{-1}$ we can write

$$L_Q(s) = (1+4s)^{-1}M_1(\theta)$$

$$\ln M_1(\theta) = \ln c_0 - \frac{1}{2}\ln\left(1-\frac{\theta}{2}\right) + \frac{\theta}{4} + \frac{9\theta}{8}\left(1-\frac{\theta}{2}\right)^{-1}$$

where

$$c_0 = \frac{e^{-\frac{5}{2}}}{\sqrt{2}}$$

Now we examine the case where $Y = Q(\boldsymbol{X})$. We have

$$d_k = \frac{1}{2^{k+1}}\left(1+\frac{9k}{2}\right), k \geq 2 \text{ and } d_1 = \frac{13}{8}.$$

Thuse,

$$d_2 = \frac{5}{4}, d_3 = \frac{29}{32}, c_1 = d_1, c_0 = \frac{13}{8}c_0$$

$$c_2 = \frac{1}{2}(d_2c_0 + d_1c_1) = \frac{249}{128}c_0$$

$$c_3 = \frac{1}{3}(d_3c_0 + d_2c_1 + d_1c_2) = \frac{6245}{3072}, c_0 = \frac{1}{\sqrt{2}}\exp\{-5/2\}.$$

Therefore

$$f(y) = \sum_{k=0}^{\infty} c_k \frac{y^k e^{-y/4}}{4^{k+1}k!}, 0 < y < \infty$$

2.1.4 Linear Combination of Chi-Squares

The distribution of linear combination of independent central chi-square random variables can be obtained as an infinite series of gamma distributions, see Moschopoulos and Canada (1984). This can be done using the straightforward manner of inverting the moment generating function. Let $\boldsymbol{c} = (c_1, ..., c_p)'$ be a vector of real and nonzero constants. Also, let $\boldsymbol{n} = (n_1, ..., n_p)'$ be a vector of positive integers. Consider the random variable

$$Q(c, n) = \sum_{i=1}^{p} c_i \chi^2(n_i),$$

where $\chi^2(n_i)$,'s are independent random variables with n_i degrees of freedom. Concerning the distribution of Q, there are various representations for its distribution. Although the exact series representation of the distribution of Q are found in the literature, the use of these representations are computationally unfeasible. There have been several proposed methods as remedies to this problem. Oman and Zacks (1981) proposed a method and showed that other methods available in the literature are less efficient than their method in the sense that they need more lengthy computations while gaining less accuracy. Moschopoulos and Canada (1984) proposed another method which is fairly accurate and computationally efficient. The method is basically based on inverting the moment generating function $M(t)$ of Q. It is well known that

$$M(t) = \prod_{i=1}^{p} (1 - 2c_i t)^{-m_i}, \quad m_i = n_i/2.$$

Then one can write

$$1 - 2c_i t = (1 - 2c_i t)(c_i/c_1)[1 - (1 - c_1/c_i)/(1 - 2c_1 t)], \quad i = 2, ..., p$$

When $|(1-c_1/c_i)/(1-2c_1 t)| < 1$, Moschopoulos and Canada (1984) obtained an asymptotic expansion as follows

$$(1 - 2c_i t)^{-m_i} = (c_1/c_i)^{m_i} \sum_{r=0}^{\infty} \frac{(m_i)_r}{r!} (1 - c_1/c_i)^r (1 - 2c_1 t)^{-(r+m_i)},$$

where $(m)_0 = 1, (m)_r = m(m+1)...(m+r-1)$, and $t < min(1/2c_i)$ is a sufficient for this expansion.

Writing $M(t)$ as an infinite series of $(1-2c_i t)^{-1}$ one can have

$$M' = \big(\sum_{i=2}^{p} b_i\big) \sum_{j=0}^{\infty} a_j (1-2c_1 t)^{-(s+j)}, \quad s = \sum_{i=1}^{p} m_i \tag{2.15}$$

where $b_i = (c_1/c_i)^{m_i}$ and a_j's satisfy the relation

$$\prod_{i=2}^{p} [\sum_{r=0}^{\infty} A(c_i, r) x^{-r}] = \sum_{j=0}^{\infty} a_j x^{-j}$$

where $A(c_i, r) = (m_i)_r (1-c_1/c_i)^r / r!$. Hence, computing a_j's recursively yields the following:

$$a_j = A_j^{(p)}, A_j^{(i)} = \sum_{k=0}^{j} A_k^{(i-1)} A(c_i, j-k)$$

for $i = 3, ..., p$, $j = 0, 1, 2, ...$, and for $r = 0, 1, 2, ...$.

By taking into consideration that $(1-2c_i t)^{-(s+j)}$ is the moment generating function of a gamma density $g_j(y) = y^{s+j-1} e^{-y/2c_1} / (2c_1)^{s+j} \Gamma(s+j)$ and inverting (2.15) term by term one can obtain the distribution function of Q as follows

$$F(w) = \big(\prod_{i=2}^{p} b_i\big) \sum_{j=0}^{\infty} a_j \int_0^w g_j(y) dy$$

where a_j's and b_j's are computed as before.

2.2 Approximations and Limiting Distributions

2.2.1 Approximations

There has been always a need for a table of percentiles for noncentral case. But regarding the existing tables not only an interpolation is frequently required, but there has been need for extensive tabulation for larger p (dimension of the random vector) which has been prohibited because of the number of parameters involved. Regarding these needs, various

methods have been developed. However, methods for computing the exact distribution of the quadratic forms in normal random variables require extensive computations. It is also one of the important obstacles for the numerical integration, even though they are accurate enough in solving the general problems. To alleviate the situation several approximations have been proposed. However, some are restricted to some the special cases of quadratic forms and certain applications. From all of them we only present normal approximation since we will use it in the next chapter.

A Normalizing Transformation

This normalizing transformation is used to make a statistic converge to a normal distribution. The procedure it takes is to make the skewness of the final normal distribution as small as possible, Mathai and Provost (1992b). Let

$$Z = \left(\frac{Q}{\theta_1}\right)^h = (1 + \frac{Q - \theta_1}{\theta_1})^h = (1 + \frac{T}{\theta_1})^h \tag{2.16}$$

where $T = Q - \theta_1$. The moments of Z using a binomial expansion are,

$$\mu_1^{'}(h) = \sum_{r=0}^{\infty} \left(\frac{h(h-1)\cdots(h-r+1)}{r!}\right) \theta_1^{-r} E(T^r)$$

For a quadratic form we can obtain the cumulants and are as follow

$$K_r = 2^{r-1}(r-1)! \sum_{j=1}^{n} \lambda_j^r (1 + r b_j^2), r = 1, 2, \cdots$$

$$K_r = 2^{r-1}(r-1)! \theta_r$$

where

$$\theta_r = \sum_{j=1}^{n} \lambda_j^r (1 + r b_j^2).$$

Using the well known relation between the means and the cumulants one can obtain

$$
\begin{aligned}
\mu_1^{'} &= K_1 = \theta_1 \\
\mu_2 &= K_2 = 2\theta_2 \\
\mu_3 &= K_3 = 8\theta_3 \\
\mu_4 &= K_4 + 3\mu_2^2 = 12(4\theta_4 + \theta_2^2).
\end{aligned}
$$

an hence it is straightforward to find first moments, second central moments and third central moment as follows

$$
\begin{aligned}
\mu_1^{'} \simeq 1 &+ \frac{1}{\theta_1}(h(h-1)\varphi_2) \\
&+ \frac{h(h-1)(h-2)}{\theta_1^2}\left\{\left(\frac{4}{3}\right)\varphi_3 + \left(\frac{h-3}{2}\right)\varphi_2^2\right\} \\
&+ \frac{h(h-1)(h-2)(h-3)}{\theta_1^3} \\
&\left\{2\varphi_4 + \frac{4(h-4)}{3}\varphi_2\varphi_3 + \frac{(h-4)(h-5)}{6}\varphi_2^3\right\}
\end{aligned}
$$

and

$$
\begin{aligned}
\mu_2(h) \simeq \quad & \frac{1}{\theta_1}[2h^2\varphi_2] + 2\frac{h^2(h-1)}{\theta_1^2}\{4\varphi_3 + (3h-5)\varphi_2^2\} + 4\frac{h^2(h-1)}{\theta_1^3} \\
& \left\{(7h-11)\varphi_4 + \frac{4(h-2)(7h-12)}{3}\varphi_2\varphi_3 + \frac{(h-2)(14h^2-60h+64)}{6}\varphi_2^3\right\}
\end{aligned}
$$

$$
\begin{aligned}
\mu_3(h) &\simeq \frac{4h^3}{\theta_1^2}(2\varphi_3 + 3(h-1)\varphi_2^2) \\
&+ \frac{h^3(h-1)}{\theta_1^3}(72\varphi_4 + 24(7h-10)\varphi_2\varphi_3 + 4(17h^2-55h+44)\varphi_2^3)
\end{aligned}
$$

Thus the skewness of the quadratic form would be obtained.

$$
\begin{aligned}
\tau_1(h) &\simeq 4(2\varphi_3 + 3(h-1)\varphi_2^2)/[\theta_1^{\frac{1}{2}}(2\varphi_2)^{\frac{3}{2}}] + (h-1)\{72\varphi_4 \\
&+ 24(7h-10)\varphi_2\varphi_3 + 4(17h^2 - 55h + 44\varphi_2^3)\}/[\theta_1^{\frac{3}{2}}(2\varphi_2)^{\frac{3}{2}}]
\end{aligned}
$$

Finally h is chosen to make the leading term of the skewness zero, and hence the distribution of the statistic is closer to a symmetric normal distribution. The solution for equation

$$2\varphi_3 + 3(h-1)\varphi_2^2 = 0$$

or

$$\frac{2\theta_3}{\theta_1} + 3(h-1)(\theta_2/\theta_1)^2 = 0$$

makes the leading term of the skewness zero. Hence,

$$h_0 = 1 - 2\theta_1\theta_3/3\theta_2^2$$

Here, the distribution of $Z = \left(\frac{Q}{\theta_1}\right)^{h_0}$ is a approximated by a normal distribution. The mean and the variance are obtained through the formulas below.

$$\mu_1^{'}(h_0) = 1 + \theta_2 h_0(h_0 - 1)/\theta_1^2 \text{ and } \mu_2(h_0) = 2\theta_2 h_0^2/\theta_1^2$$

2.2.2 Limiting Distribution

Beside the approximations to the distribution function of Q, we can also think of its limiting distribution. A brief overview of the asymptotic methods is provided below.

The proof for the convergence to normality of a linear combination of chi-square variables, $Q = \sum_{i=1}^{p} \lambda_i X_i^2$, where X_i^2 is a chi-square variable with n_i degrees of freedom is provided in the following.

Let $Z = Q/\sqrt{n}$ where $n = n_1 + n_2 + \ldots + n_p$. The moment generating function of Z, denoted as $M_Z(t)$, is given by

$$M_Z(t) = M_Q(t/\sqrt{n}) = \prod_{i=1}^{p}(1 - 2\lambda_i t/\sqrt{n})^{-n_i/2} \tag{2.17}$$

It follows that

$$\log \quad M_Q \quad (t/\sqrt{n}) = -\sum_{i=1}^{p}(n_i/2)\log(1-2\lambda_i t/\sqrt{n}) \tag{2.18}$$

$$= \quad \sum_{i=1}^{p}(n_i/2)(2\lambda_i t/\sqrt{n}) + (2\lambda_i t/\sqrt{n})^2/2 + (2\lambda_i t/\sqrt{n})^3/3 + \cdots \tag{2.19}$$

$$\text{for } |2\lambda_i t| < 1, i = 1, \cdots, n. \tag{2.20}$$

Suppose that n is large so that $\sum_{i=1}^{p}(n_i/2)(2\lambda_i/\sqrt{n})^r \to 0$, as $n \to \infty$, for $r \geq 3$, and $\sum_{i=1}^{p}(n_i/2)(2\lambda_i/\sqrt{n})^2 < \infty$. Then,

$$\log M_Q(t/\sqrt{n}) \simeq \sum_{i=1}^{p}(n_i/2)(2\lambda_i/\sqrt{n})t + \sum_{i=1}^{p}(n_i/2)\frac{(2\lambda_i/\sqrt{n})^2}{2}t^2 \tag{2.21}$$

and

$$M_Q(t/\sqrt{n}) \simeq \exp\left\{\sum_{i=1}^{p}(n_i/2)(2\lambda_i/\sqrt{n})t + (1/2)\sum_{i=1}^{p}(n_i/2)(2\lambda_i/\sqrt{n})^2 t^2\right\}. \tag{2.22}$$

Therefore, Q is approximately normally distributed with mean $\mu = \sum_{i=1} p n_i \lambda_i/\sqrt{n}$ and variance $\sigma^2 = 2\sum_{i=1}^{p} n_i(\lambda_i/\sqrt{n})^2$. Hence,

$$\left(Q - \sum_{i=1}^{p} n_i\lambda_i/\sqrt{n}\right)\left(2\sum_{i=1}^{p} n_i(\lambda_i/\sqrt{n})^2\right)^{-\frac{1}{2}} \to N(0,1), \text{ as } n \to \infty \tag{2.23}$$

when

$$\sum_{i=1}^{p}(n_i/2)(2\lambda_i/\sqrt{n})^2 < \infty \text{ and } \sum_{i=1}^{p}(n_i/2)(2\lambda_i/\sqrt{n})^2 \to 0, \tag{2.24}$$

$$n \to \infty \text{ for } r \geq 3. \tag{2.25}$$

A special case of the quadratic form $Q = \boldsymbol{Z}'\boldsymbol{A}\boldsymbol{Z} = \sum_{i=1}^{p} \lambda_i(X_i + b_i)^2$ is when all λ_i's have

the same multiplicity n and equal b_i's. Then, $Q = \sum_{j=1}^{n} A_j$, $Q_j = \sum_{i=1}^{q} \lambda_i (Y_{ij} + b_i)^2$, $p = nq$, the Y_{ij}'s are mutually independent standard normal variables, where Q_j's have common mean and variance, $\mu = \sum_{i=1}^{p} \lambda_i (1 + b_i^2)$ and $\sigma^2 = 2 \sum_{i=1}^{p} \lambda_i^2 (a + 2b_i^2)$ respectively. By the central limit theorem $[(Q - \mu)/\Sigma] \rightarrow N(0, 1)$.

Chapter 3

Hotelling's T^2, Shortcomings and Remedies

In this chapter we review Hotelling's T^2, applications, and uses for testing as well as its shortcomings. In this chapter the notation $W_m(n, \Sigma)$ denotes the Wishart distribution, where m is the dimension and n is the degrees of freedom. If $\boldsymbol{A} \sim W_m(n, \boldsymbol{\Sigma})$, then it is well known that

$$E(\boldsymbol{A}^{-1}) = \frac{\boldsymbol{\Sigma}^{-1}}{n - m - 1}. \tag{3.1}$$

We will use this to have an unbiased estimator of $\boldsymbol{\Sigma}^{-1}$ when the sample is from a multivariate normal population. Suppose that $X_1, ..., X_n$ from $N_m(\boldsymbol{\mu}, \boldsymbol{\Sigma})$, and $\boldsymbol{\Sigma} > 0$.

$$(n-1)\boldsymbol{S}_{n-1} = \sum_{i=1}^{n}(\boldsymbol{x}_i - \bar{\boldsymbol{x}})(\boldsymbol{x}_i - \bar{\boldsymbol{x}})' \sim W_m(n-1, \boldsymbol{\Sigma}) \tag{3.2}$$

where $\boldsymbol{S}$ is the sample covariance. Then, using the above result, we have:

$$E[((n-1)\boldsymbol{S}_{n-1})^{-1}] = \frac{\boldsymbol{\Sigma}^{-1}}{n-1-m-1} = \frac{\boldsymbol{\Sigma}^{-1}}{n-m-2} \tag{3.3}$$

This implied that $\left(\frac{n-m-2}{n-1}\right)\boldsymbol{S}_{n-1}^{-1}$ is an unbiased estimator of $\boldsymbol{\Sigma}^{-1}$. Now we move on to hotelling's T^2 which is multivariate generalized t distribution for the multivariate setup.

