STUDY OF UNIFIED MULTIVARIATE SKEW NORMAL DISTRIBUTION WITH APPLICATIONS IN FINANCE AND ACTUARIAL SCIENCE

Mohammad Abdus Samad Aziz

A Dissertation

Submitted to the Graduate College of Bowling Green
State University in partial fulfillment of
the requirements for the degree of

DOCTOR OF PHILOSOPHY

August 2011

Committee:

Arjun K. Gupta, Advisor

Arthur Yeh,
Graduate Faculty Representative

John T. Chen

Wei Ning

ABSTRACT

Arjun K. Gupta, Advisor

The classical work horse in finance and insurance for modeling asset returns is the Gaussian model. However, when modeling complex random phenomena, more flexible distributions are needed which are beyond the normal distribution. This is because most of the financial and economic data are skewed and have “fat tails” due to the presence of outliers. Hence symmetric distributions like normal or others may not be good choices while modeling these kinds of data. Flexible distributions like skew normal distribution allow robust modeling of high-dimensional multimodal and asymmetric data. In this dissertation, we consider a very flexible financial model to construct robust comonotonic lower convex order bounds in approximating the distribution of the sums of dependent log skew normal random variables. The dependence structure of these random variables is based on a recently developed multivariate skew normal distribution, called unified skew normal distribution. In order to accommodate the distribution to the model so considered, we first study inherent properties of this class of skew normal distribution. These properties along with the convex order and comonotonicity of random variables are then used to approximate the distribution function of terminal wealth. By calculating lower bounds in the convex order sense, we consider approximations that reduce the multivariate randomness to univariate randomness. The approximations are used to calculate the risk measure related to the distribution of terminal wealth. The accurateness of the approximation is investigated numerically. Results obtained from our methods are competitive with a more time consuming method called, Monte Carlo method. The dissertation also includes the study of quadratic forms and their distributions in the unified skew normal setting. Regarding the inferential issue of the distribution, we propose an estimation procedure based on the weighted moments approach. Results of our simulations provide an indication of the accuracy of the proposed method.

To the loving memory of my father, who inspired me to acquire knowledge and to become a nice human being. Every good deed I do is because of my father.

ACKNOWLEDGMENTS

With a deep gratitude I wish to thank my advisor Dr. Arjun Gupta for his constant support, advice and valued suggestions throughout this research. I am especially grateful for his incredible patience. His explanations were very lively and understandable. I will remember all my life his irreproachable supervision and help. I also want to extend my gratitude to the other members of my committee, Dr. John Chen, Dr. Wei Ning, and Dr. Arthur Yeh. I am thankful to Dr. Ning for his discussions and advices. I would like to take this opportunity to thank Dr. Ray Kresman, professor, Computer Science department for helping with the computer programming.

I am also grateful to the Department of Mathematics and Statistics at Bowling Green State University for the teaching assistantship and dissertation fellowship awarded to me during my study.

I wish to thank my colleagues and friends from Bowling Green, Xin Hu, Adamu, Swarup, Junvie, Ngoc, and all the other people for their friendship and supports.

My deepest gratitude goes to my mom and siblings for their special prayer. Staying away from them was always difficult. Special thanks are due to my brother Abdus Wahed and Abdul Alim for encouraging me to pursue doctoral degree.

Last, but not the least, I would like to thank my wife Ruksana who has motivated me to keep moving through difficult time. Her love, support, encouragement, patience and sacrifice was in the end what made this dissertation possible. We are also blessed with our daughter Audrika who brings all the joy and happiness to our life.

Mohammad Abdus Samad Aziz

Bowling Green, Ohio

May, 2011.

Table of Contents

List of Figures

List of Tables

CHAPTER 1

INTRODUCTION

1.1 Notations

In this dissertation, matrices will be denoted by capital letters and vectors by small bold letters. Unless otherwise stated all the vectors are column vectors. The following notation is used throughout the dissertation:

$\Re^d$: d-dimensional real space

$A_{d\times m}$: $d \times m$-dimensional matrix A

$(A)_{i,j}$: (i, j) element of the matrix A

$(A)_{i.}$: ith row of the matrix A

$(A)_{.j}$: jth column of the matrix A

A^T : transpose of the matrix A

$|A|$ *or* $det(A)$: determinant of the square matrix A

A^{-1} : inverse of the matrix A when $|A| \neq 0$

$tr(A)$: trace of a square matrix A

I_d : d-dimensional identity matrix

$\mathbf{1}_d$: d-dimensional column vector of ones

$\mathbf{0}_d$: d-dimensional column vector of zeros

$0_{d\times m}$: $d \times m$-dimensional matrix of zeros

$N_d(\boldsymbol{\mu}, \Sigma)$: d-dimensional normal density with mean $\boldsymbol{\mu}$ and variance covariance matrix Σ

$\phi_d(\boldsymbol{y} - \boldsymbol{\mu}; \Sigma)$ (or $\phi_d(\boldsymbol{y}; \boldsymbol{\mu}, \Sigma)$: density of a $N_d(\boldsymbol{\mu}, \Sigma)$ random variable evaluated at $\boldsymbol{y}$

$\phi_d(\boldsymbol{y})$: density of a $N_d(\boldsymbol{0}, I_d)$ random variable evaluated at $\boldsymbol{y}$

$\Phi_d(\boldsymbol{y} - \boldsymbol{\mu}; \Sigma)$ (or $\Phi_d(\boldsymbol{y}; \boldsymbol{\mu}, \Sigma)$: distribution of a $N_d(\boldsymbol{\mu}, \Sigma)$ random variable evaluated at $\boldsymbol{y}$

$\Phi_d(\boldsymbol{y})$: distribution of a $N_d(\boldsymbol{0}, I_d)$ random variable evaluated at $\boldsymbol{y}$

$\phi(y - \mu; \sigma^2)$(or $\phi(y; \mu, \sigma^2)$: density of a $N(\mu, \sigma^2)$ random variable evaluated at y

$\Phi(y - \mu; \sigma^2)$(or $\phi(y; \mu, \sigma^2)$: distribution of a $N(\mu, \sigma^2)$ random variable evaluated at y

$\phi(y)$: density of a $N(0, 1)$ random variable evaluated at y

$\Phi(y)$: distribution of a $N(0, 1)$ random variable evaluated at y

pdf : Probability density function

cdf : Cumulative distribution function

m.g.f : Moment generating function

jmgf : Joint moment generating function

1.2 Preleminaries

1.2.1 Vectors, Matrices and their Properties

A d-dimensional real *vector* $\boldsymbol{a}$ is an ordered array of real numbers $a_i, i = 1, 2, ..., d$, organized in a single column, written as $\boldsymbol{a} = (a_i)$.

A real *matrix* A of dimension $d \times m$ is an ordered rectangular array of real numbers a_{ij} arranged in rows $i = 1, 2, ..., d$ and columns $j = 1, 2, ..., m$, written as $A = (a_{ij})$.

For two $d \times m$ matrices A and B, the *addition* and *subtraction* are defined as $A \pm B = (a_{ij} \pm b_{ij})$. *Product* of two matrices A of dimension $d \times m$ and B of dimension $m \times n$ is defined as $AB = (c_{ij})$ where $c_{ij} = \sum_{l=1}^{d} a_{il} b_{lj}$.

A matrix A is called *square* matrix of order d if the number of columns and rows of A are equal to d.

The *transpose* of a matrix A is obtained by interchanging the rows and columns of A and represented by A^T. When $A = A^T$, we say A is *symmetric.* A $d \times d$ symmetric matrix A for which $a_{ij} = 0, i \neq j$ for $i, j = 1, 2, ..., d$ is called *diagonal* matrix of order d, and represented by $A = diag(a_{11}, a_{22}, ..., a_{dd})$; in this case if $a_{ii} = 1$ for all i, then we denote this by I_d and call it *identity* matrix.

A square matrix A of order d is called a *lower triangular* matrix if $a_{ij} = 0, i < j$ and *upper triangular* if $a_{ij} = 0, i > j$. A square matrix A of order d is called a *lower unit triangular* matrix if $a_{ij} = 1, i \leq j$ and *upper unit triangular* if $a_{ij} = 1, i \geq j$

A symmetric matrix A of order d is *positive definite* if $\boldsymbol{a}^T A \boldsymbol{a} > 0$ for all nonzero vectors $\boldsymbol{a}$. A symmetric matrix A of order d is *positive semi-definite* if $\boldsymbol{a}^T A \boldsymbol{a} \geq 0$ for all nonzero vectors $\boldsymbol{a}$.

Let A be a $d \times d$ matrix. Then the roots of the equation $det(A - \lambda I_d) = 0$ are called *charateristic roots* or *eigenvalues* of the matrix A. The *determinant* of a square matrix A is defined as the product of eigenvalues of the matrix. If $det(A) \neq 0$, then A is called *non-singular* matrix and there exists a unique matrix B such that $AB = BA = I_d$ or $B = A^{-1}$, B is called the *inverse* of A. The *trace* of a square matrix A of order d is defined as the sum of its eigenvalues and written as tr(A).

The number of nonzero eigenvalues of a square matrix is the *rank* of A and written as rank(A). If the rank of a $d \times m$ matrix is rank(A)= min(d, m) then A is called a *full rank* matrix.

Let $A = (a_{ij})$ be a $d \times m$ matrix. Then a 2×2 *partition* of A is defined as

$$A = \begin{array}{cc} \begin{array}{cc} r & m-r \end{array} & \\ \begin{pmatrix} A_{11} & A_{12} \\ A_{21} & A_{22} \end{pmatrix} & \begin{array}{l} k \\ d-k \end{array} \end{array}$$

where A_{11}, A_{12}, A_{21} and A_{22} are

$$\begin{aligned}
A_{11} &= (a_{ij}), i = 1, 2, \ldots, k, j = 1, 2, \ldots, r \\
A_{12} &= (a_{ij}), i = 1, 2, \ldots, k, j = r + 1, 2, \ldots, m \\
A_{21} &= (a_{ij}), i = k + 1, 2, \ldots, d, j = 1, 2, \ldots, r
\end{aligned}$$

and

$$A_{22} = (a_{ij}), i = k + 1, 2, \ldots, d, j = r + 1, 2, \ldots, m$$

The *Kronecker product* for any two matrices $A_{m\times n} = (a_{ij})$ and $B_{p\times q} = (b_{ij})$ denoted by $A \otimes B$ results an $mp \times nq$ matrix defined by

$$A \otimes B = \begin{bmatrix} a_{11}B & a_{12}B \ldots a_{1n}B \\ a_{21}B & a_{22}B \ldots a_{2n}B \\ \vdots a_{m1}B & a_{m2}B \ldots a_{mn}B \end{bmatrix} = (a_{ij}B)$$

For the properties of the Kronecker matrix product we refer to Gupta and Nagar (2000).

The *direct sum operator* of two matrices $A_{m\times n} = (a_{ij})$ and $B_{p\times q} = (b_{ij})$ denoted by $A \oplus B$ results an $(m + p) \times (n + q)$ matrix defined by

$$\begin{bmatrix} A & \mathbf{0} \\ \mathbf{0} & B \end{bmatrix}$$

That is matrix direct sum operator gives a block diagonal matrix.

Note: It is easy to see that $\oplus_{i=1}^{n} A = I_n \otimes A$. For other properties of matrix direct sum operator we refer to Horn and Johnson (1991).

1.2.2 Multivariate Random Variable and Its Distribution

A multivariate random variable is represented by a d dimensional column vector $\boldsymbol{y}$ where d random variables $y_1, y_2, \ldots y_d$ are observed on a sampling unit. In the univariate case, the distribution of a random variable y can be characterized by its cumulative distribution function (cdf) $F_Y(.)$, which is defined as $F_Y(y) = P(Y \leq y)$. We say that the random variable y with the cdf $F_Y(.)$ has the probability density function (pdf) $f(y)$ if $dF(y) = f(y)$. Similarly, the distribution of a random variable $\boldsymbol{y}$ in $\Re^d$ can be characterized by its joint cumulative distribution function (jcdf) $F_{\boldsymbol{y}}(.)$. For $\boldsymbol{c}$ in $\Re^d$, it is defined as $F_{\boldsymbol{y}}(\boldsymbol{c}) = P(y_1 \leq c_1, y_2 \leq c_2, \ldots, y_d \leq c_d)$. We say that the random variable $\boldsymbol{y}$ with jcdf $F_{\boldsymbol{y}}(.)$ has joint probability density function (pdf) $f(\boldsymbol{y})$ if $dF(\boldsymbol{y}) = f(\boldsymbol{y})$. The joint moment generating function (jmgf) of the random variable $\boldsymbol{y}$ is defined as $M_{\boldsymbol{y}}(\boldsymbol{t}) = E(exp(\boldsymbol{t}^T\boldsymbol{y}))$, if it exists for all $\boldsymbol{t}$ in some neighborhood of $\mathbf{0}$ in $\Re^d$. To say that a random vector has a certain distribution the notation $\boldsymbol{y} \sim F(\boldsymbol{y}), \boldsymbol{y} \sim f(\boldsymbol{y})$ or $\boldsymbol{y} \sim M_{\boldsymbol{y}}(\boldsymbol{t})$ is equivalently used. If the joint distribution of d random variables $y_1, ..., y_d$ is known, then the *marginal distribution* of any subset k of the d random variables $y_1, ..., y_d$ can be derived from this joint distribution by integrating over all possible values of the other $d-k$ variables. Similarly the conditional distribution can be derived. Suppose that the random vector $\boldsymbol{y} = (y_1, ..., y_d)^T$ is divided into two subvectors $\boldsymbol{x}$ and $\boldsymbol{z}$, where $\boldsymbol{x}$ is a k-dimensional random vector comprising k of the d random variables in $\boldsymbol{y}$, and $\boldsymbol{z}$ is an $(d-k)$-dimensional random vector comprising the other $d-k$ random variables. Suppose also that the pdf of $\boldsymbol{y}$ is f_1 and that the marginal pdf of $\boldsymbol{z}$ is f_2. Then for every given point $\boldsymbol{z} = \boldsymbol{\xi} \in \Re^{d-k}$ such that $f_2(\xi) > 0$, the conditional k-dimensional pdf g_1 of $\boldsymbol{x}$ when $\boldsymbol{z} = \boldsymbol{\xi}$ is defined as:

$$f(\boldsymbol{x}|\boldsymbol{z}) = \frac{f(\boldsymbol{x}, \boldsymbol{z})}{f_2(\boldsymbol{z})}, \boldsymbol{x} \in \Re^k.$$

1.3 Literature Review, Problem and Organization

In recent years, there has been considerable interest in the construction of the general class of skewed distributions which include the standard symmetric distributions such as the normal, t, logistic and Cauchy distributions. The key is to introduce additional parameters or parametric functions in the distributional form that accounts for the skewness of the distribution. The first of this kind of model was studied by Roberts (1966). The idea became institutionalized when Azzalini (1985) defined a class of distributions (which he referred to as skew normal (SN)) by introducing an additional skewness parameter that included the normal distribution as a special case. The name suggests that this distribution, unlike the normal distribution, is asymmetric in general and allows both positively and negatively skewed distributions. Subsequently, Azzalini and Dalla Valle (1996) came up with the multivariate version of the skew-normal distribution. A statistical application of the multivariate skew-normal distribution was considered by Azzalini and Capitanio (1999). This paper popularized the application of such distributions and led the way for others to define similar families of distributions based on other symmetric distributions.

Because of the popularity several other versions of the multivariate skew normal model are proposed, for example, Sahu et al (2003), Liseo and Loperfido (2003), Gupta, Gonzalez-Farias, Dominguez-Molina (2004), Gonzalez-Farias, Dominguez-Molina, and Gupta (2004), Arellano-Valle and Genton (2005), to name a few among many. Arellano-valle and Azzalini (2006) developed a skew normal model which includes or is at least equivalent to the earlier versions of the skew normal models. They called it unified skew normal model with the acronym SUN. The purpose of this dissertation is to study the properties and inferential issues related to SUN model which is motivated by the realization of the importance of this model. The unique feature of this study is that the methodologies and tools that are developed for SUN model, remain true for all the earlier versions of skew normal models.

The rest of the dissertation is organized as follows. In chapter 2, an overview of skew normal distribution is provided and properties of unified multivariate skew normal random vector are discussed that will be used in later chapters. In particular, it is shown that SUN density is closed under:

- linear transformations of full column (or row) rank
- marginalization
- conditional distribution
- joint distributions of independent SUN random variables, and
- sums of independent SUN random variables

Chapter 3 is devoted to study the quadratic forms under the unified skew normal setting. Quadratic forms and their distributions are important for statistical inference. They have their applications in economics and time series as well. The study in this chapter includes:

- Distribution of quadratic forms
- Expected value of quadratic forms and their functions
- Independence of quadratic forms, and independence between a linear form and a quadratic form

In chapter 4, an estimation technique, called method of weighted moments (MOWM) is developed to estimate parameters of a SUN model. It paved the way to apply the model to real data. In chapter 5, two applications of the SUN density are presented in context of finance and actuarial science. In general, the sum of log-normal and log-skew normal distribution does not have an exact distribution or follow the parent distribution. In this chapter, the concept of convex order of random variables is used to construct a lower bound in order to approximate the sum of dependent log unified skew normal random variables. The derived bounds are then used to approximate the distribution of terminal wealth and calculate the portfolio risk, called value at risk (VaR). Performance of the derived bounds is also compared with time consuming Monte Carlo method.

CHAPTER 2

PROPERTIES OF UNIFIED SKEW NORMAL RANDOM VECTOR

2.1 Introduction

Unified skew normal (SUN) density has many interesting and appealing properties and it also preserves some important properties of the multivariate normal distribution. In this chapter, we explore some of these properties. We start with a brief background description of the multivariate normal distribution, and univariate and multivariate skew normal distributions. We then describe the basic properties of the unified skew normal distribution, starting with the moment generating function that allows us to establish other important properties of interest. We show that similar to the multivariate normal distribution, the unified skew normal distribution is also closed under linear transformations, marginalization and conditioning. To be more precise, we show that for a random vector with a the unified skew normal distribution all row (column) full rank linear transformations are in the same family of distributions, marginal and conditional distribution of unified skew normal distribution belong to the same family, the joint distribution and sum of independent unified skew normal random vectors is again unified skew normal distributed random vector. Since the unified skew normal density includes most of the skew normal models developed earlier, these properties remain true for all those skew normal models.

2.2 Multivariate Normal Distribution (MND)

Definition 2.2.1. *A d-dimensional random vector* $\boldsymbol{y}$ *is said to have a d-dimensional normal distribution with parameters* $\boldsymbol{\mu} \in \Re^d$ *and* $d \times d$ *positive definite covariance matrix* Σ *if the density is given by*

$$f(\boldsymbol{y}) = \frac{1}{(2\pi)^{d/2} \ |\Sigma|^{1/2}} \quad e^{-\frac{1}{2}\boldsymbol{y}^T \Sigma^{-1} \boldsymbol{y}}. \tag{2.1}$$

We say that $\boldsymbol{y}$ *is distributed as* $N_d(\boldsymbol{\mu}, \Sigma)$*, and write* $\boldsymbol{y} \sim N_d(\boldsymbol{\mu}, \Sigma)$.

Some Properties of Multivariate Normal Distribution

Some of the important properties of the multivariate normal distribution are stated here without proof. The proofs could be found from any multivariate book.

- The m.g.f of (2.1) is given by

  $$M_{\boldsymbol{y}}(\boldsymbol{t}) = exp(\boldsymbol{\mu}^T \boldsymbol{t} + \frac{1}{2}\boldsymbol{t}^T \Sigma \boldsymbol{t}),$$

  where $\boldsymbol{t} \in \Re^d$.

- The parameters of $N_d(\boldsymbol{\mu}, \Sigma)$ have the direct interpretation as the mean vector and the variance covariance matrix of $\boldsymbol{y}$ that is $\mathbb{E}(\boldsymbol{y}) = \mu$ and $\mathbb{E}(\boldsymbol{y} - \boldsymbol{\mu})(\boldsymbol{y} - \boldsymbol{\mu})^T = \Sigma$.

- If $\boldsymbol{y} \sim N_d(\boldsymbol{\mu}, \Sigma)$, then $\boldsymbol{z} = \Sigma^{-1/2}(\boldsymbol{y} - \boldsymbol{\mu})$ has the distribution $N_d(\boldsymbol{0}, I_d)$. The m.g.f in this case becomes $\mathbb{E}(e^{\boldsymbol{t}^T \boldsymbol{z}}) = exp(-\frac{1}{2}\boldsymbol{t}^T \boldsymbol{t})$ and $x = \boldsymbol{z}^T \boldsymbol{z}$ has *chi-square* distribution with d degrees of freedom and is denoted by χ^2_d.

- If $\boldsymbol{y} \sim N_d(\boldsymbol{\mu}, \Sigma)$, then $(\boldsymbol{y} - \boldsymbol{\mu})^T \Sigma^{-1} (\boldsymbol{y} - \boldsymbol{\mu}) \sim \chi^2_d$.

- The family of the normal distributions is closed under the linear transformations, marginalization and conditioning.

- Let the $d \times 1$ random vector $\boldsymbol{y} \sim N_d(\boldsymbol{\mu}, \Sigma)$, $\boldsymbol{a}$ be a $d \times 1$ vector of constants and A be any $n \times d$ matrix of constants with rank of A is $n \leq d$. Then

  (i) $z = \boldsymbol{a}^T \boldsymbol{y} \sim N(\boldsymbol{a}^T \boldsymbol{\mu}, \boldsymbol{a}^T \Sigma \boldsymbol{a})$.

  (ii) $\boldsymbol{z} = A\boldsymbol{y} \sim N_n(A\boldsymbol{\mu}, A\Sigma A^T)$.

- If $\boldsymbol{y}$ and $\boldsymbol{x}$ are multivariate normal random vectors with $\Sigma_{\boldsymbol{yx}} \neq 0$, then the conditional distribution of $\boldsymbol{y}|\boldsymbol{x}$ is multivariate normal with the mean vector

$$\mathbb{E}(\boldsymbol{y}|\boldsymbol{x}) = \mu_{\boldsymbol{y}} + \Sigma_{\boldsymbol{yx}} \Sigma_{\boldsymbol{xx}}^{-1} (\boldsymbol{x} - \mu_{\boldsymbol{x}}),$$

  and the covariance matrix

$$Cov(\boldsymbol{y}|\boldsymbol{x}) = \Sigma_{\boldsymbol{yy}} - \Sigma_{\boldsymbol{yx}} \Sigma_{\boldsymbol{xx}}^{-1} \Sigma_{\boldsymbol{xy}}.$$

2.3 Skew Normal Distributions

The normal distribution is the most popular distribution because of its many appealing properties. Two main reasons for its popularity are: first, the effect of the central limit theorem, in most cases the distribution observations is at least approximately normal; second, normal distribution and its sampling distribution are easily tractable. However, the same family of distribution is not used frequently in modeling financial and insurance data because they do not behave in the normal sense. Most of the economic and finance data usually have outliers that produce “fat tail” distributions and in this case normal distribution is not a good model to use. Researchers have been looking for an alternative model than the normal to cope with the skewness property of the distribution of these kinds of data. The first of this kind of the models was originated from a paper by Roberts (1966). Azzalini (1985) named this class of distributions as the skew normal class. The multivariate skew normal distribution have been introduced by Azzalini and Dalla Valle (1996) and subsequently by Azzalini and Capitanio (1999), Gupta, Gonzalez-Farias and Dominguez-Molina (2004) etc. Gonzalez-Farias, Dominguez- -Molina, and Gupta (2004) defined the multivariate closed skew-normal distribution which has properties similar to the normal distribution than any other multivariate skew normal distribution. The study of the skew normal distribution explores an approach for statistical analysis without the symmetry assumption for the underlying distribution.

2.3.1 Univariate Skew Normal Distribution

Definition 2.3.1. *Let Y be a continuous random variable. Let ϕ and Φ denote the standard normal density and corresponding distribution function, respectively. Then Y, is said to have a skew-normal distribution with the parameter α,$-\infty \leq \alpha \leq \infty$ if the density of Y is*

$$f(y) = 2\phi(y)\Phi(\alpha y), \qquad -\infty \leq y \leq \infty. \tag{2.2}$$

and we write $Y \sim SN(\alpha)$.

The component α is called the shape parameter because it regulates the shape of the density function. As α increases (in absolute value), the skewness of the distribution increases. Figure 2.3.1 on the next page shows the density for different values of alpha. In practice, to fit real data, we work with an affine transformation $Z = \xi + \sigma Y$ with $\xi \in \Re$ and $\sigma > 0$. The density of Z is then written as

$$g(z; \xi, \sigma, \alpha) = \frac{2}{\sigma}\phi\Big(\frac{z-\xi}{\sigma}\Big)\Phi\Big(\alpha\frac{z-\xi}{\sigma}\Big),$$

and we write $Z \sim SN(\xi, \sigma^2, \alpha)$.

Properties of Univariate Skew-normal distribution

The density (2.2) possesses some interesting properties as noted below:

- When $\alpha = 0$, we re-obtain normal distribution.Thus normal distribution is a special

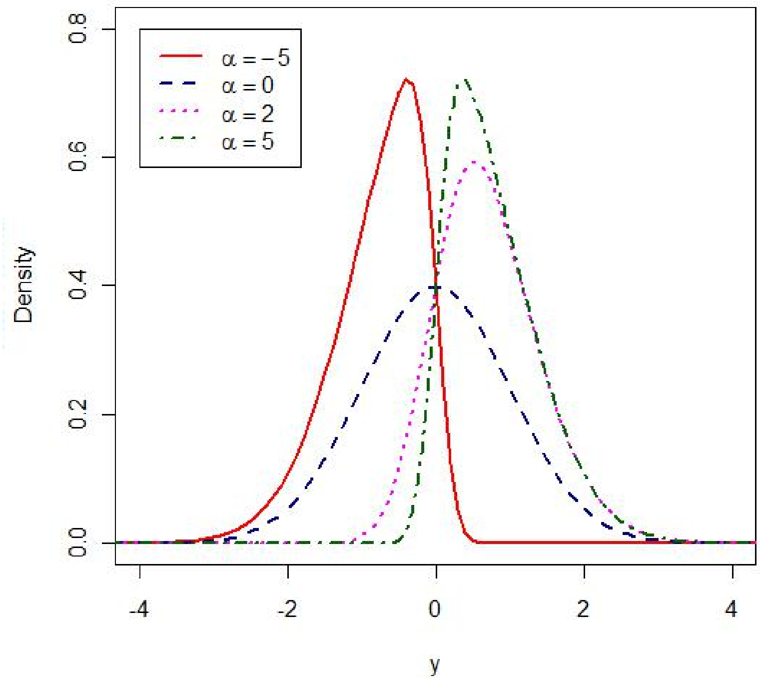

Figure 2.3.1: Univariate skew normal distribution for different values of alpha.

case of skew-normal distribution.

- As α tends to ∞, (2.2) becomes $f(y) = \phi(y), \quad 0 \leq y \leq \infty$ which is the half-normal (folded normal) pdf.
- If $Y \sim N(0,1)$ and $X \sim SN(\alpha)$, then $|Y|$ and $|X|$ have the same pdf.
- If $Y \sim SN(\alpha)$, then $Y^2 \sim \chi^2_1$.
- If $Y \sim SN(\alpha)$, then $-Y$ is a $SN(-\alpha)$.

Moment Generating function of univariate skew normal distribution

Let Y be a $SN(\alpha)$ random variable. Then the moment generating function of Y is given by

$$M(t) = 2\ exp\ (\frac{t^2}{2})\Phi(\delta t),$$

where $\delta = \alpha/\sqrt{1+\alpha^2}$. From the m.g.f the first two moments of the univariate skew normal density are obtained as

$$\mathbb{E}(Y) = \sqrt{\frac{2}{\pi}}\delta \text{ and } Var(Y) = 1 - (2/\pi)\delta^2.$$

2.3.2 Multivariate Skew Normal Distribution (MSND)

A multivariate version of skew-normal density was introduced by Azzalini and Dalla Valle (1996) and Azzalini and Capitanio (1999). Like univariate skew normal distribution, these classes of distributions include normal distribution and have some properties similar to the normal distribution.

Definition 2.3.2. *(Azzalini and Dalla Valle (1996)) A random vector $\boldsymbol{y}$ follows a multivariate skew-normal distribution if the density of $\boldsymbol{y}$ is given by*

$$f_d(\boldsymbol{y}) = 2\phi_d(\boldsymbol{y}, \Sigma)\Phi(\boldsymbol{\alpha}^T\boldsymbol{y}), \quad \boldsymbol{y}, \boldsymbol{\alpha} \in \Re^d. \tag{2.3}$$

and is usually denoted by $SN_d(\boldsymbol{\alpha}, \Sigma)$. As in the univariate case, introducing location and scale parameters, $\boldsymbol{\xi} = (\xi_1, \xi_2, \ldots, \xi_d)^T$ and $S = diag(\sigma_1, \sigma_1, \ldots \sigma_d)$ respectively, the density usually denoted by $SN_d(\boldsymbol{\xi}, S\Sigma S, \boldsymbol{\alpha})$ can be rewritten as

$$f_d(\boldsymbol{y}) = 2\phi_d(\boldsymbol{y}; \boldsymbol{\xi}, S\Sigma S)\Phi(\boldsymbol{\alpha}^T S^{-1}(\boldsymbol{y} - \boldsymbol{\xi})). \tag{2.4}$$

Properties of Multivariate Skew-normal distribution

Like univariate case the density (2.3) possesses some interesting properties as noted below:

- When $\boldsymbol{\alpha} = 0$, we re-obtain multivariate normal distribution.Thus multivariate normal distribution is a special case of multivariate skew-normal distribution.
- If $\boldsymbol{y} \sim SN_d(\boldsymbol{\alpha}, \Sigma)$ and $\boldsymbol{x} \sim N_d(\mathbf{0}, \Sigma)$, then $\boldsymbol{y}^T\Sigma^{-1}\boldsymbol{y}$ is equal to $\boldsymbol{x}^T\Sigma^{-1}\boldsymbol{x}$ in distribution.
- If $\boldsymbol{y} \sim SN_d(\boldsymbol{\alpha}, \Sigma)$, then $\boldsymbol{y}^T\Sigma^{-1}\boldsymbol{y} \sim \chi^2_d$.