Definition 3.0.1. *Suppose $\boldsymbol{S}_{m\times m} \sim W_m(n, \boldsymbol{\Sigma})$ and $\boldsymbol{d}_{m\times 1} \sim N(\boldsymbol{\delta}, c^{-1}\boldsymbol{\Sigma})$, where c is a scalar and $\boldsymbol{S}$ and $\boldsymbol{d}$ are independently distributed, the Hotelling's T^2 distribution is defined as*

$$T^2 = cn\boldsymbol{d}'\boldsymbol{S}^{-1}\boldsymbol{d}, \tag{3.4}$$

and is called Hotelling's T^2, where n degrees of freedom is associated with the Wishart distribution.

3.1 Some Properties of the Distribution of Hotelling's T^2

In order to get to the distribution of T first observe that given $\boldsymbol{d}$

$$\frac{\boldsymbol{d}'\boldsymbol{\Sigma}^{-1}\boldsymbol{d}}{\boldsymbol{d}'\boldsymbol{S}^{-1}\boldsymbol{d}} \sim \chi^2_{n-m+1} \tag{3.5}$$

and is independent of $\boldsymbol{d}$. Further, consider $\boldsymbol{d} \sim N_m(\boldsymbol{\delta}, c^{-1}\boldsymbol{\Sigma})$. Then

$$\boldsymbol{d}'(c^{-1}\boldsymbol{\Sigma})^{-1}\boldsymbol{d} \sim \chi^2_m(\tau^2) \tag{3.6}$$

where $\tau^2 = \boldsymbol{\delta}'(c^{-1}\boldsymbol{\Sigma})^{-1}\boldsymbol{\delta} = c\boldsymbol{\delta}\boldsymbol{\Sigma}^{-1}\boldsymbol{\delta}$ is the noncentrality parameter. Obviously this relation is a quadratic form in $\boldsymbol{d}$. One can conclude that $c\boldsymbol{d}'\boldsymbol{\Sigma}^{-1}\boldsymbol{d}$ is independent of $\frac{\boldsymbol{d}'\boldsymbol{\Sigma}^{-1}\boldsymbol{d}}{\boldsymbol{d}'\boldsymbol{S}^{-1}\boldsymbol{d}}$. Then,

$$\frac{cd'\boldsymbol{\Sigma}^{-1}\boldsymbol{d}/m}{\left(\frac{\boldsymbol{d}'\boldsymbol{\Sigma}^{-1}\boldsymbol{d}}{\boldsymbol{d}'\boldsymbol{S}^{-1}\boldsymbol{d}}\right)/(n-m+1)} \sim F'_{m,n-m+1}(\tau^2) \tag{3.7}$$

which is a noncentral F distribution with noncentrality parameter τ^2. From this we have

$$c.\boldsymbol{d}'\boldsymbol{S}^{-1}\boldsymbol{d}.\frac{n-m+1}{m} \sim F_{m,n-m+1}(\tau^2). \tag{3.8}$$

Hence, one can conclude

$$\frac{T^2}{n}\frac{n-m+1}{m} \sim F_{m,n-m+1}(\tau^2). \tag{3.9}$$

3.2 Applications of Hotelling's T^2 Statistic

Let's assume we have a multivariate normal population, $N(\boldsymbol{\mu}, \boldsymbol{\Sigma})$, $\boldsymbol{\Sigma} > 0$. It is desired to test the hypothesis $H_0 : \boldsymbol{\mu} = \boldsymbol{\mu}_0$ vs $H_a : \boldsymbol{\mu} \neq \boldsymbol{\mu}_0$. This is a type of testing which we frequently come across in multivariate theory.

Assume we have a sample of $X_1, ..., X_n$ from $N_m(\boldsymbol{\mu}, \boldsymbol{\Sigma})$ and it is of the interest to use this sample to test the hypothesis. It is known that $\bar{X}$ and S independent and

$$\bar{X} \sim N_m(\boldsymbol{\mu}, \boldsymbol{\Sigma}/n),$$

$$and$$

$$(n-1)\boldsymbol{S} \sim W_m(n-1, \boldsymbol{\Sigma}).$$

It is also known that under the null hypothesis $(\bar{X} - \boldsymbol{\mu}_0) \sim N_m(0, \boldsymbol{\Sigma}/n)$. Hence, Hotelling's T^2 is

$$T^2 = n(\bar{x} - \boldsymbol{\mu}_0)' \boldsymbol{S}^{-1} (\bar{x} - \boldsymbol{\mu}_0).$$

Following the same steps taken before we have

$$\frac{T^2}{n-1} . \frac{(n-1) - m + 1}{m} \sim F_{m,n-m}(\tau^2),$$

where m is the dimension of the samples taken from the population. Note that under the null hypothesis $\tau^2 = 0$. Testing the hypothesis is then based on the F distribution and in case of $m = 1$ the test statistics is reduced to t test.

Hotelling's T^2 can also be applied to testing the equality of two population means. Following the same idea of one sample test and on condition that the covariance matrices for both populations are identical one can develop the two sample test statistic as follows

$$T^2 = (\bar{x}_1 - \bar{x}_2)' (\boldsymbol{S}_p (\frac{1}{n_1} + \frac{1}{n_2}))^{-1} (\bar{x}_1 - \bar{x}_2)$$

where $\boldsymbol{S}_p$ is the pooled sample covariance matrix,

$$\boldsymbol{S}_p = \frac{(n_1 - 1)\boldsymbol{S}_1 + (n_2 - 1)\boldsymbol{S}_2}{n_1 + n_2 - 2}.$$

The distribution of T^2 in two sample case can be transformed to F distribution as the same as in the one sample problem.

$$\frac{n_1 + n_2 - m - 1}{m(n_1 + n_2 - 2)} T^2 \sim F_{m, n_1 + n_2 - m - 1}$$

where m in the dimension of the samples taken from the population. Based on the distribution of the statistic under the null hypothesis one can have construct the test. Hence, the null hypothesis $H_0 : \boldsymbol{\mu}_1 = \boldsymbol{\mu}_2$ is rejected against $H_a : \boldsymbol{\mu}_1 \neq \boldsymbol{\mu}_2$ at a significance level of α if

$$F > F_{m, n_1 + n_2 - m, \alpha}$$

3.3 Some Notes About Hotelling's T^2

The relationship between likelihood ratio test and Hotelling's T^2 is as follows:

$$\Lambda^{2/n} = \left(1 + \frac{T^2}{(n-1)}\right)^{-1}$$

where Λ is the test statistic obtained from likelihood ratio test. This relation shows the equivalence between Hotelling's T^2 and the likelihood ratio test, Johnson and Wichern (2007).

In cases where $\boldsymbol{X}$ is distributed as $N_m(\boldsymbol{\mu}, \boldsymbol{\Sigma})$ with $|\boldsymbol{\Sigma}| > 0$ then $(\boldsymbol{X} - \boldsymbol{\mu})'\boldsymbol{\Sigma}^{-1}(\boldsymbol{X} - \boldsymbol{\mu})$ has χ^2_m distribution. Consequently, it is used in testing the hypothesis $H_0 : \boldsymbol{\mu} = \boldsymbol{\mu}_0$ vs $H_a : \boldsymbol{\mu} \neq \boldsymbol{\mu}_0$. In other words the $N_m(\boldsymbol{\mu}, \boldsymbol{\Sigma})$ assigns probability $1 - \alpha$ to the solid ellipsoid $\{\boldsymbol{x} : (\boldsymbol{x} - \boldsymbol{\mu})'\boldsymbol{\Sigma}^{-1}(\boldsymbol{x} - \boldsymbol{\mu}) \leq \chi^2_m(\alpha)\}$, where $\chi^2_m(\alpha)$ denotes the upper (100α)th percentile of the χ^2_m distribution. It is due to the fact that $(\boldsymbol{X} - \boldsymbol{\mu})'\boldsymbol{\Sigma}^{-1}(\boldsymbol{X} - \boldsymbol{\mu})$ can be written as the sum of m chi-square distribution with one degree of freedom Johnson and Wichern (2007). The confidence region can also be constructed. A $100(1 - \alpha)$ confidence region is the ellipsoid determined by all $\boldsymbol{\mu}$ such that

$$n(\bar{\boldsymbol{x}} - \boldsymbol{\mu})'\boldsymbol{S}^{-1}(\bar{\boldsymbol{x}} - \boldsymbol{\mu}) \leq \frac{m(n-1)}{(n-m)} F_{m,n-m}(\alpha)$$

where $\bar{\boldsymbol{x}}$ is the sample mean and $\boldsymbol{S}$ is the sample covariance matrix.

The test of the hypothesis about $\boldsymbol{\mu}$ and confidence region when the underlying population in not normal, is also possible on condition that $n - m$ is sufficiently large. The inference about $\boldsymbol{\mu}$ is based on the fact that $n(\bar{\boldsymbol{X}} - \boldsymbol{\mu})'\boldsymbol{S}^{-1}(\bar{\boldsymbol{X}} - \boldsymbol{\mu})$ is approximately distributed as χ^2_m, where m is the dimension. The associated test statistic and confidence region are then as follows:

Consider a random sample $\boldsymbol{X}_1, ..., \boldsymbol{X}_m$ from a population with mean $\boldsymbol{\mu}$ and positive definite covariance matrix $\boldsymbol{\Sigma}$. The null hypothesis $H_0 : \boldsymbol{\mu} = \boldsymbol{\mu}_0$ is rejected against the alternative hypothesis $H_a = \boldsymbol{\mu} \neq \boldsymbol{\mu}_0$ when

$$n(\bar{\boldsymbol{x}} - \boldsymbol{\mu})'\boldsymbol{S}^{-1}(\bar{\boldsymbol{x}} - \boldsymbol{\mu}_0) > \chi^2_m(\alpha).$$

The confidence region is also based on the χ^2 distribution such that

$$P[n(\bar{\boldsymbol{X}} - \boldsymbol{\mu})'\boldsymbol{S}^{-1}(\bar{\boldsymbol{X}} - \boldsymbol{\mu}) \leq \chi^2_m(\alpha)] = 1 - \alpha,$$

where $\chi^2_m(\alpha)$ is the upper (100α)th percentile of the $\chi^2_m(\alpha)$ distribution.

3.4 Shortcomings of Hotelling's T^2

In the following some barriers against using Hotelling's T^2 are discussed briefly. Some approaches to the limitations are discussed afterwards.

Due to the wide variety of applications Hotelling's T^2, its properties have been examined in the literature. Iwashita (1997) addressed asymptotic null and nonnull distribution of Hotelling's T^2 statistics for elliptical distributions. He showed that the approximate power of T^2-test under the multivariate t-distribution and the Kotz-type distribution is

$$\begin{aligned}\beta(\delta) &= P_\delta[x \geq c_x] \\ &= (1 - G_p(c_x;\delta)) + \frac{1}{8N}\{\sum_{j=0}^{4} b_j(p,k;\delta)(1 - G_{p+2j}(c_x;\delta))\} + O(N^{-1}),\end{aligned}$$

where $G_k(x;\delta)$ denotes cumulative function of noncentral chi-squared distribution with k degrees of freedom and noncentral parameter δ,c_x is the upper $100\alpha\%$ point for Hotelling's T^2 and $b_j(p,k;\delta)$'s are as follows

$$\begin{aligned}b_0(p,k;\delta) &= -\{2p^2 + k\{2p(p+2) - \delta^2\}\} \\ b_1(p,k;\delta) &= -2\{2\{p - \delta\} - k\{2p(p+2) - 4(p+2)\delta + \delta^2\}\} \\ b_2(p,k;\delta) &= 2\{p - \delta\}\{p + 2 - \delta\} - 8\delta - k\{2p(p+2) - 12(p+2)\delta + 9\delta^2\} \\ b_d(p,k;\delta) &= 2\delta^2(1 - k)\end{aligned}$$

Bai and Saranadasa (1996) analyzed the asymptotic power of the Hotelling's T^2 test when $p < n$ and the sample covariance matrix is still invertible. In their considered setting

$y_n = \frac{p}{n} \to y \in (0,1)$, $N_1/(N_1+N_2) \to k \in (0,1)$ and $||\boldsymbol{\delta}||^2 = o(1)$, then one can have

$$\beta_H(\boldsymbol{\delta}) - \Phi(-\xi + \sqrt{\frac{n(1-y)}{2y}} k(1-k)||\boldsymbol{\delta}||^2) \to 0,$$

where p is the dimension, N_1 and N_2 are the sample size, $\boldsymbol{\delta} = \boldsymbol{\Sigma}^{-\frac{1}{2}}(\boldsymbol{\mu}_1 - \boldsymbol{\mu}_2)$, ξ_α is the $1-\alpha$ quantile of standard normal distribution, and $\beta_H(\boldsymbol{\delta})$ is the power function of Hotelling's test. The usual consideration of the alternative hypothesis in limiting theorems is to assume that $\sqrt{n}||\boldsymbol{\delta}||^2 \to a > 0$, where a is a positive constant. Under this additional assumption it follows from the above relation that the limiting power of Hotelling's test is given by $\Phi(-\xi_\alpha + ((1-y)/2y)^{1/2} k(1-k)a)$. This formula shows that the limiting power of the Hotelling's test is slowly increasing for y close to 1, as the non-centrality parameter (namely a) increases. Despite many advantages of Hotelling's T^2 such as being invariant under the linear transformation, having known exact distribution under the null hypothesis, and being powerful when dimension is sufficiently small compared with the sample size, it has also some intrinsic disadvantages.

Bai and Saranadasa (1996) showed that Hotelling's T^2 loses its power when p gets closer to n when p is still less than n. Moreover, it is clear in cases where p is larger than n the sample covariance is no longer invertible and hence not applicable in many disciplines which are so called "high dimensional" settings.

3.5 Remedies

Nowadays' it has become easier to deal with high dimensional data due to the modern computation techniques. An example of interest in high dimensional data is DNA microarray data where there are gene expressions on thousands of genes of only a few subjects. Consequently statisticians are in search of new methods and approaches to deal with these kinds of problems. Making inference about mean of a population is one of many interesting problem in high dimensional settings.

Shrinkage is an established way to circumvent the issues related to singularity of the sample

covariance. Stein (1956) noted about poor performance of the sample covariance matrix when p and p/n are large. The problem lies in the tendency of the largest eigenvalue of $\boldsymbol{S}$ to be larger and the smallest one to be smaller. It can be seen in following figures that for an identity covariance matrix with all eigenvalues equal to one, the range of the sample eigenvalues gets larger for larger values of the ration p/n, see Pourahmadi (2013).

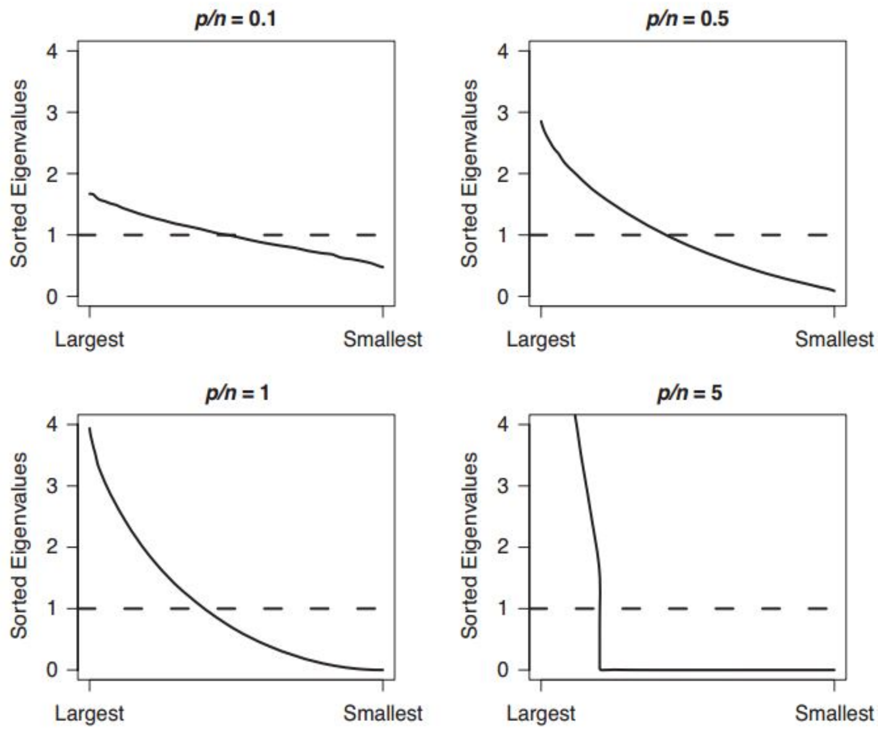

Figure 3.1: The solid line represents the distribution of the eigenvalues of the sample covariance matrix sorted from largest to smallest.

The idea of improving the sample covariance is then based on shrinking the eigenvalues toward a central value. The amount of shrinking is bases on the selected loss function to be minimized. Consider $\hat{\boldsymbol{\Sigma}} = \hat{\boldsymbol{\Sigma}}(\boldsymbol{S})$ to be the estimator of the covariance matrix. The two

common loss functions with their corresponding risk functions are as follows

$$L_1(\hat{\mathbf{\Sigma}}, \mathbf{\Sigma}) = tr(\hat{\mathbf{\Sigma}}\mathbf{\Sigma}^{-1}) - \log|\hat{\mathbf{\Sigma}}\mathbf{\Sigma}^{-1}| - p$$
$$L_2(\hat{\mathbf{\Sigma}}, \mathbf{\Sigma}) = tr(\hat{\mathbf{\Sigma}}\mathbf{\Sigma}^{-1} - I)^2$$

and

$$R_i(\hat{\mathbf{\Sigma}}, \mathbf{\Sigma}) = E_{\mathbf{\Sigma}} L_i(\hat{\mathbf{\Sigma}}, \mathbf{\Sigma}), i = 1, 2.$$

As a selection rule the smaller risk function the better estimator. The loss function L_1 is usually called entropy loss or Kullback-Liebler divergence of two multivariate multivariate normal densities corresponding to two covariance matrices.

The two stated loss functions yield different estimators. As an example it is known that among all estimators in the form $a\boldsymbol{S}, a > 0$, $\boldsymbol{S}$ is optimal under L_1, while the estimator $\frac{n\boldsymbol{S}}{(n+p+1)}$ is optimal under L_2.

The shrinkage estimation suggested by Stein (1956) is

$$\hat{\mathbf{\Sigma}} = \hat{\mathbf{\Sigma}}(\boldsymbol{S}) = P\Phi(\lambda)P'$$

where $\lambda = (\lambda_1, ..., \lambda_p)', \lambda_1 > ... > \lambda_p > 0$ are eigenvalues of $\boldsymbol{S}$. P is an orthonormal matrix with the eigenvectors as its columns and $\Phi(\lambda)$ =diag$(\varphi_1, ..., \varphi_p)$ is a diagonal matrix such that its diagonal elements are estimation of the eigenvalues of $\mathbf{\Sigma}$ from largest to smallest. For example, setting $\varphi_j = \lambda_j$ yields the usual unbiased estimator $\boldsymbol{S}$. Stein's shrinkage suggests to shrink eigenvalues of $\boldsymbol{S}$ toward a central value to fix the biases of the eigenvalues of $\boldsymbol{S}$. The modified estimator under the risk function L_1 is then

$$\varphi_j = \lambda_j.n\big(n - p + 1 + 2\lambda_j \sum_{i \neq j} \frac{1}{\lambda_j - \lambda_i}\big)^{-1}$$

Unfortunately in the above relation φ_j's can be negative and they may not be monotone. A modification to this is proposed by Lin and Perlman (1985). Based on linear combinations of the sample covariance matrix and the identity matrix. Ledoit and Wolf (2004) proposed a class of well-conditioned shrinkage estimators in the form of

$$\hat{\mathbf{\Sigma}}_{LW} = \alpha_1 \boldsymbol{I} + \alpha_2 \boldsymbol{S}$$

where α_1 and α_2 are scalars to be estimated. The corresponding risk function in this setting is Frobenius norm and α_1 and α_2 are estimated through minimizing the risk function.

Another approach to dealing with singularity of sample covariance when $p > n$ is to avoid estimating the sample covariance. This approach can be found in Dempster (1960), Srivastava (2009), Ahmad et al. (2013), and Bai and Saranadasa (1996).
Srivastava's test statistics is

$$T_1 = \frac{N\bar{\boldsymbol{x}}'\boldsymbol{D}_s^{-1}\bar{\boldsymbol{x}} - \frac{np}{n-2}}{\sqrt{2(tr\boldsymbol{R}^2 - \frac{p^2}{n})c_{p,n}}}$$

where

$$\boldsymbol{R} = \boldsymbol{D}_s^{-\frac{1}{2}}\boldsymbol{S}\boldsymbol{D}_s^{-\frac{1}{2}} = (r_{ij})$$

denotes the sample correlation matrix, and $c_{p,n}$ is an adjustment coefficient such that

$$c_{p,n} \xrightarrow{p} 1 \qquad as \quad (n,p) \to \infty.$$

One particular choice of $c_{p,n}$ proposed by Srivastava and Du (2008) is

$$c_{p,n} = 1 + \frac{tr\boldsymbol{R}^2}{p^{3/2}}.$$

Srivastava showed that under some strict assumptions the test statistic T_1 will have a standard normal distribution under the null hypothesis. While under relatively strict assumptions Srivastava's test statistic performs well for large p's, its poor performance in small p's, particularly less than 60, is a disadvantage to this test.
Dempster based his test on the statistic

$$T_2 = \frac{N\bar{\boldsymbol{x}}'\bar{\boldsymbol{x}}}{tr\boldsymbol{S}}$$

where $\bar{\boldsymbol{x}}$ is the sample mean, and $\boldsymbol{S}$ denotes sample variance as in (4.23). Its distribution is different from ours. Under the null hypothesis $H_0 : \boldsymbol{\mu} = \boldsymbol{0}$ and using orthogonal transformation it can be shown that

$$T_2 = \frac{NQ_1}{Q_2 + ... + Q_N}$$

where the Q_i's are independent and identically distributed. In fact Dempster (1958) made the assumption that each Q_i, $i = 1, ..., N$, is approximately distributed as $m\chi_r^2$, where $m > 0$ and χ_r^2 is a chi-square distribution with r degrees of freedom. By matching the first and second moment of Q_i and $m\chi_r^2$, Dempster obtained:

$$r = \frac{(tr\mathbf{\Sigma})^2}{tr\mathbf{\Sigma}^2} = p\frac{a_1^2}{a_2}$$

where $a_1 = tr\mathbf{\Sigma}/p$, and $a_2 = tr\mathbf{\Sigma}^2/p$. The estimators for a_1, a_2, and r are as follows

$$\hat{a}_1 = \frac{tr\boldsymbol{S}}{p}, \quad \hat{a}_2 = \frac{n^2}{(n-1)(n+2)}\frac{1}{p}\left[tr\boldsymbol{S}^2 - \frac{(tr\boldsymbol{S})^2}{n}\right], \quad \text{and} \quad \hat{r} = p\frac{\hat{a}_1^2}{\hat{a}_2}$$

It is proposed that T_2 is approximately distributed as F with $[\hat{r}]$ and $[n\hat{r}]$ as degrees of freedom, where $[x]$ denotes the largest number less than or equal to x.

Ahmad proposed the test statistics T_3 as is below for the mean vector of high dimensional multivariate data. The proposed test statistic is

$$T_3 = \frac{\frac{N\bar{\boldsymbol{x}}'\bar{\boldsymbol{x}}}{tr\boldsymbol{S}} - 1}{\sqrt{\frac{2}{f}}},$$

where N is the sample size and

$$f = \frac{[tr(\mathbf{\Sigma})]^2}{tr(\mathbf{\Sigma}^2)}.$$

f is estimated from the data. Ahmad proposed that under some lenient assumptions the test statistic T_3 has a standard normal distribution.

The next chapter is devoted to introducing of a new test for the mean vector in high dimensional setting. The pros and cons are discussed and the results are compared with the competing tests.

Chapter 4

A New Test for the Mean Vector in High Dimensional Setting

The Hotelling's T-square statistic for testing hypotheses about the multivariate mean vector is well known in Multivariate Analysis. It is also well known that such a test is not applicable in the case where the dimension of the multivariate population is higher than the sample size, as is often for example in Bio-Informatics problems. In these problems N is the number of individuals and on each individual we observe a very large number of gene expressions p; the case is most commonly referred to as the "p greater than N case". In situations like these, the $p \times p$ sample covariance matrix $\boldsymbol{S}$, is not invertible and hence one cannot use Hotelling's T-square. In this paper we consider the test statistic $T = N\bar{\boldsymbol{x}}'\bar{\boldsymbol{x}}/tr(\boldsymbol{S})$ where $\bar{\boldsymbol{x}}$ and $\boldsymbol{S}$ are the sample mean vector and the sample covariance matrix. Asymptotically, for large N and/or large p, the test is shown to have a normal distribution. Subsequently, we consider a power transformation of the test statistic that is shown to reduce skewness and accelerate convergence to normality. Then the test is evaluated numerically under the null hypothesis and its performance is shown to be better than other tests in the literature.

4.1 Introduction

Let $\boldsymbol{X} = (X_1, X_2, ..., X_p)'$ be a $p \times 1$ random vector with mean $\boldsymbol{\mu}$ and a positive definite covariance matrix $\boldsymbol{\Sigma}$. The representation below is useful in obtaining the distribution of the quadratic form $\boldsymbol{X}'\boldsymbol{A}\boldsymbol{X}$ where $\boldsymbol{A}$ is a positive definite matrix. Letting $\boldsymbol{X} = \boldsymbol{\Sigma}^{\frac{1}{2}}\boldsymbol{Y} + \boldsymbol{\mu}$,

the quadratic form is subsequently expressed as follows:

$$\begin{aligned}
\boldsymbol{X}'\boldsymbol{A}\boldsymbol{X} &= (\boldsymbol{\Sigma}^{\frac{1}{2}}\boldsymbol{Y}+\boldsymbol{\mu})'\boldsymbol{A}(\boldsymbol{\Sigma}^{\frac{1}{2}}\boldsymbol{Y}+\boldsymbol{\mu}) \\
&= [\boldsymbol{\Sigma}^{\frac{1}{2}}(\boldsymbol{Y}+\boldsymbol{\Sigma}^{-\frac{1}{2}}\boldsymbol{\mu})]'\boldsymbol{A}[\boldsymbol{\Sigma}^{\frac{1}{2}}(\boldsymbol{Y}+\boldsymbol{\Sigma}^{-\frac{1}{2}}\boldsymbol{\mu}) \\
&= (\boldsymbol{Y}+\boldsymbol{\Sigma}^{-\frac{1}{2}}\boldsymbol{\mu})'\boldsymbol{\Sigma}^{\frac{1}{2}}\boldsymbol{A}\boldsymbol{\Sigma}^{\frac{1}{2}}(\boldsymbol{Y}+\boldsymbol{\Sigma}^{-\frac{1}{2}}\boldsymbol{\mu}) \\
&= (\boldsymbol{Y}+\boldsymbol{\Sigma}^{-\frac{1}{2}}\boldsymbol{\mu})'\boldsymbol{P}'\boldsymbol{D}\boldsymbol{P}(\boldsymbol{Y}+\boldsymbol{\Sigma}^{-\frac{1}{2}}\boldsymbol{\mu}) \\
&= (\boldsymbol{P}\boldsymbol{Y}+\boldsymbol{P}\boldsymbol{\Sigma}^{\frac{1}{2}}\boldsymbol{\mu})'\boldsymbol{D}(\boldsymbol{P}\boldsymbol{Y}+\boldsymbol{P}\boldsymbol{\Sigma}^{\frac{1}{2}}\boldsymbol{\mu}) \\
&= (\boldsymbol{U}+\boldsymbol{b})'\boldsymbol{D}(\boldsymbol{U}+\boldsymbol{b})
\end{aligned}$$

where $\boldsymbol{P}$ is an orthogonal matrix such that $\boldsymbol{P}\boldsymbol{P}' = \boldsymbol{I}$ and $\boldsymbol{P}\boldsymbol{\Sigma}^{\frac{1}{2}}\boldsymbol{A}\boldsymbol{\Sigma}^{\frac{1}{2}}\boldsymbol{P}' = \boldsymbol{D}$ where $\boldsymbol{D} = \text{diag}(\lambda_1, \lambda_2, ..., \lambda_p)$, $\boldsymbol{U} = \boldsymbol{P}\boldsymbol{Y}$, and $\boldsymbol{b} = \boldsymbol{P}'\boldsymbol{\Sigma}^{\frac{1}{2}}\boldsymbol{\mu}$. Noting that $\boldsymbol{U}$ has mean $\boldsymbol{0}$ and covariance matrix $\boldsymbol{I}$, the quadratic form is expressed as

$$\boldsymbol{X}'\boldsymbol{A}\boldsymbol{X} = \sum_{j=1}^{p} \lambda_j (u_j + b_j)^2 \tag{4.1}$$

where the u_j's, $j = 1, ..., p$ are independent random variables with mean 0 and variance 1, and b_j's, $j = 1, ..., p$ are constants.

Now, if we assume that $\boldsymbol{X}$ is $N(\boldsymbol{\mu}, \boldsymbol{\Sigma})$, then the u_j's, $j = 1, ...p$ are independent standard normal random variables. In this case, the cumulants of the quadratic form in (1) are given by

$$K_r = 2^{r-1}(r-1)! \sum_{j=1}^{p} \lambda_j^r (1 + r b_j^2) \quad , \quad r = 1, 2, ... \tag{4.2}$$

If we let $\theta_r = \sum_{j=1}^{p} \lambda_j^r (1 + r b_j^2)$, then the first four central moments of $\boldsymbol{X}'\boldsymbol{A}\boldsymbol{X}$ are, see Mathai and Provost (1992a):

$$\mu_1' = K_1 = \theta_1 \tag{4.3}$$

$$\mu_2 = K_2 = 2\theta_2 \tag{4.4}$$

$$\mu_3 = K_3 = 8\theta_3 \tag{4.5}$$

$$\mu_4 = K_4 + 3\mu_2^2 = 12(4\theta_4 + \theta_2^2) \tag{4.6}$$

4.2 An Approximation to the Distribution of the Quadratic Form

Let Q be a positive definite quadratic form and let θ_1 be the mean of Q. Wilson and Hilferty (1931) assumed Q to be a chi-square random variable. In their pioneer work on this subject they showed that a transformation of the form $\left(\frac{Q}{\theta_1}\right)^{1/3}$ is a better normalizing transformation than Fisher's square root transformation. The fact is that $\frac{1}{3}$ gives "less" skewness than $\frac{1}{2}$. Building on this idea, Jensen and Solomon (1972), considered the transformation $\left(\frac{Q}{\theta_1}\right)^h$ where h is chosen so that it minimizes the skewness. In the latter, all moments of $\left(\frac{Q}{\theta_1}\right)^h$ are expressed as series in powers of θ_1^{-1}. Obviously, the approximation works better if θ_1 is large, as is the case in the chi-square or any positive random variable with large mean. Since then, the transformation has been used in a number of instances, see for example Mudholkar and Trivedi (1980), Moschopoulos and Mudholkar (1983), Moschopoulos (1983), Staniswalis et al. (1993). Mathai and Provost (1992a) present the power transformation specifically for a quadratic form.