- If $\boldsymbol{y} \sim SN_d(\boldsymbol{\alpha}, \Sigma)$ and B is a positive symmetric semidefinite $d \times d$ matrix with rank k such that $B\Sigma B^T = B$, then $\boldsymbol{y}^T B \boldsymbol{y} \sim \chi^2_k$.

- Suppose $\boldsymbol{y} \sim SN_d(\mathbf{0}, \Sigma, \boldsymbol{\alpha})$. Then the distribution of $\boldsymbol{y}\boldsymbol{y}^T$ is Wishart with d.f 1 and the scale parameter Σ.

Liseo and Loperfido (2003) derived a multivariate skew normal distribution in the Bayesian context and called it hierarchical skew-normal (HSN) density. Here is the formal definition.

Definition 2.3.3. *Let $\boldsymbol{\theta}|\boldsymbol{\theta}_0 \sim N_p(\boldsymbol{\theta}_0, \Sigma)$ and $\boldsymbol{\theta}_0 \sim N_p(\boldsymbol{\mu}, \Omega)$. Then under the constraints $C\boldsymbol{\theta}_0 + \boldsymbol{d} \leq \mathbf{0}$ where C is a full rank $k \times p$ matrix and $\boldsymbol{d} \in \Re^k$, the marginal distribution of θ is given by*

$$\frac{1}{\Phi_k(\mathbf{0}; C\boldsymbol{\mu} + \boldsymbol{d})} \quad \phi_p(\boldsymbol{\theta}, \boldsymbol{\mu}, \Sigma + \Omega)\Phi_k(\mathbf{0}; C\Delta(\Sigma^{-1}\boldsymbol{\theta} + \Omega^{-1}\boldsymbol{\mu}) + \boldsymbol{d}, C\Delta C^T), \tag{2.5}$$

where $\Delta^{-1} = \Sigma^{-1} + \Omega^{-1}$. The heirarchical skew normal density is denoted by $HSN_p(\boldsymbol{\mu}, \boldsymbol{d}, \Sigma, \Omega, C)$

The multivariate skew-normal distributions discussed so far do not cohere with the joint distribution of a random sample from a univaraite skew-normal distribution which is important for the sampling distribution theory. To overcome this drawback, Gupta and Chen (2004) provided a new definition of multivariate skew-normal distribution as given below:

Definition 2.3.4. *Let Σ be a $k \times k$ positive definite matrix. A $k \times 1$ random vector $\boldsymbol{y}$ is said to follow a multivariate skew-normal random vector if the density of $\boldsymbol{y}$ is of the form*

$$f_d(\boldsymbol{y}, \Sigma, \boldsymbol{d}) = 2^k \phi_k(\boldsymbol{y}, \Sigma) \prod_{j=1}^{k} \Phi({\boldsymbol{\lambda}_j}^T \boldsymbol{y}), \tag{2.6}$$

where $\boldsymbol{d} = (\delta_1, \delta_1, \ldots \delta_k)^T$ for some real numbers $\delta_1, \delta_2, \ldots \delta_k$ and $\boldsymbol{\lambda}_1, \boldsymbol{\lambda}_2, \ldots \boldsymbol{\lambda}_k$ are real vectors satisfying

$$\Lambda = \Sigma^{-1/2} diag(\delta_1, \delta_2, \ldots \delta_k).$$

Besides the multivariate models described so far, several other versions of multivariate skew normal distribution have been proposed in literature. Gupta et al (2004) defined a more general version of MSND that has the same property as (2.6). In addition, it takes care of the hidden truncation problem arises in skew normal distribution. Gonzalez-Farias, Dominguez-Molina, and Gupta (2004) defined a multivariate skew-normal distribution. Since this distribution is closed under the linear transformation, conditioning, summation and joint distribution of random variables from the same family, it is called closed skew normal distribution. For the details of this density we refer to Chapter 2 of Genton (2004).

2.4 Unified Skew Normal Distribution (SUN)

The study of SN class of distribution has been a resumption of interest because of two reasons: first, it opened the door for robustness study. Second, it includes the normal density, and has very similar properties as that of normal density. However, because of the popularity of this class of distribution there have been intense developments in the theory of this class of distribution. In fact, the developments are so numerous that sometimes it is confusing (especially for applied statisticians) which class of skew normal model is to be used. The other question of interest could be: is there any version which is better than the rest? or is there any generalized version that represents other models as a special case or that is at least equivalent to others up to some reparameterization? With this view in mind, Arellano-valle and Azzalini (2006) developed a skew normal model and named it unified skew normal model with the acronym SUN. They showed that this multivariate skew normal model includes or at least is equivalent to the earlier versions of skew normal models.

Definition 2.4.1. *Suppose* (U_o, U_1) *is a multivariate normal vector of dimension* $m{+}d$ *with* *the density*

$$U = \begin{pmatrix} U_0 \\ U_1 \end{pmatrix} \begin{matrix} m \\ d \end{matrix} \sim N_{m+d}(0, \Omega^*), \Omega^* = \begin{pmatrix} \Gamma & \Delta^T \\ \Delta & \bar{\Omega} \end{pmatrix},$$

where Ω^* *is the correlation matrix, and* $\Omega = \omega\bar{\Omega}\omega$ *is the covariance matrix with* ω*, a* $d \times d$ *diagonal matrix. Now, suppose* Ω^* *is positive definite and consider the distribution of* $Z = (U_1|U_0 + \boldsymbol{\gamma} > 0)$*. Then the density of* $\boldsymbol{y} = \boldsymbol{\mu} + \omega Z$ *is*

$$f(\boldsymbol{y}) = \phi_d(\boldsymbol{y} - \boldsymbol{\mu}; \Omega) \frac{\Phi_m(\boldsymbol{\gamma} + \Delta^T\bar{\Omega}^{-1}\omega^{-1}(\boldsymbol{y} - \boldsymbol{\mu}); \Gamma - \Delta^T\bar{\Omega}^{-1}\Delta)}{\Phi_m(\boldsymbol{\gamma}; \Gamma)}, \tag{2.7}$$

for $\boldsymbol{y} \in \Re^d$*. The notation* $\phi_d(\boldsymbol{y} - \boldsymbol{\mu}; \Omega)$ *is used to denote the* d *dimensional multivariate normal distribution with the mean vector* $\boldsymbol{\mu}$ *and the covariance matrix* Ω*,* $\Phi_d(\boldsymbol{y} - \boldsymbol{\mu}; \Omega)$ *denotes the corresponding distribution function. This density is called SUN (acronym for unified skew normal distribution) and is denoted by* $\boldsymbol{y} \sim SUN_{d,m}(\boldsymbol{\mu}, \boldsymbol{\gamma}, \bar{\boldsymbol{\omega}}, \Omega^*)$*, where* $\bar{\boldsymbol{\omega}} = \omega\mathbf{1}_{\boldsymbol{d}}$*.*

Note that if Δ equal to zero, then the density reduces to the d dimensional multivariate normal distribution. The derivation of the SUN density was given in Arellano-valle and Azzalini (2006). The density of univariate SUN distributon is given in figure 2.4.1.

2.4.1 Properties of SUN Density

Before describing the properties of SUN density, we state a lemma as follows which is useful for evaluating some integrals used in this chapter and some of the rest of the chapters.

Lemma 2.4.1. *Let* $\boldsymbol{V}$ *be a d-dimensional random vector with distribution* $N_d(\boldsymbol{\mu}, \Sigma)$ *where* $\boldsymbol{\mu}$ *is a* $d \times 1$ *vector and* Σ *is a positive definite matrix. Then*

$$\mathbb{E}_V\Big[\Phi_d(\boldsymbol{a} + B\boldsymbol{V}; \boldsymbol{\nu}, \Upsilon)\Big] = \Phi(\boldsymbol{a} - \boldsymbol{\nu} + B\boldsymbol{\mu}; \Upsilon + B\Sigma B^T),$$

where $\boldsymbol{a}$ *and* $\boldsymbol{\nu}$ *are d-dimensional vectors,* B *is a constant* $d \times d$ *matrix and* Υ *is a positive*

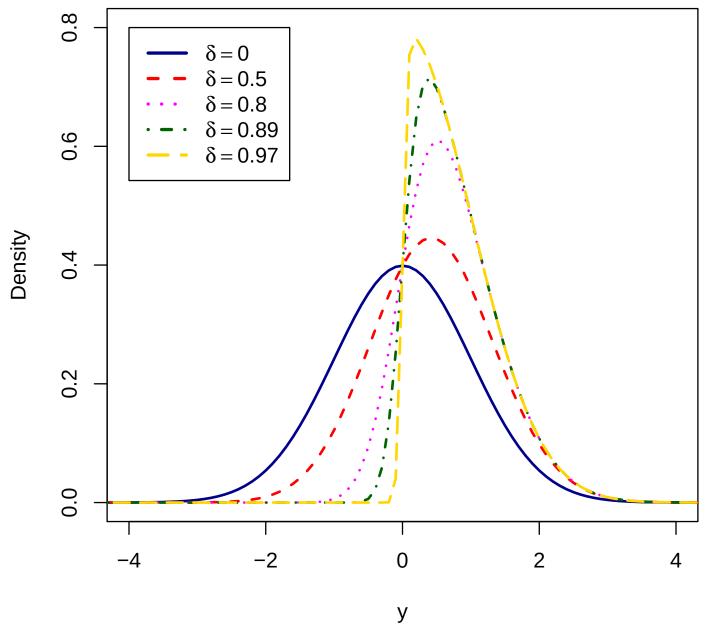

Figure 2.4.1: Univariate SUN distribution for different values of δ.

matrix of dimension $d \times d$.

Proof. For the proof of the lemma we refer to Gupta et al (2004) $\square$

Remarks: Since $\Phi_m(\boldsymbol{a}; \boldsymbol{\mu}, \Sigma) = \Phi_m(\boldsymbol{0}; \boldsymbol{\mu} - \boldsymbol{a}, \Sigma) = \Phi_m(\boldsymbol{a} - \boldsymbol{\mu}; \boldsymbol{0}, \Sigma) = \Phi_m(\boldsymbol{a} - \boldsymbol{\mu}; \Sigma)$, lemma (2.4.1) can be stated and used in a variety of ways.

2.4.1.1 The Moment Generating Function

In this section we present the moment generating function of the SUN density. In order to derive many of the most important properties of SUN distribution we need the following result.

Theorem 2.4.1. *If* $\boldsymbol{y} \sim SUN_{d,m}(\boldsymbol{\mu}, \boldsymbol{\gamma}, \bar{\boldsymbol{\omega}}, \Omega^*)$, *then its m.g.f is given by*

$$M_{\boldsymbol{y}}(\boldsymbol{t}) = exp\ (\boldsymbol{\mu}^T t + \frac{1}{2}\boldsymbol{t}^T \Omega t) \frac{\Phi_m(\boldsymbol{\gamma} + \Delta^T \omega \boldsymbol{t}; \Gamma)}{\Phi_m(\boldsymbol{\gamma}; \Gamma)}, \quad \boldsymbol{t} \in \Re^d. \tag{2.8}$$

Corollary 2.4.1. *Let $\boldsymbol{y}$ be a d dimensional random vector. Then the following results can be deduced from equation* (2.8).

- *Taking* $\Delta = \mathbf{0}$ *we get* $M_{\boldsymbol{y}}(\boldsymbol{t}) = exp\ (\boldsymbol{\mu}^T\boldsymbol{t} + \frac{1}{2}\boldsymbol{t}^T\Omega\boldsymbol{t})$ *which is the m.g.f of* $N_d(\boldsymbol{\mu}, \Omega)$, *the multivariate normal distribution.*

- *Taking m=1 ,* $\boldsymbol{\gamma} = 0$, $\Gamma = 1$ *and* $\Delta = \frac{\Sigma\boldsymbol{\alpha}}{\sqrt{(1+\boldsymbol{\alpha}^T\Sigma\boldsymbol{\alpha})}}$ *where* $\boldsymbol{\alpha}$ *is a d dimensional vector, we get*

$$M_{\boldsymbol{y}}(\boldsymbol{t}) = exp\ (\boldsymbol{\mu}^T\boldsymbol{t} + \frac{1}{2}\boldsymbol{t}^T\Omega\boldsymbol{t})\Phi\left(\frac{\Sigma\boldsymbol{\alpha}}{\sqrt{(1+\boldsymbol{\alpha}^T\Sigma\boldsymbol{\alpha})}}\right),$$

  which is the m.g.f of the distribution $SN_d(\boldsymbol{\mu}, \Omega, \boldsymbol{\alpha})$ *defined by Azzallini and Dalla Valle (1996).*

- *Taking* $\boldsymbol{\gamma} = \mathbf{0},\ \bar{\Omega} = \Sigma,\ \Delta = \Sigma D^T,\ \Gamma = \boldsymbol{I}_d + D\Sigma D^T$ *we get*

$$M_{\boldsymbol{y}}(\boldsymbol{t}) = exp\ (\boldsymbol{\mu}^T\boldsymbol{t} + \frac{1}{2}\boldsymbol{t}^T\Omega\boldsymbol{t})\frac{\Phi_d(D\Sigma t; \boldsymbol{I}_d + D\Sigma D^T)}{\Phi_d(\mathbf{0}; \boldsymbol{I}_d + D\Sigma D^T)},$$

  which is the m.g.f of the distribution $SN_d(\boldsymbol{\mu}, \Sigma, D)$ *defined by Gupta et al (2004).*

- *Taking* $\boldsymbol{\gamma} = -\boldsymbol{\nu},\ \bar{\Omega} = \Sigma,\ \Delta = \Sigma D^T,\ \Gamma = \Psi + D\Sigma D^T$ *we get*

$$M_{\boldsymbol{y}}(\boldsymbol{t}) = exp\ (\boldsymbol{\mu}^T\boldsymbol{t} + \frac{1}{2}\boldsymbol{t}^T\Sigma\boldsymbol{t})\frac{\Phi_m(D\Sigma t; \Psi + D\Sigma D^T)}{\Phi_m(\mathbf{0}; \Psi + D\Sigma D^T)},$$

  which is the m.g.f of the distribution $CSN_{d,m}(\boldsymbol{\mu}, \Sigma, D, \boldsymbol{\nu}, \Psi)$ *defined by Gonzalez-Farias, Dominguez-Molina and Gupta (2004).*

- *Taking* $\bar{\Omega} = \Upsilon + \Sigma,\ \Gamma = C\Upsilon C^T,\ \Delta = -(C\Upsilon)^T$ *and* $\boldsymbol{\gamma} = C\boldsymbol{\mu} + d$ *we get*

$$M_{\boldsymbol{y}}(\boldsymbol{t}) = exp\ (\boldsymbol{\mu}^T\boldsymbol{t} + \frac{1}{2}\boldsymbol{t}^T(\Upsilon + \Sigma)\boldsymbol{t})\frac{\Phi_m(\mathbf{0}; C\boldsymbol{\mu} + d + C\Upsilon\boldsymbol{t}, C\Upsilon C^T)}{\Phi_m(\mathbf{0}; C\boldsymbol{\mu} + d, C\Upsilon C^T)},$$

  which is the m.g.f of $HSN_d(\boldsymbol{\mu}, \boldsymbol{d}, \Sigma, \Upsilon, C)$, *the hierarchical skew-normal distribution defined by Liseo and Loperfido (2003).*

2.4.2 Cumulants and Moments

The cumulants and moments can be obtained from the m.g.f defined in (2.8). We deduce first two moments of SUN random vector. The third and fourth moments can also be derived accordingly. However, the mathematical espressions for these moments are too complex to be of any practical use. Therefore we will not derive these moments.