Let

$$Z = \left(\frac{Q}{\theta_1}\right)^h = (1 + \frac{Q-\theta_1}{\theta_1})^h = (1 + \frac{V}{\theta_1})^h \tag{4.7}$$

where $V = Q - \theta_1$. The moments of Z using a binomial expansion are,

$$\begin{aligned} \mu_r'(h) &= E(Z^{rh}) = E(1+\frac{V}{\theta_1})^{rh} \\ &= \sum_{r=0}^{\infty} \left(\frac{h(h-1)...(h-r+1)}{r!}\right) \theta_1^{-r} E(V^r) \end{aligned} \tag{4.8}$$

For $r = 1$, the above relation gives the first moment of Z, $\mu_1'(h)$ as

$$\begin{aligned} \mu_1'(h) \simeq & 1 + \frac{1}{\theta_1}(h(h-1)\varphi 2) + \frac{h(h-1)(h-2)}{\theta_1^2}\{(\frac{4}{3})\varphi_3 + (\frac{h-3}{2}\varphi_2^2\} \\ & + \frac{h(h-1)(h-2)(h-3)}{\theta_1^3}\{2\varphi_4 + \frac{4(h-4)}{3}\varphi_2\varphi_3 + \frac{(h-4)(h-5)}{6}\varphi_2^3\}, \end{aligned} \tag{4.9}$$

where $\varphi_r = \theta_r/\theta_1$, $r = 2, 3, 4$, are assumed to be bounded, Jensen and Solomon (1972). Since $\mu_r'(h) = \mu_1'(rh)$ one can obtain $\mu_r'(h)$ by substituting h with rh. After lengthy

computations and writing μ_r in terms of μ'_j, $j = 1, ..., r$, one can obtain the central moments $\mu_r(h)$. The second and third central moments are as follows:

$$\mu_2(h) \simeq \frac{1}{\theta_1}[2h^2\varphi_2] + 2\frac{h^2(h-1)}{\theta_1^2}\{4\varphi_3 + (3h-5)\varphi_2^2\} + 4\frac{h^2(h-1)}{\theta_1^3} \tag{4.10}$$
$$\{(7h-11)\varphi_4 + \frac{4(h-2)(7h-12)}{3}\varphi_2\varphi_3 + \frac{(h-2)(14h^2-60h+64)}{6}\varphi_2^3\}$$

and

$$\mu_3(h) \simeq \frac{4h^3}{\theta_1^2}(2\varphi_3 + 3(h-1)\varphi_2^2) \tag{4.11}$$
$$\frac{h^3(h-1)}{\theta_1^3}(72\varphi_4 + 24(7h-10)\varphi_2\varphi_3 + 4(17h^2-55h+44)\varphi_2^3).$$

From the above expressions one can have the skewness $\tau_1(h) = \mu_3(h)/(\mu_2(h))^{\frac{3}{2}}$. By approximating $\mu_2(h)$ and writing $\mu_2(h) \simeq \theta_1^{-1}(2h^2\varphi_2)$, one can have

$$\tau_1(h) \simeq 4(2\varphi_3 + 3(h-1)\varphi_2^2)/[\theta_1^{\frac{1}{2}}(2\varphi_2)^{\frac{3}{2}}] + (h-1)\{72\varphi 4 \tag{4.12}$$
$$+ 24(7h-10)\varphi_2\varphi_3 + 4(17h^2 - 55h + 44\varphi_2^3\}/[\theta_1^{\frac{3}{2}}(2\varphi_2)^{\frac{3}{2}}]. \tag{4.13}$$

After obtaining the skewness in powers of θ_1^{-1} and letting h to be the value such that the leading terms of $\mu_3(h)$ and $\tau_1(h)$ vanish, the distribution of $(Q/\theta_1)^h$ will be more symmetric and closer to a normal distribution. Thus

$$h_0 = 1 - 2\theta_1\theta_3/3\theta_2^2 = 1 - k_1k_3/3k_2^2. \tag{4.14}$$

Upon substituting h with h_0 in (4.9), (4.10), (4.11), and (4.12), one can compute $\mu'_1(h_0)$, $\mu_2(h_0)$, $\mu(h_0)$ and $\tau_1(h_0)$. Then the distribution of $Z = (Q/\theta_1)^{h_0}$ is approximated by a normal distribution with mean and variance as:

$$\mu'_1(h_0) = 1 + \theta_2h_0(h_0-1)/\theta_1^2 \tag{4.15}$$
$$\mu_2(h_0) = 2\theta_2h_0^2/\theta_1^2. \tag{4.16}$$

Hence, the approximation is that the statistic

$$u = \theta_1\{(Q/\theta_1)^{h_0} - 1 - \theta_2h_0(h_0-1)/\theta_1^2\}/(2\theta_2h_0^2)^{\frac{1}{2}} \tag{4.17}$$

can be approximated by a standard normal distribution. As an example, when $\lambda_1 = ... = \lambda_p = 1$, the quadratic form Q has a noncentral chi square distribution with p degrees of freedom and noncentrality parameter $\lambda = \sum_{j=1}^{p} b_j^2$. In the central case $h_0 = \frac{1}{3}$, which is confirmed by the results obtained by Wilson and Hilferty (1931). Moschopoulos (1983), offered a modification for Wilson-Hilferty by the normalizing transformation $\{(\chi_k^2(\lambda) + b)/(k+\lambda)\}^h$ in which various choices of b may lead to improvements Moschopoulos (1983).

4.3 The Quadratic form and its distribution

Let $\bar{\boldsymbol{X}}$ be the mean of a sample of size N from the normal distribution with mean $\boldsymbol{\mu}$ and covariance matrix $\boldsymbol{\Sigma}$. The statistic, $Q = N\bar{\boldsymbol{X}}'\bar{\boldsymbol{X}}$ is a quadratic form in normal random variables. Ahmad et al. (2008), showed that under the null hypothesis $H_0 : \boldsymbol{\mu} = \boldsymbol{0}$ and for large N and p, $W = \frac{Q}{tr(\boldsymbol{S})}$, has a distribution like $\frac{\chi_f^2}{f}$ where

$$S = \frac{1}{N}\sum_{k=1}^{N} \boldsymbol{X}_k \boldsymbol{X}_k' \tag{4.18}$$

and

$$f = \frac{[tr(\boldsymbol{\Sigma})]^2}{tr(\boldsymbol{\Sigma}^2)}. \tag{4.19}$$

If we let $A_k = \boldsymbol{X}_k'\boldsymbol{X}_k$ and $A_{kl} = \boldsymbol{X}_k'\boldsymbol{X}_l$, then f can be estimated by $\frac{E_2}{E_3}$, where E_2 and E_3 are as follows:

$$E_2 = \frac{1}{N(N-1)}\sum_{k=1}^{N}\sum_{l=1}^{N} A_k A_l, k \neq l \quad E_3 = \frac{1}{N(N-1)}\sum_{k=1}^{N}\sum_{l=1}^{N} A_{kl}^2, k \neq l. \tag{4.20}$$

Ahmad et al. (2013), showed that even by relaxing the assumption of normality, the distribution of W behaves like a $\frac{\chi_f^2}{f}$, under the assumptions below:

Assumption 1: $E(\boldsymbol{X}_{ks}^4) \leq \gamma < \infty$, $\forall s = 1, ..., p$, for some finite constant γ independent of p.

Assumption 2: For $p \to \infty$, let $\frac{tr(\boldsymbol{\Sigma})}{p} = O(1)$

Assumption 3: For $p \to \infty$, let $\frac{tr(\boldsymbol{\Sigma}^2)}{p^2} = O(\delta)$, where, $0 < \delta \leq 1$.

Assumption 1 states that moments of the elements of the vector $\boldsymbol{X}_k$ exist and are finite up to the fourth order. This assumption is somewhat a substitute for normality assumption, especially when we are dealing with the second moment of quadratic forms involved in the computations, see Atiqullah (1962), Rao and Kleffe (1988), Wiens (1992), and Srivastava (2009). Assumption 3 is different from what is usually stated in the literature for high dimensional settings that is, $tr(\boldsymbol{\Sigma}^2)/p = O(1)$. It can be shown that $tr(\boldsymbol{\Sigma}^2)/p$ diverges to infinity as p grows large in many important cases. In fact, in most cases assumption 3 has been made with higher power of $\boldsymbol{\Sigma}$, see for example, Srivastava (2007), Fisher et al. (2010), Chen et al. (2010), and Ledoit and Wolf (2002).

As a typical example, in the intraclass correlation matrix $\boldsymbol{\Sigma} = (1-\rho)\boldsymbol{I} + \rho\boldsymbol{J}$, where $\boldsymbol{I}$ is the identity matrix, $\boldsymbol{J}$ is a matrix of 1's, and ρ is the correlation, $tr(\boldsymbol{\Sigma}^2)/p^2 = O(1)$, as $p \to \infty$. Except for some cases such as when the covariance matrix is singular or near singular, it can be shown that the assumption 3 would not be violated in many practical cases such as in compound symmetric, autoregressive of order 1, or unstructured matrix, see Ahmad et al. (2013). For an elaborate discussion on the assumptions adopted in high dimensional literature, see Ahmad et al. (2011).

4.4 A New Test for the Multivariate Mean

The testing of hypothesis about mean vector will basically be based on the quadratic form

$$Q = N\bar{\boldsymbol{X}}'\bar{\boldsymbol{X}}$$

where $\bar{\boldsymbol{X}}$ is the mean vector with covariance matrix $\boldsymbol{\Sigma}/N$. Such a quadratic form has been considered in Dempster (1958), and more recently in Ahmad et al. (2008).

In the multivariate normal case note that the cumulants and moments of Q are immediately

available from section 2. In particular we have

$$k_1 = \theta_1 = tr(\mathbf{\Sigma})$$

$$k_2 = 2\theta_2 = 2tr(\mathbf{\Sigma}^2)$$

$$k_3 = 8\theta_3 = 8tr(\mathbf{\Sigma}^3).$$

As $\theta_1 \to \infty$, $N\bar{\boldsymbol{X}}'\bar{\boldsymbol{X}}$ converges to a normal distribution. We discussed some properties of the limiting distribution of quadratic forms in normal random variables in the previous chapter. We discussed the convergence to normality of the quadratic form $Q = \sum_{i=1}^{p} \lambda_i X_i^2$, where X_i^2 is a chi-square distribution with n_i degrees of freedom. We followed the method for the more general case of a linear combination of independent gamma variates with different parameters as in Mathai and Provost (1992b). In the case of $N\bar{\boldsymbol{X}}'\bar{\boldsymbol{X}}$ we have a similar situation. By Mann-Wald theorem $\left(\frac{N\bar{\boldsymbol{X}}'\bar{\boldsymbol{X}}}{\theta_1}\right)^h$ also converges to a normal distribution. For testing the hypothesis $H_0 : \boldsymbol{\mu} = \mathbf{0}$, we propose the statistic:

$$T = \left(\frac{N\bar{\boldsymbol{x}}'\bar{\boldsymbol{x}}}{tr\boldsymbol{S}}\right)^h. \tag{4.21}$$

where $\boldsymbol{S}$ is computed as 4.18.
Since $\left(\frac{N\bar{\boldsymbol{X}}'\bar{\boldsymbol{X}}}{\theta_1}\right)^h$ converges to a normal distribution as $\theta_1 \to \infty$ while $\frac{N\bar{\boldsymbol{X}}'\bar{\boldsymbol{X}}}{tr\boldsymbol{S}}$ behaves like $\frac{\chi_f^2}{f}$ for large N and p, and since $\frac{tr\mathbf{\Sigma}}{tr\boldsymbol{S}} \xrightarrow{p} 1$, we write:

$$\left(\frac{N\bar{\boldsymbol{X}}'\bar{\boldsymbol{X}}}{tr\boldsymbol{S}}\right)^h = \left(\frac{tr\mathbf{\Sigma}}{tr\boldsymbol{S}}\right)^h \left(\frac{N\bar{\boldsymbol{X}}'\bar{\boldsymbol{X}}}{tr\mathbf{\Sigma}}\right)^h,$$

and by use of the Slutsky's theorem, one can conclude that $\left(\frac{N\bar{\boldsymbol{X}}'\bar{\boldsymbol{X}}}{tr\boldsymbol{S}}\right)^h$ also converges to a normal distribution as $\theta_1 \to \infty$.
To choose h we use the fact that $\frac{N\bar{\boldsymbol{X}}'\bar{\boldsymbol{X}}}{tr\boldsymbol{S}}$ behaves like $\frac{\chi_f^2}{f}$. For a χ_f^2 distribution we have

$$\begin{aligned} k_1 &= f \\ k_2 &= 2f \\ k_3 &= 8f. \end{aligned}$$

Hence $h_0 = 1/3$, as we noted earlier. We summarize the approximation as follows:

Under the null hypothesis $H_0 : \boldsymbol{\mu} = \mathbf{0}$, the statistic $\left(\frac{N\bar{\boldsymbol{x}}'\bar{\boldsymbol{x}}}{tr\boldsymbol{S}}\right)^{1/3}$ is approximately normally distributed with mean $\mu'(h_0) = 1-\frac{2}{9f}$ and variance $\mu_2(h_0) = \frac{2}{9f}$, where $f = [tr(\boldsymbol{\Sigma})]^2/tr(\boldsymbol{\Sigma}^2)$.

The quantity f is estimated by $\frac{E_2}{E_3}$, where E_2 and E_3 are computed as in relation (4.20). Our test is clearly invariant under the transformation $\boldsymbol{x} \rightarrow c\Gamma\boldsymbol{x}$, $c \neq 0$, $\Gamma\Gamma' = \boldsymbol{I}_p$, $i = 1, ..., N$. The performance of the test is evaluated numerically and compared with other tests already available in the literature, see the following.

4.5 Other Tests for the Mean

Srivastava and Du (2008) proposed a test statistic for the hypothesis

$$H_0 : \boldsymbol{\mu} = \mathbf{0} \quad vs. \quad H_a : \boldsymbol{\mu} \neq \mathbf{0} \tag{4.22}$$

when the sample size N is smaller than or equal to the dimension p, that is $N \leq p$. Let $n = N - 1$ and

$$\bar{\boldsymbol{x}} = N^{-1}\sum_{i=1}^{N}\boldsymbol{x}_i \qquad \boldsymbol{S} = n^{-1}\sum_{i=1}^{N}(\boldsymbol{x}_i - \bar{\boldsymbol{x}})(\boldsymbol{x}_i - \bar{\boldsymbol{x}})' = (s_{ij}). \tag{4.23}$$

Also, let $\boldsymbol{D}_s$ =diag$(s_{11}, ..., s_{pp})$. The test statistics T_1 is

$$T_1 = \frac{N\bar{\boldsymbol{x}}'\boldsymbol{D}_s^{-1}\bar{\boldsymbol{x}} - \frac{np}{n-2}}{\sqrt{2(tr\boldsymbol{R}^2 - \frac{p^2}{n})c_{p,n}}} \tag{4.24}$$

where

$$\boldsymbol{R} = \boldsymbol{D}_s^{-\frac{1}{2}}\boldsymbol{S}\boldsymbol{D}_s^{-\frac{1}{2}} = (r_{ij}) \tag{4.25}$$

denotes the sample correlation matrix, and $c_{p,n}$ is an adjustment coefficient such that

$$c_{p,n} \xrightarrow{p} 1 \qquad as \quad (n,p) \rightarrow \infty. \tag{4.26}$$

One particular choice of $c_{p,n}$ proposed by Srivastava and Du (2008) is

$$c_{p,n} = 1 + \frac{tr\boldsymbol{R}^2}{p^{3/2}}. \tag{4.27}$$

The statistic T_1 is an invariant test statistic under the group of scalar transformation where $\boldsymbol{x} \to \boldsymbol{D}\boldsymbol{x}$, and is asymptotically the standard normal $N(0,1)$. The assumptions needed to prove the asymptotic behavior of T_1 are as follows

Assumption 1:

$$0 < \lim_{p\to\infty} \frac{tr\mathcal{R}^i}{p} < \infty \quad , \quad i = 1,2,3,4,$$

Where

$$\mathcal{R} = \boldsymbol{D}_\sigma^{-\frac{1}{2}} \boldsymbol{\Sigma} \boldsymbol{D}_\sigma^{-\frac{1}{2}} = (\rho_{ij}). \tag{4.28}$$

Assumption 2:

$$\lim_{p\to\infty} \max_{1\le i\le p} \frac{\lambda_{ip}}{\sqrt{p}} = 0$$

where λ_{ip}'s ,$i = 1,...,p$, are eigenvalues of the matrix $\mathcal{R}$ as introduced above.

Assumption 3:

$$n = O(p^\xi), \quad \frac{1}{2} < \xi \le 1.$$

Under these assumptions and the null hypothesis $H_0 : \boldsymbol{\mu} = \boldsymbol{0}$ we have:

$$\lim_{(n,p)\to\infty} P_0(T_1 \le z_{1-\alpha}) = \Phi(z_{1-\alpha}) \tag{4.29}$$

where P_0 denotes that the probability is being computed under the null hypothesis, and $\Phi(.)$ is the standard normal cumulative distribution function. For details of proofs see Srivastava and Du (2008).

Another competing test is proposed by Dempster (1958) and Dempster (1960). The test statistic is also based on

$$T_2 = \frac{N\bar{\boldsymbol{x}}'\bar{\boldsymbol{x}}}{tr\boldsymbol{S}} \tag{4.30}$$

where $\bar{\boldsymbol{x}}$ is the sample mean, and $\boldsymbol{S}$ denotes sample variance as in (4.23). Its distribution is different from ours. Under the null hypothesis $H_0 : \boldsymbol{\mu} = \mathbf{0}$ and using orthogonal transformation it can be shown that

$$T_2 = \frac{NQ_1}{Q_2 + ... + Q_N} \tag{4.31}$$

where the Q_i's are independent and identically distributed. In fact Dempster (1958) made the assumption that each $Q_i,\ i = 1, ..., N$, is approximately distributed as $m\chi^2_r$, where $m > 0$ and χ^2_r is a chi-square distribution with r degrees of freedom. By matching the first and second moment of Q_i and $m\chi^2_r$, Dempster obtained:

$$r = \frac{(tr\boldsymbol{\Sigma})^2}{tr\boldsymbol{\Sigma}^2} = p\frac{a_1^2}{a_2} \tag{4.32}$$

where $a_1 = tr\boldsymbol{\Sigma}/p$, and $a_2 = tr\boldsymbol{\Sigma}^2/p$. The estimators for a_1, a_2, and r are as follows

$$\hat{a}_1 = \frac{tr\boldsymbol{S}}{p}, \quad \hat{a}_2 = \frac{n^2}{(n-1)(n+2)}\frac{1}{p}\left[tr\boldsymbol{S}^2 - \frac{(tr\boldsymbol{S})^2}{n}\right], \quad \text{and} \quad \hat{r} = p\frac{\hat{a}_1^2}{\hat{a}_2} \tag{4.33}$$

It is proposed that T_2 is approximately distributed as F with $[\hat{r}]$ and $[n\hat{r}]$ as degrees of freedom, where $[x]$ denotes the largest number less than or equal to x.

Ahmad et al. (2013) proposed the test statistics T_3 as is below for the mean vector of high dimensional multivariate data.

$$T_3 = \frac{\frac{N\bar{\boldsymbol{x}}'\bar{\boldsymbol{x}}}{tr\boldsymbol{S}} - 1}{\sqrt{\frac{2}{f}}} \tag{4.34}$$

The test statistic is proposed to be approximated by a standard normal distribution as $N \to \infty$. Here $f = \frac{[tr\boldsymbol{\Sigma}]^2}{tr(\boldsymbol{\Sigma}^2)}$ is estimated by $\frac{E_2}{E_3}$, where E_2 and E_3 are as in equations (4.20).

Finally, a fourth test statistic is proposed by Bai and Saranadasa (1996) for testing the hypothesis $\boldsymbol{\mu} = \mathbf{0}$. Their test statistic is

$$T_4 = \frac{N\bar{\boldsymbol{x}}'\bar{\boldsymbol{x}} - tr\boldsymbol{S}}{\left[\frac{2n(n+1)}{(n-1)(n+2)}\left(tr\boldsymbol{S}^2 - \frac{(tr\boldsymbol{S})^2}{N}\right)\right]^{\frac{1}{2}}}. \tag{4.35}$$

Bai and Saranadasa (1996) showed that under the null hypothesis $\boldsymbol{\mu} = \mathbf{0}$, T_4 is asymptotically distributed as $N(0, 1)$. They have also shown that T_2 and T_4 have the same power

and is as fallows

$$\beta(T_2|\boldsymbol{\mu}) \simeq \beta(T_4|\boldsymbol{\mu}) \simeq \Phi\left(-Z_{1-\alpha} + \frac{N\boldsymbol{\mu}'}{\sqrt{2tr\boldsymbol{\Sigma}^2}}\right). \tag{4.36}$$

4.6 Numerical Comparison

First, since T_2 and T_4 have the same asymptotic power behavior as discussed in Srivastava and Du (2008), we just compare our test T with T_1, and T_2 and T_3. A comparison of attained significance levels of the our test T and the alternative tests T_1, T_2,and T_3 is presented in the following simulation tables. To compare the power of the proposed tests we rely on a simulation study under the same correlation structure and covariance matrices as in Srivastava and Du (2008). We consider the independence correlation structure $\mathcal{R} = \mathcal{I}_p$ =diag$(1, 1, ..., 1)$ and the equal correlation $\mathcal{R} = \mathcal{R}_1 = (\rho_{ij}) \quad \rho_{ij} = 0.25, \quad i \neq j$. We also simulate the tests for various scalar matrices as in Srivastava and Du (2008). We select $\mathcal{D}_\sigma = \mathcal{I}_p$, $\mathcal{D}_\sigma = \mathcal{D}_{\sigma,1} : \sigma_{11}^{\frac{1}{2}}, ..., \sigma_{pp}^{\frac{1}{2}} \overset{iid}{\sim}$ unif$(2, 3)$ and $\mathcal{D}_\sigma = \mathcal{D}_{\sigma,2} : \sigma_{11}, ..., \sigma_{pp} \overset{iid}{\sim} \chi_3^2$. For the alternative hypothesis we choose $\boldsymbol{\mu} = (\boldsymbol{\mu}_1, ..., \boldsymbol{\mu}_p)' : \boldsymbol{\mu}_{2k-1} = 0$ and $\boldsymbol{\mu}_{2k} \overset{iid}{\sim} Unif(-\frac{1}{2}, \frac{1}{2}), k = 1, ..., \frac{p}{2}$.

Under the null hypothesis the attained significance level is computed as

$$\hat{\alpha} = \frac{(\#t_H \geq z_{1-\alpha})}{m} \tag{4.37}$$

where t_H is denotes the value of the test statistics T, T_1, T_2, and T_3. $z_{1-\alpha}$ is the $\%100(1-\alpha)$ quantile of $N(0, 1)$, and m is the number of replications. To compute the empirical powers we obtain empirical critical values. We choose $(m\alpha)$th largest value of the test statistics as the empirical critical values. Then, another m replications of the data will be simulated under the alternative hypothesis. The empirical power is calculated as

$$\hat{\beta} = \frac{(\#t_A \geq \hat{z}_{1-\alpha})}{m} \tag{4.38}$$

where t_A is the value of the test statistics, obtained under the alternative hypothesis, and m is the number of replications. In the simulation study, we choose $m = 10,000$ as the number

of replications and fix the nominal significance level $\alpha = 0.05$. Hence, under the null hypothesis, the attained significance level $\hat{\alpha}$ is approximately distributed as $Binomial(10000, 0.05)$. Hence, an estimate of the standard error of $\hat{\alpha}$ is $\hat{se}(\hat{\alpha}) = \sqrt{0.05 \times 0.95/10,000} \simeq 0.0022$.

The attained significance levels and empirical powers for $\mathcal{R} = I_p$ and $\mathcal{R} = \mathcal{R}_1$ and different scalar matrices are presented in the following tables. As it is obvious from the tables, the attained significance levels of our proposed test T are around the nominal level 0.05 for most N's and p's. Our test compares favorably or is better than T_1, T_2, and T_3 that have attained alphas greater than the nominal level. Regarding the power of the tests, tables 3 and 4 present the empirical powers of the tests for $\mathcal{R} = I_p$ and $\mathcal{R} = \mathcal{R}_1$ respectively. Our test is again as good as or better than the other tests.

Table 4.1: Attained significance levels of T, T_1, T_2, and T_3, $\mathcal{R} = I_p$

		$\mathcal{R} = \mathcal{R}_p$											
		$\mathcal{D}_\sigma = I_p$				$\mathcal{D}_\sigma = \mathcal{D}_{\sigma,1}$				$\mathcal{D}_\sigma = \mathcal{D}_{\sigma,2}$			
p	N	T	T1	T2	T3	T	T1	T2	T3	T	T1	T2	T3
10	5	0.0277	0.1670	0.0669	0.0473	0.0257	0.1671	0.0643	0.0455	0.0325	0.1663	0.0781	0.0560
	9	0.0441	0.1159	0.0559	0.0615	0.0372	0.1130	0.0506	0.0568	0.0430	0.1095	0.0612	0.0639
20	5	0.0262	0.1799	0.0638	0.0399	0.0257	0.1838	0.0673	0.0402	0.0270	0.1779	0.0670	0.0449
	10	0.0433	0.1025	0.0545	0.0571	0.0433	0.1082	0.0548	0.0566	0.0508	0.1092	0.0664	0.0649
	15	0.0496	0.0891	0.0569	0.0643	0.0398	0.0801	0.0457	0.0523	0.0498	0.0853	0.0558	0.0643
30	5	0.0281	0.1833	0.0684	0.0413	0.0249	0.1853	0.0650	0.0393	0.0288	0.1872	0.0765	0.0444
	10	0.0394	0.1048	0.0517	0.0495	0.0416	0.1000	0.0524	0.0521	0.0452	0.1069	0.0577	0.0582
	20	0.0463	0.0678	0.0505	0.0566	0.0517	0.0751	0.0548	0.0635	0.0462	0.0744	0.0508	0.0605
	25	0.0469	0.0657	0.0501	0.0585	0.0482	0.0655	0.0519	0.0608	0.0506	0.0618	0.0531	0.0636
40	5	0.0271	0.1902	0.0621	0.0384	0.0263	0.1863	0.0657	0.0392	0.0308	0.1933	0.0733	0.0448
	10	0.0448	0.1055	0.0558	0.0555	0.0406	0.0991	0.0521	0.0511	0.0442	0.1043	0.0589	0.0577
	20	0.0445	0.0693	0.0494	0.0545	0.0464	0.0694	0.0501	0.0568	0.0489	0.0658	0.0527	0.0598
	30	0.0480	0.0616	0.0494	0.0569	0.0479	0.0597	0.0505	0.0598	0.0516	0.0614	0.0541	0.0630
	35	0.0461	0.0578	0.0479	0.0568	0.0485	0.0578	0.0503	0.0593	0.0467	0.0566	0.0493	0.0594
50	5	0.0299	0.1878	0.0632	0.0385	0.0282	0.1870	0.0642	0.0372	0.0332	0.1878	0.0740	0.0441
	10	0.0415	0.1008	0.0537	0.0509	0.0414	0.0992	0.0551	0.0502	0.0416	0.1015	0.0560	0.0514
	20	0.0444	0.0671	0.0482	0.0536	0.0471	0.0656	0.0508	0.0561	0.0477	0.0630	0.0526	0.0600
	40	0.0501	0.0580	0.0522	0.0593	0.0460	0.0558	0.0474	0.0559	0.0492	0.0535	0.0505	0.0605
	45	0.0455	0.0526	0.0463	0.0544	0.0477	0.0523	0.0484	0.0560	0.0515	0.0541	0.0528	0.0637
60	5	0.0290	0.1931	0.0690	0.0386	0.0285	0.1898	0.0663	0.0378	0.0293	0.1865	0.0709	0.0408
	10	0.0420	0.0969	0.0526	0.0504	0.0393	0.0943	0.0493	0.0469	0.0444	0.0944	0.0578	0.0542
	20	0.0443	0.0618	0.0476	0.0524	0.0465	0.0650	0.0504	0.0540	0.0460	0.0639	0.0513	0.0569
	40	0.0481	0.0540	0.0498	0.0563	0.0505	0.0580	0.0519	0.0583	0.0499	0.0539	0.0518	0.0600
	50	0.0497	0.0555	0.0510	0.0593	0.0477	0.0532	0.0489	0.0572	0.0564	0.0551	0.0574	0.0656
	55	0.0500	0.0542	0.0512	0.0606	0.0504	0.0520	0.0514	0.0591	0.0513	0.0515	0.0525	0.0629
70	5	0.0267	0.1848	0.0649	0.0358	0.0312	0.1880	0.0705	0.0400	0.0307	0.1918	0.0734	0.0427
	10	0.0407	0.0936	0.0527	0.0490	0.0395	0.0936	0.0499	0.0466	0.0460	0.0903	0.0583	0.0539
	20	0.0457	0.0636	0.0505	0.0545	0.0432	0.0622	0.0491	0.0525	0.0475	0.0618	0.0521	0.0573
	40	0.0512	0.0559	0.0522	0.0576	0.0480	0.0549	0.0499	0.0572	0.0494	0.0540	0.0516	0.0588
	60	0.0469	0.0493	0.0476	0.0552	0.0469	0.0486	0.0475	0.0537	0.0520	0.0544	0.0539	0.0632
	65	0.0520	0.0549	0.0527	0.0606	0.0452	0.0491	0.0465	0.0531	0.0508	0.0525	0.0520	0.0599
80	5	0.0290	0.1899	0.0647	0.0370	0.0312	0.1880	0.0687	0.0392	0.0329	0.1894	0.0721	0.0428
	10	0.0402	0.0902	0.0508	0.0490	0.0428	0.0893	0.0565	0.0517	0.0419	0.0918	0.0558	0.0506
	20	0.0445	0.0607	0.0484	0.0523	0.0486	0.0621	0.0532	0.0552	0.0491	0.0588	0.0558	0.0590
	40	0.0472	0.0501	0.0488	0.0541	0.0500	0.0538	0.0514	0.0582	0.0486	0.0531	0.0505	0.0577
	60	0.0495	0.0536	0.0504	0.0588	0.0496	0.0529	0.0506	0.0567	0.0514	0.0500	0.0517	0.0593
	75	0.0474	0.0473	0.0480	0.0555	0.0523	0.0566	0.0528	0.0606	0.0528	0.0471	0.0532	0.0607
90	5	0.0261	0.1898	0.0626	0.0327	0.0255	0.1944	0.0635	0.0333	0.0301	0.1932	0.0709	0.0391
	10	0.0428	0.0900	0.0561	0.0494	0.0423	0.0902	0.0551	0.0501	0.0440	0.0905	0.0565	0.0527
	20	0.0434	0.0590	0.0488	0.0507	0.0465	0.0599	0.0511	0.0543	0.0466	0.0574	0.0521	0.0558
	40	0.0426	0.0491	0.0443	0.049	0.0489	0.0524	0.0511	0.0575	0.0483	0.0476	0.0505	0.0567
	60	0.0464	0.0477	0.0483	0.0535	0.0473	0.0491	0.0481	0.0542	0.0512	0.0519	0.0527	0.0606
	80	0.0501	0.0532	0.0506	0.0576	0.0464	0.0513	0.0474	0.0553	0.0502	0.0496	0.0507	0.0604
	85	0.0517	0.0512	0.0522	0.0594	0.0516	0.0527	0.0523	0.0578	0.0498	0.0488	0.0504	0.0586
100	5	0.0311	0.1888	0.0678	0.0381	0.0300	0.1915	0.0679	0.0378	0.0347	0.1927	0.0724	0.0441
	10	0.0430	0.0867	0.0563	0.0499	0.0410	0.0884	0.0549	0.0491	0.0467	0.0828	0.0608	0.0555
	20	0.0496	0.0609	0.0559	0.0575	0.0483	0.0568	0.0528	0.0542	0.0512	0.0572	0.0552	0.0590
	40	0.0519	0.0544	0.0533	0.0587	0.0434	0.0483	0.0458	0.0517	0.0476	0.0497	0.0509	0.0559
	60	0.0479	0.0503	0.0492	0.0554	0.0477	0.0474	0.0490	0.0549	0.0492	0.0469	0.0509	0.0576
	80	0.0490	0.0496	0.0497	0.0585	0.0494	0.0496	0.0503	0.0573	0.0512	0.0523	0.0520	0.0601
	90	0.0532	0.0534	0.0543	0.0624	0.0518	0.0515	0.0522	0.0599	0.0456	0.0464	0.0465	0.0538
	95	0.0474	0.0508	0.0480	0.0542	0.0497	0.0509	0.0500	0.0562	0.0516	0.0483	0.0522	0.0601