Mean and Variance of SUN random vector

The mean and variance for the SUN random vector is computed from (2.8) by successive differentiation with respect to the vector $\boldsymbol{t}$. Taking first derivative with respect to $\boldsymbol{t}$, we get,

$$\begin{aligned}\frac{\partial}{\partial \boldsymbol{t}} M_{\boldsymbol{y}}(\boldsymbol{t}) = (\boldsymbol{\mu} + \Omega \boldsymbol{t}) exp\ (\boldsymbol{\mu}^T \boldsymbol{t} + \frac{1}{2} \boldsymbol{t}^T \Omega \boldsymbol{t}) \frac{\Phi_m(\boldsymbol{\gamma} + \Delta^T \omega t; \Gamma)}{\Phi_m(\boldsymbol{\gamma}; \Gamma)} + \\ \frac{\Phi_m^*(\boldsymbol{\gamma} + \Delta^T \omega \boldsymbol{t}; \Gamma)}{\Phi_m(\boldsymbol{\gamma}; \Gamma)} exp\ (\boldsymbol{\mu}^T \boldsymbol{t} + \frac{1}{2} \boldsymbol{t}^T \Omega \boldsymbol{t})\end{aligned}$$

where

$$\Phi_m^*(\boldsymbol{\gamma} + \Delta^T \omega \boldsymbol{t}; \Gamma) = \frac{\partial}{\partial \boldsymbol{t}} \Phi_m(\boldsymbol{\gamma} + \Delta^T \omega \boldsymbol{t}; \Gamma)$$

and

$$\Phi_m^*(\boldsymbol{\gamma}; \Gamma) = \frac{\partial}{\partial \boldsymbol{t}} \Phi_m(\boldsymbol{\gamma} + \Delta^T \omega \boldsymbol{t}; \Gamma) \Big|_{\boldsymbol{t}=0}$$

Therefore,

$$\mathbb{E}\boldsymbol{y} = \frac{\partial}{\partial \boldsymbol{t}} M_{\boldsymbol{y}}(\boldsymbol{t}) \Big|_{\boldsymbol{t}=0} = \boldsymbol{\mu} + \frac{\Phi_m^*(\boldsymbol{\gamma}; \Gamma)}{\Phi_m(\boldsymbol{\gamma}; \Gamma)} \tag{2.9}$$

Now using lemma B.1 given in the Appendix of Dominguez-Molina et al (2001) we have

$$\Phi_m^*(\boldsymbol{\gamma}; \Gamma) = \sum_{i=1}^{d} \sum_{j=1}^{m} (\Delta^T \omega)_{ij}\ \Phi_m^{\{j\}}(\boldsymbol{\gamma}; \Gamma)\ \boldsymbol{e}_i,$$

where $(\Delta^T \omega)_{ij}$ is the (i,j) element of the matrix $\Delta^T \omega$, $\boldsymbol{e}_i$ is a $d \times 1$ vector with one in the ith position and zero elsewhere, and

$$\Phi_m^{\{j\}}(\boldsymbol{\gamma}; \Gamma) = \phi(\boldsymbol{\gamma}_j; \Gamma_{ij})\ \Phi_{m-1}(\boldsymbol{\gamma}_{-j}; \Gamma | \boldsymbol{\gamma}_j)$$

where $\boldsymbol{\gamma}_{-j}$ is the vector $\boldsymbol{\gamma}$ without the jth element.

Next taking second derivative of $M_{\boldsymbol{y}}(\boldsymbol{t})$ in (2.8) we get,

$$
\begin{aligned}
&\frac{\partial}{\partial \boldsymbol{t} \partial \boldsymbol{t}^T} M_{\boldsymbol{y}}(\boldsymbol{t}) \\
= \;& (\boldsymbol{\mu} + \Omega t)\Bigg[exp\ (\boldsymbol{\mu}^T t + \frac{1}{2} t^T \Omega t) \frac{\Phi_m^{*T}(\boldsymbol{\gamma} + \Delta^T \omega t; \Gamma)}{\Phi_m(\boldsymbol{\gamma}; \Gamma)} + exp\ (\boldsymbol{\mu}^T t + \frac{1}{2} t^T \Omega t) \times \\
& \frac{\Phi_m(\boldsymbol{\gamma} + \Delta^T \omega t; \Gamma)}{\Phi_m(\boldsymbol{\gamma}; \Gamma)} (\boldsymbol{\mu} + \Omega t)^T \Bigg] + \frac{\Phi_m(\boldsymbol{\gamma} + \Delta^T \omega t; \Gamma)}{\Phi_m(\boldsymbol{\gamma}; \Gamma)} exp\ (\boldsymbol{\mu}^T t + \frac{1}{2} t^T \Omega t) \Omega + \\
& \frac{\Phi_m^{**}(\boldsymbol{\gamma} + \Delta^T \omega t; \Gamma)}{\Phi_m(\boldsymbol{\gamma}; \Gamma)} exp\ (\boldsymbol{\mu}^T t + \frac{1}{2} t^T) + \frac{\Phi_m^{*}(\boldsymbol{\gamma} + \Delta^T \omega t; \Gamma)}{\Phi_m(\boldsymbol{\gamma}; \Gamma)} exp\ (\boldsymbol{\mu}^T t + \frac{1}{2} t^T)(\boldsymbol{\mu} + \Omega t)^T,
\end{aligned}
$$

where $\Phi_m^{*T}(\boldsymbol{\gamma} + \Delta^T \omega t; \Gamma) = \Big[\Phi_m^{*}(\boldsymbol{\gamma} + \Delta^T \omega t; \Gamma)\Big]^T$ and

$$
\Phi_m^{**}(\boldsymbol{\gamma} + \Delta^T \omega t; \Gamma) = \frac{\partial}{\partial \boldsymbol{t} \partial \boldsymbol{t}^T} (\boldsymbol{t}) \Phi_m(\boldsymbol{\gamma} + \Delta^T \omega t; \Gamma).
$$

Hence,

$$
\frac{\partial}{\partial \boldsymbol{t} \partial \boldsymbol{t}^T} M_{\boldsymbol{y}}(\boldsymbol{t}) \Big|_{\boldsymbol{t}=0} = \boldsymbol{\mu} \frac{\Phi_m^{*T}(\boldsymbol{\gamma}; \Gamma)}{\Phi_m(\boldsymbol{\gamma}; \Gamma)} + \boldsymbol{\mu}\boldsymbol{\mu}^T + \Omega + \frac{\Phi_m^{**}(\boldsymbol{\gamma}; \Gamma)}{\Phi_m(\boldsymbol{\gamma}; \Gamma)} + \frac{\Phi_m^{**}(\boldsymbol{\gamma}; \Gamma)}{\Phi_m(\boldsymbol{\gamma}; \Gamma)} \boldsymbol{\mu}^T.
$$

and we get,

$$
\mathbb{E}\boldsymbol{y}\boldsymbol{y}^T = \Omega + \boldsymbol{\mu}\boldsymbol{\mu}^T + \boldsymbol{\mu} \frac{\Phi_m^{*T}(\boldsymbol{\gamma}; \Gamma)}{\Phi_m(\boldsymbol{\gamma}; \Gamma)} + \frac{\Phi_m^{**}(\boldsymbol{\gamma}; \Gamma)}{\Phi_m(\boldsymbol{\gamma}; \Gamma)} \boldsymbol{\mu}^T + \frac{\Phi_m^{**}(\boldsymbol{\gamma}; \Gamma)}{\Phi_m(\boldsymbol{\gamma}; \Gamma)}. \tag{2.10}
$$

Finally from (2.9) and (2.10) the variance of the SUN density is

$$
\begin{aligned}
& Var(\boldsymbol{y}) = \mathbb{E}\boldsymbol{y}\boldsymbol{y}^T - (\mathbb{E}\boldsymbol{y})(\mathbb{E}\boldsymbol{y})^T \\
= \;& \Omega + \boldsymbol{\mu}\boldsymbol{\mu}^T + \boldsymbol{\mu} \frac{\Phi_m^{*T}(\boldsymbol{\gamma}; \Gamma)}{\Phi_m(\boldsymbol{\gamma}; \Gamma)} + \frac{\Phi_m^{**}(\boldsymbol{\gamma}; \Gamma)}{\Phi_m(\boldsymbol{\gamma}; \Gamma)} \boldsymbol{\mu}^T + \frac{\Phi_m^{**}(\boldsymbol{\gamma}; \Gamma)}{\Phi_m(\boldsymbol{\gamma}; \Gamma)} - \Big[\boldsymbol{\mu} + \frac{\Phi_m^{*}(\boldsymbol{\gamma}; \Gamma)}{\Phi_m(\boldsymbol{\gamma}; \Gamma)}\Big] \Big[(\boldsymbol{\mu} + \frac{\Phi_m^{*}(\boldsymbol{\gamma}; \Gamma)}{\Phi_m(\boldsymbol{\gamma}; \Gamma)}\Big]^T \\
= \;& \Omega + \frac{\Phi_m^{**}(\boldsymbol{\gamma}; \Gamma)}{\Phi_m(\boldsymbol{\gamma}; \Gamma)} - \frac{\Phi_m^{*}(\boldsymbol{\gamma}; \Gamma)}{\Phi_m(\boldsymbol{\gamma}; \Gamma)} \frac{\Phi_m^{*T}(\boldsymbol{\gamma}; \Gamma)}{\Phi_m(\boldsymbol{\gamma}; \Gamma)}.
\end{aligned}
$$

The variance of the SUN density can also be expressed as

$$Var(\boldsymbol{y}) = \Omega + \frac{\Phi_m^{**}(\boldsymbol{\gamma};\Gamma)}{\Phi_m(\boldsymbol{\gamma};\Gamma)} - \mathbb{E}(\boldsymbol{y}-\boldsymbol{\mu})\mathbb{E}(\boldsymbol{y}-\boldsymbol{\mu})^T, \tag{2.11}$$

where

$$\Phi_m^{**}(\boldsymbol{\gamma};\Gamma) = \frac{\partial}{\partial \boldsymbol{t}\partial \boldsymbol{t}^T}\Phi_m(\boldsymbol{\gamma}+\Delta^T\omega\boldsymbol{t};\Gamma)\Big|_{\boldsymbol{t}=0}.$$

Example 2.4.1. *Consider the univariate SUN density $SUN_{1,1}(\mu,0,w,\Omega^*)$, where*

$$\Omega^* = \begin{pmatrix} g^2 & \delta \\ \delta & v \end{pmatrix}$$

is the correlation matrix. Then the corresponding density is given by

$$f(y) = 2\phi(y;\mu,\sigma^2)\Phi(\delta v^{-1}w^{-1}(y-\mu);g^2-\delta^2/v^{-1}).$$

The mean of this density from (2.9) is

$$\mathbb{E}y = \mu + 2\Phi_1^*(0;g^2) = \mu + 2\ \delta w\frac{1}{\sqrt{2\pi}g} = \mu + \frac{\delta w}{g}\sqrt{\frac{2}{\pi}}, \tag{2.12}$$

where $\Phi_1^*(0;g^2) = \frac{\partial}{\partial t}\Phi_1(\delta;g^2)\Big|_{t=0} = \delta w\phi(0;g^2) = \delta w\frac{1}{\sqrt{2\pi}g}$.

The variance of the density from (2.11) is

$$Var(y) = \sigma^2 + 2\Phi_1^{**}(0;g^2) - (2\Phi_1^*(0;g^2))^2 = \sigma^2 - \left(\frac{\delta w}{g}\sqrt{\frac{2}{\pi}}\right)^2 = \sigma^2 - \frac{2}{\pi}\ \frac{\delta^2\ w^2}{g^2}, \tag{2.13}$$

where

$$\Phi_1^{**}(0;g^2) = \frac{\partial}{\partial^2 t}\Phi_1(\delta wt;g^2)\Big|_{t=0}.$$

Note: *Taking $\delta = 0$ we get the mean and variance of the $MND_1(\mu,\sigma^2)$.*

Cumulants and moments in a special case

As mentioned in the previous section deriving moments from (2.8) could be cumbersome. The difficulty arises because of the absence of the analytical representations for the derivative of $\Phi_m(.;.)$. The derivation of cumulants and moments is simplified when Γ is taken as $\Gamma = diag(\tau_1^2, ..., \tau_m^2)$. With Γ defined as $diag(\tau_1^2, ..., \tau_m^2)$, the cumulant generating function from (2.1) is given as

$$K(\boldsymbol{t}) = logM(\boldsymbol{t}) = \boldsymbol{\mu}^T t + \frac{1}{2}\boldsymbol{t}^T\Omega\boldsymbol{t} + \sum_{j=1}^{m} log\Phi(\tau_j^{-1}\gamma_j + \tau_j^{-1}\boldsymbol{\delta}_{.j}^T\omega\boldsymbol{t}) - log\Phi(\gamma;\Gamma),$$

where $\boldsymbol{\delta}_{.1}, ..., \boldsymbol{\delta}_{.m}$ are the columns of Δ. Now taking derivative with respect to $\boldsymbol{t} \in \Re^d$ we get:

$$K^{'}(\boldsymbol{t}) = \boldsymbol{\mu} + \Omega\boldsymbol{t} + \sum_{j=1}^{m} \Phi^{-1}(\tau_j^{-1}\gamma_j + \tau_j^{-1}\boldsymbol{\delta}_{.j}^T\omega\boldsymbol{t})\phi(\tau_j^{-1}\gamma_j + \tau_j^{-1}\boldsymbol{\delta}_{.j}^T\omega\boldsymbol{t})\tau_j^{-1}\omega\boldsymbol{\delta}_{.j}.$$

Hence the first cumulant (or first raw moment) or mean of SUN distribution is obtained as

$$\kappa_1 = M_1 = \mathbb{E}(\boldsymbol{y}) = K^{'}(0) = \boldsymbol{\mu} + \sum_{j=1}^{m} \Phi^{-1}(\tau_j^{-1}\gamma_j)\phi(\tau_j^{-1}\gamma_j)\tau_j^{-1}\omega\boldsymbol{\delta}_{.j}.$$

Next, taking second derivative we get:

$$K^{''}(\boldsymbol{t}) = \Omega + \sum_{j=1}^{m} \frac{\Phi(\tau_j^{-1}\gamma_j + \tau_j^{-1}\boldsymbol{\delta}_{.j}^T\omega\boldsymbol{t})\phi^{'}(\tau_j^{-1}\gamma_j + \tau_j^{-1}\boldsymbol{\delta}_{.j}^T\omega\boldsymbol{t}) - [\phi(\tau_j^{-1}\gamma_j + \tau_j^{-1}\boldsymbol{\delta}_{.j}^T\omega\boldsymbol{t})]^2}{[\Phi(\tau_j^{-1}\gamma_j + \tau_j^{-1}\boldsymbol{\delta}_{.j}^T\omega\boldsymbol{t})]^2}\tau_j^{-2}\omega\boldsymbol{\delta}_{.j}\boldsymbol{\delta}_{.j}^T\omega.$$

Therefore, the 2nd cumulant(or 2nd central moment) or variance of SUN is

$$\kappa_2 = m_2 = var(\boldsymbol{y}) = K^{''}(0) = \Omega + \sum_{j=1}^{m} \frac{\Phi(\tau_j^{-1}\gamma_j)\phi^{'}(\tau_j^{-1}\gamma_j) - [\phi(\tau_j^{-1}\gamma_j]^2}{[\Phi(\tau_j^{-1}\gamma_j)]^2}\tau_j^{-2}\omega\boldsymbol{\delta}_{.j}\boldsymbol{\delta}_{.j}^T\omega.$$

Now suppose, $\zeta_r(x)$ is the rth derivative of $\zeta_0(x) = log\{\Phi(x)\}$. Then the mean and variance of SUN density can be expressed respectively as:

$$\mathbb{E}(\boldsymbol{y}) = \boldsymbol{\mu} + \sum_{j=1}^{m} \zeta_1(\tau_j^{-1}\gamma_j)\tau_j^{-1}\omega\boldsymbol{\delta}_{.j}, \tag{2.14}$$

and

$$Var(\boldsymbol{y}) = \Omega + \sum_{j=1}^{m} \zeta_2(\tau_j^{-1}\gamma_j)\tau_j^{-2}\omega\boldsymbol{\delta}_{.j}\boldsymbol{\delta}_{.j}^T\omega. \tag{2.15}$$

Thus we have the following two theorems:

Theorem 2.4.2. *Let $\boldsymbol{y}$ be a random vector with a unified skew normal distribution, $\boldsymbol{y} \sim SUN_{d,m}(\boldsymbol{\mu}, \gamma, \bar{\boldsymbol{\omega}}, \Omega^*)$, where*

$$\Omega^* = \begin{pmatrix} \Gamma & \Delta^T \\ \Delta & \bar{\Omega} \end{pmatrix}.$$

Suppose $\Gamma = diag(\tau_1^2, \ldots.., \tau_m^2)$. Then the first two central moments of y are:

$$(a)\ m_1 = \sum_{j=1}^{m} \zeta_1(\tau_j^{-1}\gamma_j)\tau_j^{-1}\omega\boldsymbol{\delta}_{.j}$$

$$(b)\ m_2 = \Omega + \sum_{j=1}^{m} \zeta_2(\tau_j^{-1}\gamma_j)\tau_j^{-2}\omega\boldsymbol{\delta}_{.j}\boldsymbol{\delta}_{.j}^T\omega.$$

Corollary 2.4.2. *Let $\boldsymbol{y}$ be a random vector with a unified skew normal distribution, $y \sim SUN_{d,m}(\boldsymbol{\mu}, \mathbf{0}, \bar{\boldsymbol{\omega}}, \Omega^*)$, where*

$$\Omega^* = \begin{pmatrix} I_m & \Delta^T \\ \Delta & \bar{\Omega} \end{pmatrix}.$$

Then the first two central moments of $\boldsymbol{y}$ are:

$$(a)\ m_1 = \sqrt{\frac{2}{\pi}}\omega\Delta\mathbf{1}_m.$$

$$(b)\ m_2 = \Omega - \frac{2}{\pi}\omega\Delta\Delta^T\omega.$$

Proof. With $\boldsymbol{\gamma} = \mathbf{0}$ and $\Gamma = I_m$, it could be easily shown that $\zeta_1(\tau_j^{-1}\gamma_j) = \sqrt{\frac{2}{\pi}}$ and $\zeta_2(\tau_j^{-1}\gamma_j) = -\frac{2}{\pi}$. The proof then follows from plugging in these values in Theorem 2.4.2. □

Theorem 2.4.3. *Let* $\boldsymbol{y}$ *be a random vector with a unified skew normal distribution,* $\boldsymbol{y} \sim SUN_{d,m}(\boldsymbol{\mu}, \gamma, \bar{\boldsymbol{\omega}}, \Omega^*)$, *where*

$$\Omega^* = \begin{pmatrix} \Gamma & \Delta^T \\ \Delta & \bar{\Omega} \end{pmatrix}.$$

Suppose $\Gamma = diag(\tau_1^2, ..., \tau_m^2)$. *Then the first two raw moments of* $\boldsymbol{y}$ *are:*

$$(a) \quad M_1 = \boldsymbol{\mu} + \sum_{j=1}^{m} \zeta_1(\tau_j^{-1}\gamma_j)\tau_j^{-1}\omega\boldsymbol{\delta}_{.j}.$$

$$(b) \quad M_2 = \Omega + \sum_{j=1}^{m} \zeta_2(\tau_j^{-1}\gamma_j)\tau_j^{-2}\omega\boldsymbol{\delta}_{.j}\boldsymbol{\delta}_{.j}^T\omega + (\boldsymbol{\mu} + \sum_{j=1}^{m} \zeta_1(\tau_j^{-1}\gamma_j)\tau_j^{-1}\omega\boldsymbol{\delta}_{.j})$$
$$(\boldsymbol{\mu} + \sum_{j=1}^{m} \zeta_1(\tau_j^{-1}\gamma_j)\tau_j^{-1}\omega\boldsymbol{\delta}_{.j})^T.$$

Note that taking $\boldsymbol{\delta}_{.j} = 0$ for all j in (a) and (b) we obtain the moments of multivariate normal density(MND).

Corollary 2.4.3. *Let* $\boldsymbol{y}$ *be a random vector with a unified skew normal distribution,* $y \sim SUN_{d,m}(\boldsymbol{\mu}, \mathbf{0}, \bar{\boldsymbol{\omega}}, \Omega^*)$, *where*

$$\Omega^* = \begin{pmatrix} I_m & \Delta^T \\ \Delta & \bar{\Omega} \end{pmatrix}.$$

Then the first two raw moments of $\boldsymbol{y}$ *are:*

$$(a) \quad M_1 = \boldsymbol{\mu} + \sqrt{\frac{2}{\pi}}\omega\Delta\mathbf{1}_m.$$

$$(b) \quad M_2 = \Omega - \frac{2}{\pi}\omega\Delta\Delta^T\omega + \left(\boldsymbol{\mu} + \sqrt{\frac{2}{\pi}}\omega\Delta\mathbf{1}_m\right)\left(\boldsymbol{\mu} + \sqrt{\frac{2}{\pi}}\omega\Delta\mathbf{1}_m\right)^T.$$

Proof. The proof follows from Corollary 2.4.2 using the relationship between raw moments and central moments. □

2.4.3 Linear Transformation

The SUN density is closed under translations and sclar multiplications. Thus

- If $\boldsymbol{y} \sim SUN_{d,m}(\boldsymbol{\mu}, \boldsymbol{\gamma}, \bar{\boldsymbol{\omega}}, \Omega^*)$ and $\boldsymbol{a}$ be a real vector of dimension d, then

$$\boldsymbol{y} + \boldsymbol{a} \sim SUN_{d,m}(\boldsymbol{\mu} + \boldsymbol{a}, \boldsymbol{\gamma}, \bar{\boldsymbol{\omega}}, \Omega^*).$$

- If $c \in \Re$ then

$$c\boldsymbol{y} \sim SUN_{d,m}(c\boldsymbol{\mu}, \boldsymbol{\gamma}, \boldsymbol{c}\bar{\boldsymbol{\omega}}, \Omega^*).$$

The above two results can be easily verified by m.g.f given in (2.8). Next we establish two most important properties of SUN density namely:

- The SUN family is closed under the full row rank linear transformation
- The SUN family is closed under the full column rank linear transformation (defining singular SUN density).

The first property, the closure under the linear transformations property is useful to establish joint distribution of the independent random variables from the same family and to establish closure under marginalization, conditional distribution and summation properties of SUN density.

Theorem 2.4.4. *Let $\boldsymbol{y} \sim SUN_{d,m}(\boldsymbol{\mu}, \boldsymbol{\gamma}, \bar{\boldsymbol{\omega}}, \Omega^*)$ and A be an $n \times d$ ($n \leq d$) matrix with rank n. Then*

$$A\boldsymbol{y} \sim SUN_{n,m}(\boldsymbol{\mu}_A, \boldsymbol{\gamma}, \bar{\omega}_A, \Omega_A^*),$$

where

$$\boldsymbol{\mu}_A = A\boldsymbol{\mu}, \quad \bar{\boldsymbol{\omega}}_A = \omega_A 1_n, \quad \omega_A = A\omega A^T \quad \Omega_A^* = \begin{pmatrix} \Gamma & \Delta_A^T \\ \Delta_A & \bar{\Omega}_A \end{pmatrix}, \quad \Delta_A = (A\omega A^T)^{-1} A\omega\Delta,$$
$$\bar{\Omega}_A = (A\omega A^T)^{-1} A\Omega A^T (A\omega A^T)^{-1} \quad \text{and} \quad \Omega_A = A\Omega A^T = \omega_A \bar{\Omega}_A \omega_A.$$

Proof. For $\boldsymbol{t} \in \Re^n$ the m.g.f of $A\boldsymbol{y}$ is given by:

$$M_{A\boldsymbol{y}}(\boldsymbol{t}) = M_{\boldsymbol{y}}(A^T\boldsymbol{t}) = exp\,(\boldsymbol{\mu}^T A^T \boldsymbol{t} + \frac{1}{2}\boldsymbol{t}^T A\Omega A^T \boldsymbol{t})\frac{\Phi_m(\boldsymbol{\gamma} + \Delta^T \omega A^T \boldsymbol{t}; \Gamma)}{\Phi_m(\boldsymbol{\gamma}; \Gamma)}.$$

Now by noting that:

$$\Phi_m(\boldsymbol{\gamma} + \Delta^T \omega A^T \boldsymbol{t}; \Gamma) = \Phi_m\Big(\boldsymbol{\gamma} + \big((A\omega A^T)^{-1} A\omega\Delta\big)^T (A\omega A^T)\boldsymbol{t}; \Gamma\Big)$$

and using $\boldsymbol{\mu}_A$, $\bar{\omega}_A$, Ω_A, Δ_A, and Ω^*_A as defined above, we obtain:

$$M_{A\boldsymbol{y}}(\boldsymbol{t}) = exp\,(\boldsymbol{\mu}_A^T \boldsymbol{t} + \frac{1}{2}\boldsymbol{t}^T \Omega_A \boldsymbol{t})\frac{\Phi_m(\boldsymbol{\gamma} + \Delta_A^T \omega_A t; \Gamma)}{\Phi_m(\boldsymbol{\gamma}; \Gamma)}, \tag{2.16}$$

which is the m.g.f of $SUN_{n,m}(\boldsymbol{\mu}_A, \boldsymbol{\gamma}, \bar{\boldsymbol{\omega}}_A, \Omega^*_A)$. □

Remark 2.4.1. *If $n = 1$ in Theorem 2.4.4 and if $\boldsymbol{a}$ is a non-zero vector in $\Re^d$ then*

$$\boldsymbol{a}^T \boldsymbol{y} \sim SUN_{1,m}(\mu_{\boldsymbol{a}}, \boldsymbol{\gamma}, \bar{\omega}_{\boldsymbol{a}}, \Omega^*_{\boldsymbol{a}}),$$

where

$$\mu_{\boldsymbol{a}} = \boldsymbol{a}^T \boldsymbol{\mu}, \quad \bar{\omega}_{\boldsymbol{a}} = \boldsymbol{a}^T \omega \boldsymbol{a}, \quad \Delta_{\boldsymbol{a}} = (\boldsymbol{a}^T \omega \boldsymbol{a})^{-1} \boldsymbol{a}^T \omega \Delta,$$

$$\bar{\Omega}_{\boldsymbol{a}} = (\boldsymbol{a}^T \omega \boldsymbol{a})^{-1} \boldsymbol{a} \Omega \boldsymbol{a}^T (\boldsymbol{a}^T \omega \boldsymbol{a})^{-1}, \quad \Omega_{\boldsymbol{a}} = \boldsymbol{a}^T \Omega \boldsymbol{a}, \quad \textit{and} \quad \Omega^*_{\boldsymbol{a}} = \begin{pmatrix} \Gamma & \Delta_{\boldsymbol{a}}^T \\ \Delta_{\boldsymbol{a}} & \bar{\Omega}_{\boldsymbol{a}} \end{pmatrix}.$$

As mentioned before, the result in Theorem 2.4.4 remains true for most of the earlier versions of skew normal models. As an example, the following corollary shows that closed skew normal distribution (CSN) when written as a special case of SUN distribution, is closed under the linear transformation.