Table 4.2: Attained significance levels of T, T_1, T_2, and T_3, $\mathcal{R} = \mathcal{R}_1$

		$\mathcal{R} = \mathcal{R}_1$											
		$\mathcal{D}_\sigma = I_p$				$\mathcal{D}_\sigma = \mathcal{D}_{\sigma,1}$				$\mathcal{D}_\sigma = \mathcal{D}_{\sigma,2}$			
p	N	T	T1	T2	T3	T	T1	T2	T3	T	T1	T2	T3
10	5	0.0306	0.1632	0.0783	0.0523	0.0295	0.1632	0.0792	0.0575	0.0321	0.1669	0.0809	0.0566
	9	0.0500	0.1110	0.0669	0.0666	0.0445	0.1087	0.0637	0.0692	0.0422	0.1098	0.0602	0.0616
20	5	0.0361	0.1734	0.0852	0.0534	0.0366	0.1849	0.0920	0.0607	0.0342	0.1682	0.0866	0.0555
	10	0.0477	0.1022	0.0674	0.0627	0.0495	0.1008	0.0673	0.0852	0.0512	0.1088	0.0703	0.0674
	15	0.0493	0.0767	0.0606	0.0643	0.0539	0.0821	0.0655	0.0878	0.0517	0.0809	0.0628	0.0676
30	5	0.0391	0.1833	0.0950	0.0552	0.0384	0.1786	0.0970	0.0608	0.0369	0.1760	0.0907	0.0557
	10	0.0546	0.1050	0.0776	0.0697	0.0508	0.1010	0.0732	0.0802	0.0538	0.1024	0.0726	0.0687
	20	0.0534	0.0682	0.0629	0.0661	0.0547	0.0718	0.0639	0.0997	0.0487	0.0671	0.0562	0.0605
	25	0.0512	0.0611	0.0562	0.0630	0.0535	0.0628	0.0594	0.1100	0.0533	0.0664	0.0580	0.0674
40	5	0.0394	0.1831	0.0977	0.0573	0.0393	0.1805	0.1029	0.0637	0.0388	0.1761	0.0941	0.0555
	10	0.0529	0.0964	0.0743	0.0642	0.0498	0.0933	0.0736	0.0768	0.0490	0.0948	0.0700	0.0626
	20	0.0502	0.0627	0.0595	0.0627	0.0565	0.0721	0.0668	0.1063	0.0519	0.0709	0.0607	0.0647
	30	0.0527	0.0586	0.0580	0.0634	0.0534	0.0578	0.0586	0.1375	0.0552	0.0609	0.0601	0.0671
	35	0.0539	0.0571	0.0580	0.0656	0.0517	0.0560	0.0554	0.1138	0.0531	0.0567	0.0569	0.0659
50	5	0.0458	0.1800	0.1066	0.0614	0.0407	0.1789	0.1000	0.0691	0.0378	0.1734	0.0981	0.0558
	10	0.0579	0.0976	0.0805	0.0703	0.0566	0.0986	0.0820	0.0834	0.0518	0.0968	0.0747	0.0659
	20	0.0583	0.0687	0.0671	0.0673	0.0548	0.0653	0.0652	0.1094	0.0552	0.0673	0.0641	0.0676
	40	0.0564	0.0552	0.0604	0.0672	0.0552	0.0553	0.0593	0.1925	0.0564	0.0584	0.0595	0.0683
	45	0.0555	0.0557	0.0594	0.0666	0.0563	0.0545	0.0605	0.1439	0.0541	0.0520	0.0564	0.0664
60	5	0.0459	0.1761	0.1030	0.0602	0.0470	0.1858	0.1108	0.0753	0.0438	0.1834	0.1038	0.0634
	10	0.0558	0.0984	0.0821	0.0682	0.0604	0.0966	0.0812	0.0886	0.0547	0.0988	0.0763	0.0691
	20	0.0572	0.0670	0.0674	0.0679	0.0592	0.0714	0.0686	0.0968	0.0567	0.0689	0.0652	0.0675
	40	0.0533	0.0531	0.0574	0.0645	0.0541	0.0532	0.0575	0.1338	0.0551	0.0568	0.0593	0.0671
	50	0.0552	0.0523	0.0587	0.0670	0.0549	0.0519	0.0576	0.1919	0.0593	0.0538	0.0621	0.0738
	55	0.0548	0.0505	0.0566	0.0642	0.0564	0.0526	0.0576	0.1430	0.0553	0.0543	0.0570	0.0678
70	5	0.0430	0.1733	0.1034	0.0585	0.0458	0.1846	0.1100	0.0716	0.0427	0.1873	0.1047	0.0602
	10	0.0582	0.0940	0.0820	0.0703	0.0599	0.0990	0.0878	0.0862	0.0528	0.0948	0.0762	0.0662
	20	0.0571	0.0640	0.0662	0.0675	0.0566	0.0675	0.0673	0.1009	0.0577	0.0661	0.0671	0.0694
	40	0.0579	0.0561	0.0616	0.0694	0.0542	0.0516	0.0579	0.1177	0.0548	0.0530	0.0587	0.0654
	60	0.0535	0.0471	0.0557	0.0643	0.0532	0.0487	0.0557	0.1903	0.0548	0.0517	0.0564	0.0683
	65	0.0552	0.0491	0.0576	0.0653	0.0560	0.0503	0.0583	0.2708	0.0560	0.0510	0.0575	0.0663
80	5	0.0500	0.1815	0.1140	0.0648	0.0427	0.1730	0.1089	0.0709	0.0439	0.1834	0.1027	0.0599
	10	0.0575	0.0895	0.0834	0.0692	0.0531	0.0845	0.0769	0.0918	0.0523	0.0932	0.0778	0.0653
	20	0.0582	0.0669	0.0692	0.0692	0.0581	0.0655	0.0695	0.0988	0.0556	0.0658	0.0661	0.0671
	40	0.0601	0.0568	0.0651	0.0710	0.0538	0.0533	0.0587	0.1125	0.0539	0.0533	0.0580	0.0658
	60	0.0601	0.0530	0.0621	0.0694	0.0532	0.0470	0.0555	0.2844	0.0564	0.0488	0.0582	0.0680
	75	0.0590	0.0504	0.0602	0.0697	0.0576	0.0505	0.0595	0.2231	0.0530	0.0473	0.0548	0.0655
90	5	0.0490	0.1773	0.1155	0.0644	0.0462	0.1741	0.1107	0.0703	0.0444	0.1800	0.1077	0.0620
	10	0.0569	0.0936	0.0836	0.0692	0.0514	0.0874	0.0788	0.0852	0.0558	0.0948	0.0792	0.0680
	20	0.0568	0.0629	0.0676	0.0671	0.0585	0.0648	0.0683	0.1021	0.0554	0.0631	0.0678	0.0677
	40	0.0563	0.0526	0.0616	0.0668	0.0532	0.0515	0.0587	0.1492	0.0536	0.0494	0.0571	0.0625
	60	0.0522	0.0463	0.0554	0.0638	0.0557	0.0486	0.0586	0.2233	0.0549	0.0491	0.0571	0.0653
	80	0.0557	0.0471	0.0566	0.0676	0.0573	0.0484	0.0585	0.3228	0.0514	0.0468	0.0523	0.0616
	85	0.0553	0.0465	0.0566	0.0671	0.0540	0.0461	0.0551	0.4147	0.0539	0.0472	0.0543	0.0637
100	5	0.0463	0.1763	0.1146	0.0613	0.0459	0.1813	0.1135	0.0748	0.0426	0.1752	0.1040	0.0590
	10	0.0594	0.0938	0.0864	0.0710	0.0621	0.0960	0.0876	0.0845	0.0589	0.0949	0.0831	0.0702
	20	0.0583	0.0631	0.0694	0.0688	0.0537	0.0606	0.0661	0.0971	0.0568	0.0628	0.0676	0.0689
	40	0.0602	0.0543	0.0649	0.0700	0.0579	0.0535	0.0636	0.1731	0.0542	0.0521	0.0577	0.0640
	60	0.0611	0.0521	0.0633	0.0728	0.0537	0.0467	0.0551	0.2520	0.0565	0.0484	0.0583	0.0675
	80	0.0539	0.0452	0.0553	0.0646	0.0572	0.0463	0.0582	0.3064	0.0552	0.0451	0.0559	0.0651
	90	0.0600	0.0476	0.0602	0.0703	0.0573	0.0474	0.0580	0.2827	0.0543	0.0441	0.0548	0.0645
	95	0.0560	0.0472	0.0568	0.0657	0.0544	0.0454	0.0551	0.6957	0.0575	0.0478	0.0579	0.0671

Table 4.3: Empirical powers of the tests T, T_1, T_2, and T_3, $\mathcal{R} = I_p$

		$\mathcal{R} = \mathcal{R}_p$											
		$\mathcal{D}_\sigma = I_p$				$\mathcal{D}_\sigma = \mathcal{D}_{\sigma,1}$				$\mathcal{D}_\sigma = \mathcal{D}_{\sigma,2}$			
p	N	T	T1	T2	T3	T	T1	T2	T3	T	T1	T2	T3
10	5	0.0893	0.0721	0.0935	0.0838	0.0557	0.0544	0.0571	0.0516	0.0634	0.0605	0.0639	0.0695
	9	0.3039	0.2334	0.3064	0.3419	0.0721	0.0709	0.0717	0.0793	0.0581	0.0643	0.0552	0.0726
20	5	0.1650	0.1009	0.1777	0.1411	0.0612	0.0478	0.0608	0.0496	0.0749	0.1107	0.0736	0.0673
	10	0.2213	0.1931	0.2279	0.2379	0.0797	0.0779	0.0797	0.0885	0.0798	0.1726	0.0752	0.1080
	15	0.2498	0.2217	0.2577	0.2900	0.0963	0.0871	0.0942	0.0992	0.0872	0.1913	0.0890	0.1133
30	5	0.1659	0.1016	0.1832	0.1436	0.0700	0.0594	0.0703	0.0565	0.0711	0.0724	0.0638	0.0608
	10	0.3114	0.2301	0.3093	0.3098	0.0711	0.0693	0.0721	0.0745	0.0879	0.1105	0.0866	0.1017
	20	0.8520	0.8185	0.8533	0.8680	0.1123	0.1191	0.1159	0.1393	0.2822	0.8647	0.2828	0.3146
	25	0.8561	0.8256	0.8588	0.8724	0.1411	0.1397	0.1383	0.1627	0.1907	0.5765	0.1850	0.2292
40	5	0.1517	0.0845	0.1505	0.1198	0.0667	0.0523	0.0642	0.0512	0.0741	0.1717	0.0747	0.0672
	10	0.3262	0.2593	0.3348	0.3462	0.0973	0.0879	0.0968	0.0983	0.0898	0.1133	0.0910	0.1025
	20	0.8787	0.8350	0.8808	0.8880	0.1271	0.1143	0.1298	0.1404	0.2222	0.3892	0.2154	0.2510
	30	0.9245	0.9069	0.9255	0.9332	0.1867	0.1954	0.1872	0.2075	0.2226	0.7088	0.2224	0.2617
	35	0.9988	0.9979	0.9987	0.9988	0.1776	0.1816	0.1777	0.2009	0.5448	1.0000	0.5406	0.5855
50	5	0.1468	0.0890	0.1620	0.1203	0.0726	0.0649	0.0749	0.0552	0.0693	0.0692	0.0662	0.0595
	10	0.4039	0.3124	0.4159	0.4082	0.0988	0.0841	0.0955	0.0993	0.1343	0.1969	0.1329	0.1381
	20	0.9708	0.9513	0.9717	0.9724	0.1635	0.1675	0.1632	0.1762	0.1898	0.7112	0.1917	0.2135
	40	0.9995	0.9995	0.9995	0.9998	0.2893	0.2906	0.2903	0.3092	0.6187	0.9954	0.6157	0.6567
	45	1.0000	1.0000	1.0000	1.0000	0.5201	0.5775	0.5219	0.5458	0.5646	1.0000	0.5632	0.6155
60	5	0.2206	0.1076	0.2409	0.1867	0.0648	0.0589	0.0661	0.0511	0.0994	0.1226	0.0959	0.0871
	10	0.4905	0.3723	0.5025	0.4930	0.1052	0.0919	0.1016	0.0982	0.1074	0.3889	0.1065	0.1180
	20	0.9662	0.9458	0.9668	0.9679	0.1344	0.1392	0.1348	0.1452	0.3893	0.9573	0.3953	0.4112
	40	1.0000	1.0000	1.0000	1.0000	0.3925	0.3976	0.3925	0.4228	0.6849	0.9979	0.6813	0.7256
	50	1.0000	1.0000	1.0000	1.0000	0.3833	0.4113	0.3788	0.4041	0.5620	0.9770	0.5599	0.6318
	55	1.0000	1.0000	1.0000	1.0000	0.6224	0.7209	0.6246	0.6578	0.8707	0.9444	0.8719	0.8925
70	5	0.1792	0.1029	0.1902	0.1406	0.0680	0.0600	0.0728	0.0546	0.0824	0.1547	0.0834	0.0695
	10	0.4167	0.3141	0.4268	0.4131	0.1124	0.0914	0.1092	0.1070	0.1292	0.3888	0.1298	0.1391
	20	0.9806	0.9641	0.9809	0.9824	0.1981	0.1887	0.1998	0.2029	0.3623	0.9997	0.3617	0.3954
	40	1.0000	1.0000	1.0000	1.0000	0.3933	0.3887	0.3918	0.4141	0.4307	0.9831	0.4245	0.4678
	60	1.0000	1.0000	1.0000	1.0000	0.5603	0.6073	0.5602	0.5748	0.8407	0.9993	0.8389	0.8711
	65	1.0000	1.0000	1.0000	1.0000	0.7875	0.8264	0.7851	0.7962	0.9913	1.0000	0.9912	0.9937
80	5	0.2296	0.1057	0.2523	0.1829	0.0663	0.0547	0.0691	0.0528	0.0954	0.1405	0.0928	0.0812
	10	0.6954	0.5725	0.7051	0.6926	0.0958	0.0933	0.0988	0.0989	0.1558	0.3533	0.1571	0.1585
	20	0.9979	0.9940	0.9976	0.9995	0.1991	0.2062	0.1981	0.2162	0.3094	0.9622	0.3107	0.3408
	40	1.0000	1.0000	1.0000	1.0000	0.6154	0.6220	0.6172	0.6424	0.8241	0.9958	0.8231	0.8419
	60	1.0000	1.0000	1.0000	1.0000	0.6798	0.7080	0.6830	0.7035	0.9809	1.0000	0.9808	0.9869
	75	1.0000	1.0000	1.0000	1.0000	0.9134	0.9480	0.9139	0.9272	0.9986	1.0000	0.9984	0.9991
90	5	0.3351	0.1321	0.3652	0.2694	0.0659	0.0569	0.0663	0.0470	0.0929	0.1038	0.0882	0.0720
	10	0.5878	0.4632	0.5971	0.5875	0.1003	0.0904	0.1045	0.1005	0.1195	0.4301	0.1184	0.1239
	20	0.9964	0.9925	0.9967	0.9964	0.2172	0.1982	0.2144	0.2259	0.2115	0.8239	0.4379	0.2274
	40	1.0000	1.0000	1.0000	1.0000	0.4657	0.4903	0.4703	0.4925	0.7852	0.9985	0.8678	0.8041
	60	1.0000	1.0000	1.0000	1.0000	0.6795	0.7111	0.6798	0.6923	0.9991	1.0000	0.9961	0.9993
	80	1.0000	1.0000	1.0000	1.0000	0.7373	0.7389	0.7379	0.7493	0.9961	1.0000	0.9962	0.9972
	85	1.0000	1.0000	1.0000	1.0000	0.9085	0.9219	0.9058	0.9198	0.9986	1.0000	0.9986	0.9992
100	5	0.2575	0.1199	0.2794	0.2117	0.0665	0.0600	0.0677	0.0469	0.0988	0.1203	0.1038	0.0874
	10	0.7261	0.5913	0.7372	0.7264	0.1057	0.1008	0.1046	0.1038	0.1196	0.9432	0.1207	0.1299
	20	0.9693	0.9850	0.9928	0.9933	0.1870	0.1857	0.1874	0.1996	0.4090	0.9973	0.4064	0.4485
	40	0.9923	1.0000	1.0000	1.0000	0.5005	0.5286	0.5009	0.5069	0.7527	0.9989	0.7483	0.7734
	60	0.9990	1.0000	1.0000	1.0000	0.7996	0.8297	0.8003	0.8123	0.9732	1.0000	0.9740	0.9778
	80	1.0000	1.0000	1.0000	1.0000	0.6663	0.6914	0.6644	0.6901	0.9848	1.0000	0.9847	0.9884
	90	1.0000	1.0000	1.0000	1.0000	0.9678	0.9739	0.9675	0.9737	0.9976	1.0000	0.9975	0.9978
	95	1.0000	1.0000	1.0000	1.0000	0.9875	0.9951	0.9877	0.9898	1.0000	1.0000	1.0000	1.0000