Corollary 2.4.4. *With the reparameterization scheme $\boldsymbol{\gamma} = -\boldsymbol{\nu}$, $\bar{\Omega} = \Sigma$, $\Delta = \Sigma D^T$, $\Gamma =$*

$\Psi + D\Sigma D^T$,

$$A\boldsymbol{y} \sim CSN_{n,m}(\boldsymbol{\mu}_A, \Sigma_A, D_A, \boldsymbol{\nu}, \Psi_A),$$

where $\boldsymbol{y} \sim CSN_{n,m}(\boldsymbol{\mu}, \Sigma, D, \boldsymbol{\nu}, \Psi)$ *and* $\boldsymbol{\mu}_A = A\boldsymbol{\mu}, \quad \Sigma_A = A\Sigma A^T, \quad D_A = D\Sigma A^T(A\Sigma A^T)^{-1}$ *and*

$$\Psi_A = \Psi + D\Sigma D^T - D\Sigma A^T(A\Sigma A^T)^{-1}(A\Sigma D^T).$$

Proof. Using $\boldsymbol{\mu}_A, \Sigma_A, D_A, \boldsymbol{\nu}, \Psi_A$ as defined above and taking $\boldsymbol{\gamma} = -\boldsymbol{\nu},\ \bar{\Omega} = \Omega = \Sigma$ that is $\omega = I_d,\ \Delta = \Sigma D^T,\ \Gamma = \Psi + D\Sigma D^T$ in (2.16) we get,

$$M_{A\boldsymbol{y}}(\boldsymbol{t}) = exp(\boldsymbol{\mu}^T A^T \boldsymbol{t} + \frac{1}{2}\boldsymbol{t}^T A\Sigma A^T \boldsymbol{t})\frac{\Phi_m(D\Sigma A^T \boldsymbol{t}; \nu, \Psi + D\Sigma D^T)}{\Phi_m(\mathbf{0}; \nu, \Psi + D\Sigma D^T)}. \tag{2.17}$$

Now the numerator in the fraction of the equation(2.17) can be written as,

$$\begin{aligned}
& \Phi_m(D\Sigma A^T(A\Sigma A^T)^{-1}(A\Sigma A^T)\boldsymbol{t}; \boldsymbol{\nu}, \Psi + D\Sigma D^T - D\Sigma A^T(A\Sigma A^T)^{-1}(A\Sigma D^T) + \\
& D\Sigma A^T(A\Sigma A^T)^{-1}(A\Sigma A^T)(A\Sigma A^T)^{-1}(A\Sigma D^T)) \\
= \; & \Phi_m(D_A\Sigma_A \boldsymbol{t}; \boldsymbol{\nu}, \Psi_A + D_A\Sigma_A D_A^T).
\end{aligned}$$

and the denominator can be written as

$$\begin{aligned}
& \Phi_m(\mathbf{0}; \nu, \Psi + D\Sigma D^T - D\Sigma A^T(A\Sigma A^T)^{-1}(A\Sigma D^T) + \\
& D\Sigma A^T(A\Sigma A^T)^{-1}(A\Sigma A^T)(A\Sigma A^T)^{-1}(A\Sigma D^T)) \\
= \; & \Phi_m(\mathbf{0}; \boldsymbol{\nu}, \Psi_A + D_A\Sigma_A D_A^T).
\end{aligned}$$

Therefore (2.17) becomes

$$M_{A\boldsymbol{y}}(\boldsymbol{t}) = exp(\mu_A^T \boldsymbol{t} + \frac{1}{2}\boldsymbol{t}^T \Sigma_A \boldsymbol{t})\frac{\Phi_m(D_A\Sigma_A \boldsymbol{t}; \boldsymbol{\nu}, \Psi_A + D_A\Sigma_A D_A^T)}{\Phi_m(\mathbf{0}; \boldsymbol{\nu}, \Psi_A + D_A\Sigma_A D_A^T)},$$

which is the m.g.f of $CSN_{n,m}(\boldsymbol{\mu}_A, \Sigma_A, D_A, \boldsymbol{\nu}, \Psi_A)$. $\square$

Theorem 2.4.5. *(The singular SUN distribution)*

Let $\boldsymbol{y} \sim SUN_{d,m}(\boldsymbol{\mu}, \boldsymbol{\gamma}, \bar{\boldsymbol{\omega}}, \Omega^*)$ *and* A *be an* $n \times d$ $(n > d)$ *matrix with rank* d*. Then*

$$A\boldsymbol{y} \sim SUN_{n,m}(\boldsymbol{\mu}_A, \boldsymbol{\gamma}, \bar{\omega}_A, \Omega^*_A),$$

where

$$\boldsymbol{\mu}_A = A\boldsymbol{\mu}, \quad \Omega^*_A = \begin{pmatrix} \Gamma & \Delta_A^T \\ \Delta_A & \bar{\Omega}_A \end{pmatrix}, \quad \Delta_A = A(A^TA)^{-1}\Delta$$

and

$$\omega_A = A\omega A^T, \quad \bar{\omega}_A = \omega_A 1_d, \quad \Omega_A = A\Omega A^T = \omega_A \bar{\Omega}_A \omega_A.$$

Proof. For $\boldsymbol{t} \in \Re^n$ the m.g.f of $A\boldsymbol{y}$ is given by:

$$\begin{aligned} M_{A\boldsymbol{y}}(\boldsymbol{t}) &= M_{\boldsymbol{y}}(A^T\boldsymbol{t}) \\ &= exp\,(\boldsymbol{\mu}^T A^T \boldsymbol{t} + \frac{1}{2}\boldsymbol{t}^T A\Omega A^T \boldsymbol{t}) \frac{\Phi_m(\boldsymbol{\gamma} + \Delta^T \omega A^T \boldsymbol{t}; \Gamma)}{\Phi_m(\boldsymbol{\gamma}; \Gamma)}. \end{aligned}$$

By noting that,

$$\Phi_m(\boldsymbol{\gamma} + \Delta^T \omega A^T \boldsymbol{t}; \Gamma) = \Phi_m\Big(\boldsymbol{\gamma} + \big(A(A^TA)^{-1}\Delta\big)^T (A\omega A^T)\boldsymbol{t}; \Gamma\Big)$$

and using $\boldsymbol{\mu}_A$, Ω^*_A, Δ_A, and Ω_A as defined above we obtain:

$$M_{A\boldsymbol{y}}(\boldsymbol{t}) = exp\,(\boldsymbol{\mu}_A^T \boldsymbol{t} + \frac{1}{2} t^T \Omega_A \boldsymbol{t}) \frac{\Phi_m(\boldsymbol{\gamma} + \Delta_A^T \omega_A \boldsymbol{t}; \Gamma)}{\Phi_m(\boldsymbol{\gamma}; \Gamma)}, \tag{2.18}$$

which is the m.g.f of singular $SUN_{n,m}(\boldsymbol{\mu}_A, \boldsymbol{\gamma}, \bar{\omega}_A, \Omega^*_A)$. □

Corollary 2.4.5. *(Singular closed skew normal density) With the reparameterization scheme* $\boldsymbol{\gamma} = -\boldsymbol{\nu}$, $\bar{\Omega} = \Sigma$, $\Delta = \Sigma D^T$, $\Gamma = \Psi + D\Sigma D^T$,

$$A\boldsymbol{y} \sim CSN_{n,m}(\boldsymbol{\mu}_A, \Sigma_A, D_A, \boldsymbol{\nu}, \Psi),$$

where $\boldsymbol{y} \sim CSN_{n,m}(\boldsymbol{\mu}, \Sigma, D, \boldsymbol{\nu}, \Psi)$ *and* $\boldsymbol{\mu}_A = A\boldsymbol{\mu}, \quad \Sigma_A = A\Sigma A^T, \quad D_A = D(A^T A)^{-1}A^T.$

Proof. Using $\boldsymbol{\mu}_A, \Sigma_A, D_A$ as defined above and taking $\boldsymbol{\gamma} = -\boldsymbol{\nu},\ \bar{\Omega} = \Omega = \Sigma$ that is $\omega = I_d,\ \Delta = \Sigma D^T,\ \Gamma = \Psi + D\Sigma D^T$ in (2.19) we get,

$$
\begin{aligned}
& M_{A\boldsymbol{y}}(\boldsymbol{t}) \\
= \; & exp(\boldsymbol{\mu}^T A^T \boldsymbol{t} + \frac{1}{2}\boldsymbol{t}^T A\Sigma A^T \boldsymbol{t}) \frac{\Phi_m\Big(-\boldsymbol{\nu} + \big(A(A^TA)^{-1}\Sigma D^T\big)^T AA^T\boldsymbol{t}; \Psi + D\Sigma D^T\Big)}{\Phi_m(-\nu; \Psi + D\Sigma D^T)} \\
= \; & exp(\boldsymbol{\mu}^T A^T \boldsymbol{t} + \frac{1}{2}\boldsymbol{t}^T A\Sigma A^T \boldsymbol{t}) \frac{\Phi_m\Big(D\Sigma A^T\boldsymbol{t}; \boldsymbol{\nu}, \Psi + D\Sigma D^T\Big)}{\Phi_m(\mathbf{0}; \boldsymbol{\nu}, \Psi + D\Sigma D^T)} \\
= \; & exp(\boldsymbol{\mu}^T A^T \boldsymbol{t} + \frac{1}{2}\boldsymbol{t}^T A\Sigma A^T \boldsymbol{t}) \quad \times \\
& \frac{\Phi_m\Big(D(A^TA)^{-1}A(A\Sigma A^T)\boldsymbol{t}; \boldsymbol{\nu}, \Psi + D(A^TA)^{-1}A^T(A\Sigma A^T)A(A^TA)^{-1}D^T\Big)}{\Phi_m(\mathbf{0}; \boldsymbol{\nu}, \Psi + D(A^TA)^{-1}A^T(A\Sigma A^T)A(A^TA)^{-1}D^T)} \\
= \; & exp(\mu_A^T\boldsymbol{t} + \frac{1}{2}\boldsymbol{t}^T\Sigma_A\boldsymbol{t}) \frac{\Phi_m(D_A\Sigma_A\boldsymbol{t}; \boldsymbol{\nu}, \Psi + D_A\Sigma_A D_A^T)}{\Phi_m(\mathbf{0}; \boldsymbol{\nu}, \Psi + D_A\Sigma_A D_A^T)},
\end{aligned}
$$

which is the m.g.f of singular $CSN_{n,m}(\boldsymbol{\mu}_A, \Sigma_A, D_A, \boldsymbol{\nu}, \Psi)$. □

2.4.4 Characterization

In the following theorem a characterization for the unified multivariate skew normal distribution is provided which is similar to the characterization of multivariate normal distribution.

Theorem 2.4.6. *The vector* $\boldsymbol{y} \sim SUN_{d,m}(\boldsymbol{\mu}, \boldsymbol{\gamma}, \bar{\boldsymbol{\omega}}, \Omega^*)$ *if, and only if ,* $\boldsymbol{a}^T\boldsymbol{y} \sim SUN_{1,m}(\mu_{\boldsymbol{a}}, \boldsymbol{\gamma}, \bar{\omega}_{\boldsymbol{a}}, \Omega^*_{\boldsymbol{a}})$ *, for every non-null vector* $\boldsymbol{a} \in \Re^d$*, where* $\mu_{\boldsymbol{a}}, \boldsymbol{\gamma}, \bar{\omega}_{\boldsymbol{a}}, \Omega^*_{\boldsymbol{a}}$ *are given in Remark 2.4.1.*

Proof. We will only prove the sufficiency since the proof of the necessity is straightforward. Note that, if $\boldsymbol{a}^T\boldsymbol{y} \sim SUN_{1,m}(\mu_{\boldsymbol{a}}, \boldsymbol{\gamma}, \bar{\omega}_{\boldsymbol{a}}, \Omega^*_{\boldsymbol{a}})$ for every non-zero vector $\boldsymbol{a}$, then for $t \in \Re$ using (2.8) we obtain,

$$
M_{\boldsymbol{a}^T\boldsymbol{y}}(t) = exp\ (\mu_{\boldsymbol{a}}t + \frac{1}{2}t^2\Omega_{\boldsymbol{a}}) \frac{\Phi_m(\boldsymbol{\gamma} + \Delta^T_{\boldsymbol{a}}\omega_{\boldsymbol{a}}t; \Gamma)}{\Phi_m(\boldsymbol{\gamma}; \Gamma)}.
$$

Taking $t = 1$ and using the identity $M_{\boldsymbol{a}^T\boldsymbol{y}}(t) = M_{\boldsymbol{y}}(\boldsymbol{a}t)$, we get

$$M_{\boldsymbol{y}}(\boldsymbol{a}) = exp\ (\mu_{\boldsymbol{a}} + \frac{1}{2}\Omega_{\boldsymbol{a}})\frac{\Phi_m(\boldsymbol{\gamma} + \Delta_{\boldsymbol{a}}^T\omega_{\boldsymbol{a}}; \Gamma)}{\Phi_m(\boldsymbol{\gamma}; \Gamma)}.$$

Using $\mu_{\boldsymbol{a}},\ \omega_{\boldsymbol{a}},\ \Omega_{\boldsymbol{a}}$, as defined in remark 2.4.1 and observing that

$$\Delta_{\boldsymbol{a}}^T\omega_{\boldsymbol{a}} = \Delta^T\omega\boldsymbol{a}(\boldsymbol{a}^T\omega\boldsymbol{a})^{-1}(\boldsymbol{a}^T\omega\boldsymbol{a}) = \Delta^T\omega\boldsymbol{a},$$

the expression for $M_{\boldsymbol{y}}(\boldsymbol{a})$ reduces to

$$M_{\boldsymbol{y}}(\boldsymbol{a}) = exp\ (\boldsymbol{a}^T\boldsymbol{\mu} + \frac{1}{2}\boldsymbol{a}^T\Omega\boldsymbol{a})\frac{\Phi_m(\boldsymbol{\gamma} + \Delta^T\omega\boldsymbol{a}; \Gamma)}{\Phi_m(\boldsymbol{\gamma}; \Gamma)}. \tag{2.19}$$

Since $\boldsymbol{a}$ is arbitrary, the right-hand side of (3.4) is then the m.g.f of $\boldsymbol{y} \sim SUN_{d,m}(\boldsymbol{\mu}, \boldsymbol{\gamma}, \bar{\boldsymbol{\omega}}, \Omega^*)$ and by (2.8) the proof is complete. $\square$

2.4.5 Marginal and Conditional Distributions

The SUN family is closed under the marginalization, conditional distributions and joint distribution of independent random variables in this family. These three results are stated and established as follows:

Theorem 2.4.7. *Let $\boldsymbol{y}$ be a random vector distributed as $SUN_{d,m}(\boldsymbol{\mu}, \boldsymbol{\gamma}, \bar{\boldsymbol{\omega}}, \Omega^*)$ and be partitioned as $\boldsymbol{y} = \begin{array}{c}\begin{pmatrix} \boldsymbol{y}_1 \\ \boldsymbol{y}_2 \end{pmatrix}\end{array}\begin{array}{c} k \\ d-k \end{array}$. Consider a $k \times d$ matrix $A = (I_k \quad 0)$, with a $k \times k$ identity matrix I_k and a $k \times (d-k)$ zero matrix 0. Then the marginal distribution of $\boldsymbol{y}_1 = A\boldsymbol{y}$ is $SUN_{k,m}(\boldsymbol{\mu}_1, \boldsymbol{\gamma}, \bar{\omega}_1, \Omega_1^*)$,*

where $\boldsymbol{\mu} = \begin{pmatrix} \boldsymbol{\mu}_1 \\ \boldsymbol{\mu}_2 \end{pmatrix}\begin{array}{c} k \\ d-k \end{array},\quad \bar{\boldsymbol{\omega}} = \begin{pmatrix} \bar{\boldsymbol{\omega}}_1 \\ \bar{\boldsymbol{\omega}}_2 \end{pmatrix}\begin{array}{c} k \\ d-k \end{array},\quad \Delta = \begin{pmatrix} \Delta_1 \\ \Delta_2 \end{pmatrix}\begin{array}{c} k \\ d-k \end{array},$

$$\Omega_1^* = \begin{pmatrix} \Gamma & \Delta_1^T \\ \Delta_1 & \bar{\Omega}_{11} \end{pmatrix}, \quad and \quad \bar{\Omega} = \begin{array}{c} \begin{pmatrix} \bar{\Omega}_{11} & \bar{\Omega}_{12} \\ \bar{\Omega}_{21} & \bar{\Omega}_{22} \end{pmatrix} \end{array} \begin{array}{c} k \\ d-k \end{array}.$$