Table 4.4: Empirical powers of the tests T, T_1, T_2, and T_3, $\mathcal{R} = \mathcal{R}_1$

		$\mathcal{R} = \mathcal{R}_1$											
		$\mathcal{D}_\sigma = I_p$				$\mathcal{D}_\sigma = \mathcal{D}_{\sigma,1}$				$\mathcal{D}_\sigma = \mathcal{D}_{\sigma,2}$			
p	N	T	T1	T2	T3	T	T1	T2	T3	T	T1	T2	T3
10	5	0.1229	0.0950	0.1253	0.1263	0.0559	0.0573	0.0512	0.0575	0.0496	0.0580	0.0510	0.0576
	9	0.1720	0.1408	0.1630	0.2231	0.0548	0.0547	0.0523	0.0692	0.0754	0.2035	0.0705	0.0918
20	5	0.1000	0.0777	0.0896	0.1051	0.0545	0.0507	0.0529	0.0607	0.0713	0.0713	0.0651	0.0784
	10	0.2837	0.2100	0.2458	0.3305	0.0675	0.0676	0.0649	0.0852	0.0664	0.1053	0.0656	0.0901
	15	0.1421	0.1295	0.1365	0.1710	0.0668	0.0696	0.0617	0.0878	0.0683	0.1213	0.0636	0.0933
30	5	0.1090	0.0820	0.0926	0.1190	0.0539	0.0559	0.0538	0.0608	0.0571	0.0563	0.0545	0.0656
	10	0.1619	0.1347	0.1371	0.2159	0.0617	0.0596	0.0609	0.0802	0.0720	0.2329	0.0694	0.0971
	20	0.3678	0.3132	0.3296	0.4474	0.0727	0.0732	0.0726	0.0997	0.1135	0.1822	0.1044	0.1385
	25	0.3763	0.3296	0.3586	0.4478	0.0842	0.0836	0.0826	0.1100	0.1478	0.8246	0.1358	0.1928
40	5	0.1074	0.0778	0.0953	0.1207	0.0557	0.0505	0.0477	0.0637	0.0659	0.0874	0.0637	0.0731
	10	0.2139	0.1674	0.1680	0.2693	0.0626	0.0646	0.0605	0.0768	0.0875	0.1460	0.0763	0.1043
	20	0.5145	0.4260	0.4778	0.5866	0.0789	0.0735	0.0723	0.1063	0.0828	0.2087	0.0803	0.1076
	30	0.6934	0.6125	0.6914	0.7794	0.1069	0.1148	0.1023	0.1375	0.1140	0.1778	0.1028	0.1538
	35	0.8906	0.8326	0.9110	0.9433	0.0925	0.0887	0.0849	0.1138	0.1660	0.2026	0.1560	0.2129
50	5	0.1374	0.0948	0.1088	0.1686	0.0629	0.0499	0.0571	0.0691	0.0580	0.0599	0.0584	0.0639
	10	0.1776	0.1427	0.1429	0.2402	0.0614	0.0598	0.0580	0.0834	0.0812	0.1172	0.0752	0.1062
	20	0.5555	0.4576	0.6405	0.6918	0.0834	0.0817	0.0753	0.1094	0.1160	0.2776	0.1038	0.1558
	40	0.9957	0.9856	0.9986	0.9992	0.1431	0.1382	0.1323	0.1925	0.4220	0.9669	0.3934	0.5380
	45	0.9972	0.9886	0.9984	0.9988	0.1090	0.1067	0.1015	0.1439	0.2700	0.9215	0.2574	0.3531
60	5	0.1100	0.0920	0.0885	0.1343	0.0606	0.0592	0.0546	0.0753	0.0647	0.0800	0.0603	0.0788
	10	0.2215	0.1730	0.1816	0.2910	0.0613	0.0643	0.0593	0.0886	0.0891	0.2196	0.0743	0.1200
	20	0.5357	0.4342	0.4780	0.6455	0.0673	0.0633	0.0609	0.0968	0.0836	0.2423	0.0801	0.1191
	40	0.9710	0.9319	0.9885	0.9875	0.1031	0.1014	0.0977	0.1338	0.2608	0.8561	0.2363	0.3458
	50	0.9984	0.9947	0.9998	0.9998	0.1462	0.1389	0.1396	0.1919	0.2193	0.8106	0.1915	0.3214
	55	0.9999	0.9999	1.0000	1.0000	0.1073	0.1099	0.1040	0.1430	0.4337	0.8203	0.4215	0.5516
70	5	0.1157	0.0892	0.0957	0.1311	0.0579	0.0524	0.0578	0.0716	0.0633	0.0782	0.0625	0.0766
	10	0.1630	0.1346	0.1190	0.2275	0.0614	0.0574	0.0592	0.0862	0.0959	0.1370	0.0763	0.1186
	20	0.4156	0.3410	0.3683	0.5191	0.0776	0.0715	0.0745	0.1009	0.1534	0.6383	0.1345	0.2091
	40	0.9685	0.9268	0.9823	0.9892	0.0905	0.0883	0.0843	0.1177	0.1714	0.6791	0.1578	0.2380
	60	1.0000	1.0000	1.0000	1.0000	0.1454	0.1536	0.1357	0.1903	0.4713	0.9999	0.4402	0.6050
	65	1.0000	1.0000	1.0000	1.0000	0.2013	0.1973	0.1864	0.2708	0.3435	0.9696	0.3184	0.4595
80	5	0.1184	0.0882	0.0884	0.1532	0.0628	0.0592	0.0581	0.0709	0.0601	0.0796	0.0577	0.0723
	10	0.3208	0.2351	0.2272	0.4008	0.0697	0.0692	0.0650	0.0918	0.0856	0.1739	0.0754	0.1107
	20	0.5154	0.4055	0.4557	0.6417	0.0734	0.0708	0.0678	0.0988	0.1106	0.5039	0.1024	0.1460
	40	0.9793	0.9464	0.9931	0.9955	0.0859	0.0785	0.0824	0.1125	0.2296	0.8207	0.2116	0.3038
	60	0.9992	0.9972	1.0000	1.0000	0.2224	0.2288	0.2004	0.2844	0.2513	0.9910	0.2284	0.3552
	75	1.0000	1.0000	1.0000	1.0000	0.1564	0.1480	0.1437	0.2231	0.4895	0.9631	0.4638	0.6067
90	5	0.1273	0.0868	0.0899	0.1590	0.0583	0.0554	0.0540	0.0703	0.0671	0.0616	0.0614	0.0805
	10	0.2285	0.1706	0.1535	0.2947	0.0686	0.0665	0.0623	0.0852	0.0791	0.1721	0.0703	0.1037
	20	0.4340	0.3476	0.3669	0.5460	0.0745	0.0718	0.0693	0.1021	0.1265	0.1897	0.1048	0.1642
	40	0.9912	0.9665	0.9979	0.9983	0.1155	0.1111	0.1085	0.1492	0.3146	0.9691	0.2925	0.4053
	60	1.0000	1.0000	1.0000	1.0000	0.1626	0.1973	0.1492	0.2233	0.5103	0.9996	0.4845	0.6391
	80	1.0000	1.0000	1.0000	1.0000	0.2384	0.2593	0.2193	0.3228	0.5920	1.0000	0.5888	0.7028
	85	1.0000	1.0000	1.0000	1.0000	0.3125	0.3257	0.3019	0.4147	0.8580	0.9999	0.8776	0.9323
100	5	0.1201	0.0817	0.0841	0.1412	0.0624	0.0548	0.0599	0.0748	0.0741	0.0918	0.0663	0.0877
	10	0.2369	0.1633	0.1582	0.3136	0.0548	0.0521	0.0567	0.0845	0.0646	0.1349	0.0603	0.0946
	20	0.6659	0.5233	0.6744	0.7873	0.0763	0.0745	0.0785	0.0971	0.1061	0.1926	0.0997	0.1436
	40	0.9683	0.9248	0.9859	0.9930	0.1271	0.1192	0.1134	0.1731	0.2722	1.0000	0.2335	0.3570
	60	1.0000	0.9999	1.0000	1.0000	0.1922	0.1923	0.1765	0.2520	0.5421	0.9992	0.5274	0.6819
	80	1.0000	1.0000	1.0000	1.0000	0.2218	0.2519	0.1979	0.3064	0.6709	0.9254	0.6620	0.7879
	90	1.0000	1.0000	1.0000	1.0000	0.2025	0.2290	0.1875	0.2827	0.8873	1.0000	0.9045	0.9508
	95	1.0000	1.0000	1.0000	1.0000	0.5619	0.5892	0.5404	0.6957	0.9912	1.0000	0.9959	0.9986