Proof. By Theorem 2.4.4,

$$\boldsymbol{y}_1 = A\boldsymbol{y} \sim SUN_{n,m}(\boldsymbol{\mu}_A, \boldsymbol{\gamma}, \bar{\omega}_A, \Omega_A^*).$$

Now,

$$\boldsymbol{\mu}_A = A\boldsymbol{\mu} = (I_k \;\; 0) \begin{pmatrix} \boldsymbol{\mu}_1 \\ \boldsymbol{\mu}_2 \end{pmatrix} = \boldsymbol{\mu}_1, \quad \bar{\boldsymbol{\omega}}_A = \omega_A \mathbf{1}_d = A\omega A^T \mathbf{1}_d = \omega_1 \mathbf{1}_d,$$

$$\Delta_A = (A\omega A^T)^{-1} A\omega\Delta = \omega_1^{-1}\omega_1\Delta_1 = \Delta_1,$$

$$\bar{\Omega}_A = (A\omega A^T)^{-1} A\Omega A^T (A\omega A^T)^{-1} = \omega_1^{-1}\Omega_{11}\omega_1^{-1} = \bar{\Omega}_{11}.$$

Therefore,

$\boldsymbol{y}_1 = A\boldsymbol{y}$ is $SUN_{k,m}(\boldsymbol{\mu}_1, \boldsymbol{\gamma}, \bar{\omega}_1, \Omega_1^*)$, where

$$\Omega_1^* = \begin{pmatrix} \Gamma & \Delta_1^T \\ \Delta_1 & \bar{\Omega}_{11} \end{pmatrix}.$$

□

We provide an alternative proof here as follows:

For $\boldsymbol{t} \in \Re^k$ the m.g.f of $A\boldsymbol{y}$ is given by,

$$\begin{aligned} M_{A\boldsymbol{y}}(\boldsymbol{t}) &= M_{\boldsymbol{y}}(A^T\boldsymbol{t}) \\ &= exp\,(\boldsymbol{\mu}^T A^T t + \frac{1}{2}\boldsymbol{t}^T A\Omega A_T \boldsymbol{t}) \frac{\Phi_m(\boldsymbol{\gamma} + \Delta^T\omega A^T\boldsymbol{t}; \Gamma)}{\Phi_m(\boldsymbol{\gamma}; \Gamma)}. \end{aligned}$$

Since, $A = (I_k \;\; 0)$ and considering the partion above,

$$\boldsymbol{\mu}^T A^T = \boldsymbol{\mu}_1^T, \quad A\Omega A^T = \Omega_{11}, \quad \Delta^T \omega A^T = \Delta_1^T \omega_1.$$

Then

$$M_{A\boldsymbol{y}}(\boldsymbol{t}) = M_{\boldsymbol{y}_1}(\boldsymbol{t}) = exp\,(\boldsymbol{\mu}_1^T \boldsymbol{t} + \frac{1}{2}\boldsymbol{t}^T \Omega_{11} \boldsymbol{t}) \frac{\Phi_m(\boldsymbol{\gamma} + \Delta_1^T \omega_1 \boldsymbol{t}; \Gamma)}{\Phi_m(\boldsymbol{\gamma}; \Gamma)},$$

which is the m.g.f of $SUN_{k,m}(\boldsymbol{\mu}_1, \boldsymbol{\gamma}, \bar{\omega}_1, \Omega_1^*)$.

Corollary 2.4.6. *With the reparameterization scheme* $\boldsymbol{\gamma} = -\boldsymbol{\nu}$, $\bar{\Omega} = \Sigma$, $\Delta = \Sigma D^T$, $\Gamma = \Psi + D\Sigma D^T$,

$$\boldsymbol{y_1} \sim CSN_{k,m}(\boldsymbol{\mu}_1, \Sigma_{11}, D^*, \boldsymbol{\nu}, \Psi^*),$$

where

$$D^* = D_1 + D_2 \Sigma_{21} \Sigma_{11}^{-1}, \quad \Psi^* = \Psi + D_2 \Sigma_{22.1} D_2^T \textit{ and } \Sigma_{22.1} = \Sigma_{22} - \Sigma_{21} \Sigma_{11}^{-1} \Sigma_{12}.$$

The parameters $\boldsymbol{\mu}_1, \quad \Sigma_{11}, \quad \Sigma_{22}, \quad \Sigma_{12}, \quad \Sigma_{21}$ *come from the partition*

$$\boldsymbol{\mu} = \begin{pmatrix} \boldsymbol{\mu}_1 \\ \boldsymbol{\mu}_2 \end{pmatrix} \begin{matrix} k \\ d-k \end{matrix}, \quad D = \begin{pmatrix} D_1 \\ D_2 \end{pmatrix} \begin{matrix} k \\ d-k \end{matrix}, \quad \Sigma = \begin{pmatrix} \Sigma_{11} & \Sigma_{12} \\ \Sigma_{21} & \Sigma_{22} \end{pmatrix} \begin{matrix} k \\ d-k \end{matrix}.$$

Theorem 2.4.8. *Let* $\boldsymbol{y}$ *be a random vector distributed as* $y \sim SUN_{d,m}(\boldsymbol{\mu}, \boldsymbol{\gamma}, \bar{\boldsymbol{\omega}}, \Omega^*)$. *Consider two subvectors* $\boldsymbol{y}_1$ *and* $\boldsymbol{y}_2$, *where* $\boldsymbol{y}^T = (\boldsymbol{y}_1^T \;\; \boldsymbol{y}_2^T)$, $\boldsymbol{y}_1$ *is* k *dimensional. Suppose* $\boldsymbol{\mu}$, $\bar{\omega}$, Δ *and* $\bar{\Omega}$ *are partitioned as in above Theorem. Then the conditional distribution of* $\boldsymbol{y}_1$ *given* $\boldsymbol{y}_2 = \boldsymbol{y}_{10}$ *is* $SUN_{k,m}(\boldsymbol{\mu}_{1.2}, \boldsymbol{\gamma}_{1.2}, \bar{\omega}_1, \Omega_{11.2}^*)$, *where*

$$\boldsymbol{\mu}_{1.2} = \boldsymbol{\mu}_1 + \Omega_{12} \Omega_{22}^{-1} (y_{10} - \boldsymbol{\mu}_2), \boldsymbol{\gamma}_{1.2} = \boldsymbol{\gamma} + \Delta_2^T \bar{\Omega}_{22}^{-1} \omega_2^{-1} (y_{10} - \boldsymbol{\mu}_2), \bar{\Omega}_{11.2} = \bar{\Omega}_{11} - \bar{\Omega}_{12} \bar{\Omega}_{22}^{-1} \bar{\Omega}_{21},$$

$$\Omega_{11.2} = \omega_1 \bar{\Omega}_{11.2} \omega_1 = \Omega_{11} - \Omega_{12} \Omega_{22}^{-1} \Omega_{21}, \textit{with } \Omega_{ij} = \omega_i \bar{\Omega}_{ij} \omega_j, \; i,j = 1,2, \; \Delta_{1.2} = \Delta_1 - \bar{\Omega}_{12} \bar{\Omega}_{22}^{-1} \Delta_2,$$

$$\Gamma_{1.2} = \Gamma - \Delta_2^T \bar{\Omega}_{22}^{-1} \Delta_2 \quad \textit{and} \quad \Omega_{1.2}^* = \begin{pmatrix} \Gamma_{1.2} & \Delta_{1.2}^T \\ \Delta_{1.2} & \bar{\Omega}_{11.2} \end{pmatrix}.$$

Proof. The proof follows from direct calculations as shown in Arellano-Valle and Azzalini (2006) page 571. □

2.4.6 Joint Distribution of Independent SUN Random Vectors

In this section we will show that if we have a collection of n independent SUN random variables then the joint distribution of the n random variables is again SUN distributed random variable. As mentioned by Gupta et al (2004), this property does not hold for the multivariate skew normal distribution defined by Azzallini and Dalla Valle (1996).

Theorem 2.4.9. *Suppose $\boldsymbol{y}_1, \ldots, \boldsymbol{y}_n$ are independent random vectors with $\boldsymbol{y}_i \sim SUN_{d_i,m_i}(\boldsymbol{\mu}_i, \boldsymbol{\gamma}_i, \bar{\boldsymbol{\omega}}_i, \Omega_i^*)$. Then the joint distribution of $\boldsymbol{y}_1, \ldots, \boldsymbol{y}_n$ is given by*

$$\boldsymbol{y} = (\boldsymbol{y}_1^T, \ldots \boldsymbol{y}_n^T)^T \sim SUN_{d^\dagger, m^\dagger}(\boldsymbol{\mu}^\dagger, \boldsymbol{\gamma}^\dagger, \bar{\boldsymbol{\omega}}^\dagger, \Omega^{*\dagger}),$$

where

$$d^\dagger = \sum_{i=1}^n d_i, \quad m^\dagger = \sum_{i=1}^n m_i, \quad \boldsymbol{\mu}^\dagger = (\boldsymbol{\mu}_1^T, \ldots \boldsymbol{\mu}_n^T)^T, \quad \boldsymbol{\gamma}^\dagger = (\boldsymbol{\gamma}_1^T, \ldots \boldsymbol{\gamma}_n^T)^T, \quad \bar{\boldsymbol{\omega}}^\dagger = (\bar{\boldsymbol{\omega}}_1^T, \ldots \bar{\boldsymbol{\omega}}_n^T)^T,$$

and

$$\omega^\dagger = \bigoplus_{i=1}^n \omega_i \quad \Omega^\dagger = \bigoplus_{i=1}^n \Omega_i, \quad \bar{\Omega}^\dagger = \bigoplus_{i=1}^n \bar{\Omega}_i, \quad \Gamma^\dagger = \bigoplus_{i=1}^n \Gamma_i, \quad \Delta^\dagger = \bigoplus_{i=1}^n \Delta_i, \quad \Omega^{*\dagger} = \begin{pmatrix} \Gamma^\dagger & {\Delta^\dagger}^T \\ \Delta^\dagger & \bar{\Omega}^\dagger \end{pmatrix}.$$

Proof. For $\boldsymbol{y} = (\boldsymbol{y}_1^T, \ldots \boldsymbol{y}_n^T)^T$, $\boldsymbol{y}_i \in \Re^{d_i}$, the density function of $\boldsymbol{y}$ is given by,

$$g(\boldsymbol{y}) = \prod_{i=1}^n f_{d_i,m_i}(\boldsymbol{y}_i; \boldsymbol{\mu}_i, \boldsymbol{\gamma}_i, \bar{\boldsymbol{\omega}}_i, \Omega_i^*)$$

$$
\begin{aligned}
&= \prod_{i=1}^{n} \phi_{d_i}(\boldsymbol{y}_i - \boldsymbol{\mu}_i; \Omega_i) \frac{\Phi_{m_i}(\boldsymbol{\gamma}_i + \Delta_i^T \bar{\Omega}_i^{-1} \omega_i^{-1}(\boldsymbol{y}_i - \boldsymbol{\mu}_i); \Gamma_i - \Delta_i^T \bar{\Omega}_i^{-1} \Delta_i)}{\Phi_{m_i}(\boldsymbol{\gamma}_i; \Gamma_i)} \\
&= \prod_{i=1}^{n} \phi_{d_i}(\boldsymbol{y}_i - \boldsymbol{\mu}_i; \Omega_i) \frac{\prod_{i=1}^{n} \Phi_{m_i}(\boldsymbol{\gamma}_i + \Delta_i^T \bar{\Omega}_i^{-1} \omega_i^{-1}(\boldsymbol{y}_i - \boldsymbol{\mu}_i); \Gamma_i - \Delta_i^T \bar{\Omega}_i^{-1} \Delta_i)}{\prod_{i=1}^{n} \Phi_{m_i}(\boldsymbol{\gamma}_i; \Gamma_i)} \\
&= \phi_{d^\dagger}(\boldsymbol{y} - \boldsymbol{\mu}^\dagger; \Omega^\dagger) \frac{\Phi_{m^\dagger}(\boldsymbol{\gamma}^\dagger + \Delta^{\dagger^T} \bar{\Omega}^{\dagger^{-1}} \omega^{\dagger^{-1}}(\boldsymbol{y} - \boldsymbol{\mu}^\dagger); \Gamma^\dagger - \Delta^{\dagger^T} \bar{\Omega}^{\dagger^{-1}} \Delta^\dagger)}{\Phi_{m^\dagger}(\boldsymbol{\gamma}^\dagger; \Gamma^\dagger)},
\end{aligned}
$$

where

$$
\begin{aligned}
\bigoplus_{i=1}^{n} (\Gamma_i - \Delta_i^T \bar{\Omega}_i^{-1} \Delta_i) &= \bigoplus_{i=1}^{n} \Gamma_i - \bigoplus_{i=1}^{n} (\Delta_i^T \bar{\Omega}_i^{-1} \Delta_i) \\
&= \bigoplus_{i=1}^{n} \Gamma_i - (\bigoplus_{i=1}^{n} \Delta_i^T)(\bigoplus_{i=1}^{n} \bar{\Omega}_i^{-1})(\bigoplus_{i=1}^{n} \Delta_i) \\
&= \Gamma^\dagger - \Delta^{\dagger^T} \bar{\Omega}^{\dagger^{-1}} \Delta^\dagger.
\end{aligned}
$$

□

Corollary 2.4.7. *If $\boldsymbol{y}_1, \ldots, \boldsymbol{y}_n$ are independent and identically distributed (iid) random vectors from the $SUN_{d,m}(\boldsymbol{\mu}, \boldsymbol{\gamma}, \bar{\boldsymbol{\omega}}, \Omega^*)$ distribution, then the joint distribution of $\boldsymbol{y}_1, \ldots, \boldsymbol{y}_n$ is*

$$
Y = (\boldsymbol{y}_1^T, \ldots \boldsymbol{y}_n^T)^T \sim SUN_{d^\dagger, m^\dagger}(\boldsymbol{\mu}^\dagger, \boldsymbol{\gamma}^\dagger, \bar{\boldsymbol{\omega}}^\dagger, \Omega^{*\dagger}),
$$

where

$$
d^\dagger = nd, \quad m^\dagger = nm, \quad \boldsymbol{\mu}^\dagger = \mathbf{1}_n \otimes \boldsymbol{\mu}, \quad \boldsymbol{\gamma}^\dagger = \mathbf{1}_n \otimes \boldsymbol{\gamma}, \quad \bar{\boldsymbol{\omega}}^\dagger = \mathbf{1}_n \otimes \bar{\boldsymbol{\omega}},
$$

and

$$
\Omega^\dagger = I_n \otimes \Omega, \quad \bar{\Omega}^\dagger = I_n \otimes \bar{\Omega}, \quad \Delta^\dagger = I_n \otimes \Delta, \quad \Gamma^\dagger = I_n \otimes \Gamma, \quad \Omega^{*\dagger} = \begin{pmatrix} \Gamma^\dagger & \Delta^{\dagger^T} \\ \Delta^\dagger & \bar{\Omega}^\dagger \end{pmatrix}.
$$

In the above two theorems, we showed that the SUN family is closed under the linear transformation, and the joint distribution of random sample belongs to the same family. These two properties will help us to obtain an important property of SUN density: the sum of independent SUN distributions again follows the SUN distribution.