Chapter 5

R-code

The simulation was rendered using R programming. The following is a sample of the code used in the simulation in case of intraclass correlation structure and $D_\sigma = D_{\sigma,2}$ where $\mathcal{D}_\sigma = \mathcal{D}_{\sigma,2} : \sigma_{11}, ..., \sigma_{pp} \overset{iid}{\sim} \chi^2_3$.

```
library(numDeriv)
library(pscl)
library(MASS)
library(mvtnorm)
library(trust)
library(psych)



repetition=10^4
alpha=0.05
P_N = list()      (List of P's_and_N's)
T1_alpha=list()      (List of attained significance levels of T1)
T1_Critical=list()      (List of empirical critical values of T1)
T1_power=list()      (List of empirical powers of T1)
T2_alpha=list()      (List of attained significance levels of T2)
T2_Critical=list()      (List of empirical critical values of T2)
T2_power=list()      (List of empirical powers of T2)
Ahmad_alpha=list()      (List of attained significance levels of T3)
Ahmad_Critical=list()      (List of empirical critical values of T3)
```

```
Ahmad_power=list()     (List of empirical powers of T3)
Ours_alpha=list()     (List of attained significance levels of T)
Ours_Critical=list()     (List of empirical critical values of T)
Ours_power=list()     (List of empirical powers of T)


Rho=0.25     (Correlation in intraclass correlation matrix)
up=qnorm(1-(alpha))

for (idx in 2:20)\{     (Initializing the lists)
  P_N = c(P_N, list(c(seq(5,(idx-1)*5, by = 5), idx*5 - 1)))
  T1_alpha=c(T1_alpha,list(rep(0,idx)))
  T1_Critical=c(T1_Critical,list(rep(0,idx)))
  T1_power=c(T1_power,list(rep(0,idx)))

  T2_alpha=c(T2_alpha,list(rep(0,idx)))
  T2_Critical=c(T2_Critical,list(rep(0,idx)))
  T2_power=c(T2_power,list(rep(0,idx)))

  Ahmad_alpha=c(Ahmad_alpha,list(rep(0,idx)))
  Ahmad_Critical=c(Ahmad_Critical,list(rep(0,idx)))
  Ahmad_power=c(Ahmad_power,list(rep(0,idx)))

  Ours_alpha=c(Ours_alpha,list(rep(0,idx)))
  Ours_Critical=c(Ours_Critical,list(rep(0,idx)))
  Ours_power=c(Ours_power,list(rep(0,idx)))

}
```

```
for (pp in 1:length(P_N))
{
p=5*pp+5      (Setting P)


for (nn in 1:length(P_N[[pp]]))
{
    N=P_N[[pp]][nn]      (Setting N)
    n=N-1
    Mu=rep(0,p)      (Setting mean)
    Re=Rho*rep(1,p)%*%t(rep(1,p))+(1-Rho)*diag(1,p)
   (Correlation Structure)

    D=diag(rchisq(p,3))
    (Scalar matrix)
    sigma=(D^0.5)%*%Re%*%(D^0.5)
    (Covariance matrix)

(We first obtain attained significance levels and critical values)

########## Attained Alpha and Critical Values########
    T_1=rep(0,repetition)
    T_2=rep(0,repetition)
    f=rep(0,repetition)
    Ahmad=c(0)
    Y=c(0)
```

```
Ti=c(0)
li=c(0)
mu1s=c(0)
mu2s=c(0)
for(i in 1:repetition)
{
  X=mvrnorm(N,Mu,sigma)



  Xbar=apply(X,2,mean)
  X_Xbar=sweep(X,2,Xbar)
  (It subtracts ith element of X_bar from ith coloumn of X)

  S=(t(X_Xbar)%*%X_Xbar)/n
  D_s=diag(diag(S),ncol=p)
  D_sin=solve(D_s)
  D_sinh=diag(sqrt(1/diag(S)),ncol=p)
  R=D_sinh%*%S%*%D_sinh
  trR2=tr(R%*%R)
  C_pn=1+(trR2/p^(3/2))
  T_1[i]=((N*t(Xbar)%*%D_sin%*%Xbar)-
  (n*p/(n-2)))/sqrt(2*(trR2-(p^2/n))*C_pn)




  a1hat=tr(S)/p
  a2hat=(n^2/((n-1)*(n+2)))*(1/p )*(tr(S%*%S)-((tr(S))^2/n ))
  rhat=p*(a1hat^2/a2hat)
```

```
dg1=floor(rhat)
dg2=floor(n*rhat)
f[i]=qf((1-alpha),dg1,dg2)
T_2[i]=(N*t(Xbar)%*%Xbar)/(tr(S))
```

```
A=X%*%t(X)
A2=A^2
E3=(sum(A2)-sum(diag(A2)))/(N*(N-1))
B2=diag(A)%*%t(diag(A))
E2=(sum(B2)-sum(diag(B2)))/(N*(N-1))
s=(1/N)*(t(X)%*%X)
Xbar=apply(X,2,mean)
Q=N*t(Xbar)%*%Xbar
mm=Q/(tr(s))
Ahmad[i]=(mm-1)/sqrt(2*E3/E2)
```

```
fs=E2/E3
(k1s=fs)
(k2s=2*fs)
(k3s=8*fs)
(phi2s=k2s/k1s)
(phi3s=k3s/k1s)
```

```
(hs=1/3)
(mu1s[i]=1+(hs/k1s)*(((hs-1)*phi2s/2)  ))
(mu2s[i]=(hs^2*phi2s/k1s))
sig2=sum(X^2)/(p*N)
Ti[i]=(Q/(sig2*p))^hs
li [i]=(Ti[i]-mu1s[i])/sqrt(mu2s[i])


}
############## Alphas and Critical Values#############################
T1_alpha[[pp]][nn]=length(T_1[(T_1>up)])/repetition
T1_Critical[[pp]][nn]=sort(T_1)[9501]

Diff=T_2-f
T2_alpha[[pp]][nn]=length(Diff[Diff>=0])/repetition
T2_Critical[[pp]][nn]=sort(T_2)[9501]

Ahmad_alpha[[pp]][nn]=length(Ahmad[Ahmad>up])/repetition
Ahmad_Critical[[pp]][nn]=sort(Ahmad)[9501]

Ours_alpha[[pp]][nn]=length(Ti[(Ti-mu1s)/sqrt(mu2s)>up])/repetition
Ours_Critical[[pp]][nn]=sort(li)[9501]

############### Printing Tables of Alpha and Critical Values############
print(pp)
T11 = T1_alpha
T12= T1_Critical
T21 = T2_alpha
```

```
T22= T2_Critical
T31 = Ahmad_alpha
T32= Ahmad_Critical
T41 = Ours_alpha
T42= Ours_Critical
n_rows = length(T11)
le = length(T11[[n_rows]])
for (i in 1:n_rows)
{
  line11 = T11[[i]]
  line12 = T12[[i]]
  line21 = T21[[i]]
  line22 = T22[[i]]
  line31 = T31[[i]]
  line32 = T32[[i]]
  line41 = T41[[i]]
  line42 = T42[[i]]
  nNA = le - length(line11)
  title = toString(paste('p =',5*(i+1)))
  line11 = c(title, line11, rep(' ', nNA))
  line12 = c(title, line12, rep(' ', nNA))
  line21 = c(title, line21, rep(' ', nNA))
  line22 = c(title, line22, rep(' ', nNA))
  line31 = c(title, line31, rep(' ', nNA))
  line32 = c(title, line32, rep(' ', nNA))
  line41 = c(title, line41, rep(' ', nNA))
  line42 = c(title, line42, rep(' ', nNA))
  T11[[i]] = line11
```

```
  T12[[i]] = line12
  T21[[i]] = line21
  T22[[i]] = line22
  T31[[i]] = line31
  T32[[i]] = line32
  T41[[i]] = line41
  T42[[i]] = line42
}

  write.table(T11, 'T1_R_Intraclass_D_Chi2_Alpha.csv'
  , sep = ',', col.names = F, row.names = F)
  write.table(T12, 'T1_R_Intraclass_D_Chi2_Critical_Values.csv'
  , sep = ',', col.names = F, row.names = F)

  write.table(T21, 'T2_R_Intraclass_D_Chi2_Alpha.csv'
  , sep = ',', col.names = F, row.names = F)
  write.table(T22, 'T2_R_Intraclass_D_Chi2_Critical_Values.csv'
  , sep = ',', col.names = F, row.names = F)

  write.table(T31, 'Ahmad_R_Intraclass_D_Chi2_Alpha.csv'
  , sep = ',', col.names = F, row.names = F)
  write.table(T32, 'Ahmad_R_Intraclass_D_Chi2_Critical_Values.csv'
  , sep = ',', col.names = F, row.names = F)

  write.table(T41, 'Ours_R_Intraclass_D_Chi2_Alpha.csv'
  , sep = ',', col.names = F, row.names = F)
  write.table(T42, 'Ours_R_Intraclass_D_Chi2_Critical_Values.csv'
  , sep = ',', col.names = F, row.names = F)
```

```
(After having the empirical points, now we obtain empirical powers)
(We repeat sampling then we read the empirical critical values form-
the table we made in the previous step and then we calculate powers)
############ Powers of the Tests###########
Mu=(runif(p,-0.5,0.5))      (Mean under the alternative hypothesis)
Mu[seq(1,p,by=2)]=0

T_1=rep(0,repetition)
T_2=rep(0,repetition)
f=rep(0,repetition)
Ahmad=c(0)
Y=c(0)
Ti=c(0)
li=c(0)
mu1s=c(0)
mu2s=c(0)

for(i in 1:repetition)
{
  X=mvrnorm(N,Mu,sigma)
  Xbar=apply(X,2,mean)
  X_Xbar=sweep(X,2,Xbar)
 (It subtract ith element of X_bar from ith coloumn of X)

  (We compute T1 statistic under the alternative hypothesis)
  S=(t(X_Xbar)%*%X_Xbar)/n
```

```
D_s=diag(diag(S),ncol=p)
D_sin=solve(D_s)
D_sinh=diag(sqrt(1/diag(S)),ncol=p)
R=D_sinh%*%S%*%D_sinh
trR2=tr(R%*%R)
C_pn=1+(trR2/p^(3/2))
T_1[i]=((N*t(Xbar)%*%D_sin%*%Xbar)-
(n*p/(n-2)))/sqrt(2*(trR2-(p^2/n))*C_pn)


(We compute T2 statistic under the alternative hypothesis)
a1hat=tr(S)/p
a2hat=(n^2/((n-1)*(n+2)))*(1/p )*(tr(S%*%S)-((tr(S))^2/n ))
rhat=p*(a1hat^2/a2hat)
dg1=floor(rhat)
dg2=floor(n*rhat)
f[i]=qf((1-alpha),dg1,dg2)
T_2[i]=(N*t(Xbar)%*%Xbar)/(tr(S))


(We compute T3 statistic under the alternative hypothesis)
A=X%*%t(X)
A2=A^2
E3=(sum(A2)-sum(diag(A2)))/(N*(N-1))
B2=diag(A)%*%t(diag(A))
E2=(sum(B2)-sum(diag(B2)))/(N*(N-1))
s=(1/N)*(t(X)%*%X)
```

```
Xbar=apply(X,2,mean)
Q=N*t(Xbar)%*%Xbar
mm=Q/(tr(s))
Ahmad[i]=(mm-1)/sqrt(2*E3/E2)



(We compute T statistic under the alternative hypothesis)
fs=E2/E3
(k1s=fs)
(k2s=2*fs)
(k3s=8*fs)
(phi2s=k2s/k1s)
(phi3s=k3s/k1s)
(hs=1/3)
(mu1s[i]=1+(hs/k1s)*(((hs-1)*phi2s/2) ))
(mu2s[i]=(hs^2*phi2s/k1s))
sig2=sum(X^2)/(p*N)
Ti[i]=(Q/(sig2*p))^hs
li[i]=(Ti[i]-mu1s[i])/sqrt(mu2s[i])

}

(Now we compute the empirical powers)
```

```
T1_power[[pp]][nn]=length(T_1[T_1>T1_Critical[[pp]][nn]])/repetition
T2_power[[pp]][nn]=length(T_2[T_2>T2_Critical[[pp]][nn]])/repetition
Ahmad_power[[pp]][nn]=length(Ahmad[Ahmad>up])/repetition
Ours_power[[pp]][nn]=length(li[li>Ours_Critical[[pp]][nn]])/repetition

T1p = T1_power
T2p = T2_power
T3p= Ahmad_power
T4p = Ours_power
n_rows = length(T11)
le = length(T11[[n_rows]])
for (i in 1:n_rows)
{
  line1 = T1p[[i]]
  line2 = T2p[[i]]
  line3 = T3p[[i]]
  line4 = T4p[[i]]
  nNA = le - length(line1)
  title = toString(paste('p =',5*(i+1)))
  line1 = c(title, line1, rep(' ', nNA))
  line2 = c(title, line2, rep(' ', nNA))
  line3 = c(title, line3, rep(' ', nNA))
  line4 = c(title, line4, rep(' ', nNA))

  T1p[[i]] = line1
  T2p[[i]] = line2
  T3p[[i]] = line3
  T4p[[i]] = line4
```

```
}

write.table(T1p, 'T1_R_Intraclass_D_Chi2_power.csv', sep = ',',
col.names = F, row.names = F)
write.table(T2p, 'T2_R_Intraclass_D_Chi2_power.csv', sep = ',',
col.names = F, row.names = F)
write.table(T3p, 'Ahmad_Intraclass_D_Chi2_power.csv', sep = ',',
col.names = F, row.names = F)
write.table(T4p, 'Ours_Intraclass_D_Chi2_power.csv', sep = ',',
col.names = F, row.names = F)


(We repeat the procedure for each p and N)
}##for nn
}### for pp
```

Bibliography

Ahmad, M. R., Ohlson, M., and Rosen, D. V. (2011). Tests for covariance matrices, particularly for high-dimensional data. Technical report, Scandinavian Journal of Statistics.

Ahmad, M. R., Rosen, D. V., and Singull, M. (2013). A note on mean testing for high dimensional multivariate data under non-normality. *Statistica neerlandica*, 67(1):81–99.

Ahmad, M. R., Werner, C., and Brunner, E. (2008). Analysis of high dimensional repeated measures desings: The one sample case. *Computational Statistics and Data Analysis*, 53:416–427.

Atiqullah, M. (1962). The estimation of residual variance in quadratically balanced least-squares problems and the robustness of f-test. *Biometrika*, 49:83–91.

Bai, Z. and Saranadasa, H. (1996). Effect of high dimension: By and example of a two sample problem. *Statistica Sinica*, 6:311–329.

Chen, S. X., Zhang, L.-X., and Zhong, P.-S. (2010). Tests for high dimensional covariance matrices. *American Statistical Association*, 105:810–819.

Dempster, A. P. (1958). A high dimensional two sample significance test. *Annals of Mathematical Statistics*, 29:995–1010.

Dempster, A. P. (1960). A significance test for the separation of two highly multivariate small samples. *Biometrics*, 16:41–50.

Fisher, T. J., Sun, X., and Gallagher, C. M. (2010). A new test for sphericity of the covariance matrix for high dimensional data. *Journal of Multivariate Analysis*, 101:2554–2570.

Iwashita, T. (1997). Asymptotic null and nonnull distribution of hotelling's statistic under the elliptical distribution. *Journal of Statistical Planning and Inference*, 61:85–104.

Jensen, D. R. and Solomon, H. (1972). A gaussian approximation to the distribution of a definite quadratic form. *American Statistical Association*, 67(340):898–902.

Johnson, R. A. and Wichern, D. W. (2007). *Applied Multivariate Statistical Analysis*. Pearson Prentice Hall.

Ledoit, O. and Wolf, M. (2002). Some hypothesis tests for the covariance matrix when the dimension is large compared to the sample size. *The Annals of Statistics*, 30:1081–1102.

Ledoit, O. and Wolf, M. (2004). A well-conditioned estimator for large-dimensional covariance matrices. *J Multivariate Anal*, 88:365–411.

Lin, S. and Perlman, M. (1985). A Monte Carlo comparison of four estimators of a covariance matrix. *Journal of Multivariate Analysis*.

Mathai, A. and Provost, S. (1992a). *Quadratic Forms in Random Variables*. Statistics: A Series of Textbooks and Monographs. Taylor & Francis.

Mathai, A. M. and Provost, S. B. (1992b). *Quadratic forms in random variables: Theory and applications*. M. Dekker, New York.

Moschopoulos, P. and Canada, W. (1984). The distribution function of a linear combination of chi-squares. *Computers & Mathematics with Applications*, 10(4):383 – 386.

Moschopoulos, P. G. (1983). On a new transformation to normality. *Communications in Statistics*, 12(16):1873–1875.

Moschopoulos, P. G. and Mudholkar, G. S. (1983). A likelihood-ratio-based normal approximation for the non-null distribution of the multiple correlation coefficient. *Communications in Statistics - Simulation and Computation*, 12(3):355–371.

Mudholkar, G. S. and Trivedi, M. C. (1980). A normal approximation for the distribution of the likelihood ratio statistic in multivariate analysis of variance. *Biometrika*, 67(2):485–488.

Oman, S. D. and Zacks, S. (1981). A mixture approximation to the distribtuion of a weighted sum of cchi-square variables. *Journal of Statistical Computation and Simulation*, 13:215–224.

Pachares, J. (1955). Note on the distribution of a definite quadratic form. *The Annals of Mathematical Statistics*, 26:128–131.

Pourahmadi, M. (2013). *High-Dimensional Covariance Estimation: With High-Dimensional Data*. Wiley Series in Probability and Statistics. Wiley.

Rao, C. R. and Kleffe, J. (1988). Estimation of variance components and applications. *Elsevier science publishers*.

Ruben, H. (1960). Probability content of regions under spherical normal distribtuions. *The Anllas of Mathematical Statistics*, 31:598–619.

Ruben, H. (1962). Probability content of regions under the sperical normal distribtuion: The distribtuion of homogeneous and nonhomogenous quadratic ffunction in normal variables. *The Annals of Mathematical Statistics*, 33:542–570.

Shah, B. K. and Khatri, C. G. (1961). Distribution of definite quadratic form for non-central normal variates. *The Annals of Mathematical Statistics*, 32:883–887. (Correction 34, 673).

Srivastava, M. S. (2007). Multivariate theory for analyzing high dimensional data. *Japan Statistical Society*, 37:53–86.

Srivastava, M. S. (2009). A test for the mean vector with fewer observations than the dimension under non-normality. *Multivariate Analysis*, 100:518–532.

Srivastava, M. S. and Du, M. (2008). A test for the mean vector with fewer observations than the dimension. *Journal of Multivariate Analysis*, 99(3):386–402.

Staniswalis, J. G., Severini, T. A., and Moschopoulos, P. G. (1993). On a data-based power transformation for reducing skewness. *Journal of Statistical Computation and Simulation*, 46:91–100.

Stein, C. (1956). Inadmissibility of the usual estimator for the mean of a multivariate normal distribution. In *Proceedings of the Third Berkeley Symposium on Mathematical Statistics and Probability, Volume 1: Contributions to the Theory of Statistics*, pages 197–206, Berkeley, Calif. University of California Press.

Wiens, D. P. (1992). On moments of quadratic forms in non-spherically distributed random variables. *Statistics*, 23:265–270.

Wilson, E. B. and Hilferty, M. M. (1931). The distribution of chi-square. In 17, editor, *National Academy of Sciences*, pages 684–688.

Curriculum Vitae

Behzad Aalipur Hafshejani was born on November 21, 1988. He was a Bronze Medalist at National Physics Olympiad in 2006. He graduated from NODET (National Organization for Development of Exceptional Talents) High School in the spring of 2007, Shahrekord, Iran.

He ranked in top 1% among one million student applicants in national university entrance exam and entered Sharif University of Technology in the fall of 2007. He became a member of National Elite Foundation in 2011. He graduated in spring of 2012 with a bachelor of science in industrial engineering.

He ranked in top 1% in Nationwide M.Sc Entrance Exam in Statistics in 2013, but he started his master's at the university of Texas at El Paso in spring 2014 majoring in Statistics. He worked as a Teaching Assistant, and as a Research Assistant during his studies. He received a certificate of merit from Phi-Kappa-Phi in 2015 and graduated as an outstanding graduate student in spring of 2016.

He was one of the program committee members in IJCAI 2016 Workshop on Knowledge Discovery in Healthcare Data. He joined PhD program in UW-Madison in fall 2016 to continue his studies in statistics.

Permanent address: No. 63 Kashani Street, Hafshejan, Iran