2.4.7 Sum of Independent SUN Random Vectors

In this section we present the main result regarding the additive properties of SUN random vector. More precisely, we show that the sum of independent SUN random vectors is again a SUN random vector.

Theorem 2.4.10. *If* $\boldsymbol{y}_1, \ldots, \boldsymbol{y}_n$ *are independent random vectors with* $\boldsymbol{y}_i \sim SUN_{d,m_i}(\boldsymbol{\mu}_i, \boldsymbol{\gamma}_i, \bar{\boldsymbol{\omega}}_i, \Omega_i^*),\ i = 1, \ldots, n,$ *then*

$$\sum_{i=1}^{n} \boldsymbol{y}_i \sim SUN_{d,m^*}(\boldsymbol{\mu}^*, \boldsymbol{\gamma}^*, \bar{\boldsymbol{\omega}}^*, \Omega^{**}),$$

where

$$m^* = \sum_{i=1}^{n} m_i, \quad \boldsymbol{\mu}^* = \sum_{i=1}^{n} \boldsymbol{\mu}_i, \quad \boldsymbol{\gamma}^* = (\boldsymbol{\gamma}_1^T, \ldots \boldsymbol{\gamma}_n^T)^T, \quad \omega^* = \sum_{i=1}^{n} \omega_i, \quad \bar{\boldsymbol{\omega}}^* = \omega^* \mathbf{1}_{nd}, \quad \Omega^* = \sum_{i=1}^{n} \Omega_i$$

and

$$\Gamma^* = \bigoplus_{i=1}^{n} \Gamma_i, \quad \Delta^* = (\mathbf{1}_n \otimes \omega_d^{-1})(\bigoplus_{i=1}^{n} \omega_i \Delta_i), \quad \Omega^{**} = \begin{pmatrix} \Gamma^* & \Delta^{*T} \\ \Delta^* & \bar{\Omega}^* \end{pmatrix}.$$

Proof. Let $\boldsymbol{y} = (\boldsymbol{y}_1^T, \ldots \boldsymbol{y}_n^T)^T$ and $A = \mathbf{1}_n^T \otimes I_d$. Note that $\sum_{i=1}^n \boldsymbol{y}_i = A\boldsymbol{y}$, where $\boldsymbol{y}$ is a $nd \times 1$ vector and A is a $d \times nd$ matrix of rank d. Then by Theorems 2.4.4 and 2.4.9,

$$A\boldsymbol{y} \sim SUN_{d,nm}(\boldsymbol{\mu}_A^\dagger, \boldsymbol{\gamma}_A^\dagger, \bar{\boldsymbol{\omega}_A}^\dagger, \Omega_A^{*\dagger}),$$

where

$$\boldsymbol{\mu}_A^\dagger = A\boldsymbol{\mu}^\dagger, \quad \boldsymbol{\gamma}_A^\dagger = \boldsymbol{\gamma}^\dagger, \quad \omega_A^\dagger = A\omega^\dagger A^T, \quad \bar{\boldsymbol{\omega}_A}^\dagger = \omega_A^\dagger \mathbf{1}_d, \quad \Omega_A^\dagger = A\Omega^\dagger A^T,$$

and

$$\bar{\Omega}_A^{\dagger} = (A\omega^{\dagger}A^T)^{-1}\Omega_A^{\dagger}(A\omega^{\dagger}A^T)^{-1}, \quad \Gamma_A^{\dagger} = \Gamma^{\dagger}, \quad \Delta_A^{\dagger} = (A\omega^{\dagger}A^T)^{-1}A\omega^{\dagger}\Delta^{\dagger}, \quad \Omega_A^{*\dagger} = \begin{pmatrix} \Gamma_A^{\dagger} & \Delta_A^{\dagger^T} \\ \Delta_A^{\dagger} & \bar{\Omega}_A^{\dagger} \end{pmatrix},$$

where $\boldsymbol{\mu}^{\dagger},\ \boldsymbol{\gamma}^{\dagger},\ \bar{\boldsymbol{\omega}}^{\dagger},\ \bar{\Omega}^{\dagger},\ \Delta^{\dagger}$ and $\Gamma^{\dagger}$ are given in theorem 2.4.9. Now it is easily observed that,

$$A\boldsymbol{\mu}^{\dagger} = \sum_{i=1}^{n} \boldsymbol{\mu_i}, \quad A\omega^{\dagger}A^T = \sum_{i=1}^{n} \omega_i, \quad A\Omega^{\dagger}A^T = \sum_{i=1}^{n} \Omega_i$$

and,

$$\begin{aligned}
\Delta_A^{\dagger} &= (\sum_{i=1}^{n} \omega_i)^{-1} A(\bigoplus_{i=1}^{n} \omega_i)(\bigoplus_{i=1}^{n} \Delta_i) \\
&= (\sum_{i=1}^{n} \omega_i)^{-1} A(\bigoplus_{i=1}^{n} \omega_i \Delta_i) \\
&= (\mathbf{1}_n \otimes \omega_d^{-1})(\bigoplus_{i=1}^{n} \omega_i \Delta_i).
\end{aligned}$$

$\square$

CHAPTER 3

QUADRATIC FORMS IN UNIFIED SKEW NORMAL RANDOM VECTOR

3.1 Introduction

There is a rich literature on the distribution of quadratic form on the multivariate normal random vector. Earlier works were due to Cochran (1934), Craig (1943) and Rao (1973a). Recent references includes the book “Quadratic Forms in Random Variable” by Mathai and Provost (1992). The quadratic forms of multivariate skew normal distribution was studied by Genton et al (2001), Loperfido (2001), Gupta and Huang (2002), Huang and Chen (2006), and most recently by Wang, Li and Gupta (2009). In this chapter we study the quadratic forms under unified skew normal settings. We will explore their distributions, moments and conditions under which quadratic forms are independent.

3.2 Moment Generating Function of a Quadratic Form of SUN Random Vector

Theorem 3.2.1. *Let* $\boldsymbol{y} \sim SUN_{d,m}(\boldsymbol{\mu}, \boldsymbol{\gamma}, \bar{\boldsymbol{\omega}}, \Omega^*)$. *Let* A *be a* $d\times d$ *symmetric matrix. Consider the quadratic form,* $Q(\boldsymbol{y}) = \boldsymbol{y}^T A\boldsymbol{y}$. *Then the m.g.f of* Q *is*

$$M_Q(t) = \frac{|I_d - 2tA\Omega|^{-\frac{1}{2}} \quad \Phi_m(\boldsymbol{\gamma}; \Gamma + 2t\Delta^T\omega(I_d - 2tA\Omega)^{-1}A\omega\Delta)}{exp\ (t\boldsymbol{\mu}^T A\Omega(I_d - 2tA\Omega)^{-1}\Omega^{-1}\boldsymbol{\mu}) \quad \Phi_m(\boldsymbol{\gamma}; \Gamma)}. \tag{3.1}$$

Proof. By the definition of m.g.f for $t \in \Re$ we have,

$$\begin{aligned} &M_Q(t) = E(e^{t\boldsymbol{y}^T A\boldsymbol{y}}) \\ = \ & C\int_{\Re^d} e^{t\boldsymbol{y}^T A\boldsymbol{y}}\phi_d(\boldsymbol{y}-\boldsymbol{\mu};\Omega)\Phi_m(\boldsymbol{\gamma} + \Delta^T\bar{\Omega}^{-1}\omega^{-1}(\boldsymbol{y}-\boldsymbol{\mu}); \Gamma - \Delta^T\bar{\Omega}^{-1}\Delta)\ \mathrm{d}\boldsymbol{y} \\ = \ & K\int_{\Re^d} e^{t\boldsymbol{y}^T A\boldsymbol{y}-\frac{1}{2}(\boldsymbol{y}-\boldsymbol{\mu})^T\Omega^{-1}(\boldsymbol{y}-\boldsymbol{\mu})}\Phi_m(\boldsymbol{\gamma} + \Delta^T\bar{\Omega}^{-1}\omega^{-1}(\boldsymbol{y}-\boldsymbol{\mu}); \Gamma - \Delta^T\bar{\Omega}^{-1}\Delta)\ \mathrm{d}\boldsymbol{y}. \end{aligned}$$

By expanding $(\boldsymbol{y}-\boldsymbol{\mu})^T\Omega^{-1}(\boldsymbol{y}-\boldsymbol{\mu})$ and rearranging the terms we have,

$$\begin{aligned} & t\boldsymbol{y}^T A\boldsymbol{y} - \frac{1}{2}(\boldsymbol{y}-\boldsymbol{\mu})^T\Omega^{-1}(\boldsymbol{y}-\boldsymbol{\mu}) \\ = \ & -\frac{1}{2}\boldsymbol{\mu}^T\Omega^{-1}\boldsymbol{\mu} + \frac{1}{2} + \boldsymbol{\mu}^T A\Omega(I_d - 2tA\Omega)^{-1}\Omega^{-1}\boldsymbol{\mu}) - \frac{1}{2}(\boldsymbol{y}-\boldsymbol{a})^T\Omega^{-1} - 2tA)^{-1}(\boldsymbol{y}-\boldsymbol{a}), \end{aligned}$$

where

$$\boldsymbol{a} = (\Omega^{-1} - 2tA)^{-1}\Omega^{-1}\boldsymbol{\mu}.$$

Therefore

$$\begin{aligned} & M_Q(t) \\ = \ & \frac{|I_d - 2tA\Omega|^{-\frac{1}{2}}}{exp\ (t\boldsymbol{\mu}^T A\Omega(I_d - 2tA\Omega)^{-1}\Omega^{-1}\boldsymbol{\mu}) \quad \Phi_m(\boldsymbol{\gamma}; \Gamma)}\int_{\Re^d}\phi_d\Big(\boldsymbol{y} - (\Omega^{-1} - 2tA)^{-1}\Omega^{-1}\boldsymbol{\mu}; \\ & \qquad (\Omega^{-1} - 2tA)^{-1}\Big) \times \Phi_m(\boldsymbol{\gamma} + \Delta^T\bar{\Omega}^{-1}\omega^{-1}(\boldsymbol{y}-\boldsymbol{\mu}); \Gamma - \Delta^T\bar{\Omega}^{-1}\Delta)\ \mathrm{d}\boldsymbol{y} \end{aligned}$$

$$= \quad \frac{|I_d - 2tA\Omega|^{-\frac{1}{2}}}{exp\ (t\boldsymbol{\mu}^T A\Omega(I_d - 2tA\Omega)^{-1}\Omega^{-1}\boldsymbol{\mu})\ \ \Phi_m(\boldsymbol{\gamma};\Gamma)} E_U[\Phi_m(\boldsymbol{\gamma} + \Delta^T\bar{\Omega}^{-1}\omega^{-1}U; \Gamma - \Delta^T\bar{\Omega}^{-1}\Delta)]$$

where

$$U \sim N_d(0, (\Omega^{-1} - 2tA)^{-1}).$$

With the lemma (2.4.1) we get,

$$M_Q(t) = \frac{|I_d - 2tA\Omega|^{-\frac{1}{2}}\Phi_m(\boldsymbol{\gamma}; \Gamma - \Delta^T\bar{\Omega}^{-1}\Delta + \Delta^T\bar{\Omega}^{-1}\omega^{-1}(\Omega^{-1} - 2tA)^{-1}{\omega^{-1}}\bar{\Omega}^{-1}\Delta)}{exp\ (t\boldsymbol{\mu}^T A\Omega(I_d - 2tA\Omega)^{-1}\Omega^{-1}\boldsymbol{\mu})\ \ \Phi_m(\boldsymbol{\gamma};\Gamma)}. \quad (3.2)$$

Noting that

$$(\Omega^{-1} - 2tA)^{-1} = \Omega\sum_{j=0}^{\infty}(2t)^j(A\Omega)^j$$

for $||2tA\Omega|| < 1$, where $||.||$ is the matrix norm. The last term in $M_Q(t)$ becomes,

$$\Phi_m(\boldsymbol{\gamma}; \Gamma - \Delta^T\bar{\Omega}^{-1}\Delta + \Delta^T\bar{\Omega}^{-1}\omega^{-1}(\Omega^{-1} - 2tA)^{-1}{\omega^{-1}}\bar{\Omega}^{-1}\Delta)$$

$$= \Phi_m(\boldsymbol{\gamma}; \Gamma - \Delta^T\bar{\Omega}^{-1}\Delta + \Delta^T\bar{\Omega}^{-1}\omega^{-1}\Omega\omega^{-1}\bar{\Omega}^{-1}\Delta + 2t\Delta^T\bar{\Omega}^{-1}\omega^{-1}\Omega(I_d - 2tA\Omega)^{-1}A\Omega{\omega^{-1}}\bar{\Omega}^{-1}\Delta)$$

$$= \Phi_m(\boldsymbol{\gamma}; \Gamma + 2t\Delta^T\omega(I_d - 2tA\Omega)^{-1}A\omega\Delta).$$

The last identity was obtained from the relation $\Omega = \omega\bar{\Omega}\omega$ or equivalently $\bar{\Omega} = \omega^{-1}\Omega\omega^{-1}$ Finally, from (3.2) we get,

$$M_Q(t) = \frac{|I_d - 2tA\Omega|^{-\frac{1}{2}}\ \ \Phi_m(\boldsymbol{\gamma}; \Gamma + 2t\Delta^T\omega(I_d - 2tA\Omega)^{-1}A\omega\Delta)}{exp\ (t\boldsymbol{\mu}^T A\Omega(I_d - 2tA\Omega)^{-1}\Omega^{-1}\boldsymbol{\mu})\ \ \Phi_m(\boldsymbol{\gamma};\Gamma)}.$$

$\square$

Corollary 3.2.1. *Let $\boldsymbol{y} \sim SUN_{d,m}(\boldsymbol{\mu}, \boldsymbol{\gamma}, \bar{\boldsymbol{\omega}}, \Omega^*)$. Let A be a $d \times d$ symmetric matrix and $Q(\boldsymbol{y}) = \boldsymbol{y}^T A\boldsymbol{y}$. Then the following results can be deduced from equation (3.1)*

(i) *Suppose* $\boldsymbol{\mu} = \mathbf{0}$*, then the m.g.f of* Q *becomes*

$$M_Q(t) = \frac{|I_d - 2tA\Omega|^{-\frac{1}{2}} \quad \Phi_m(\boldsymbol{\gamma}; \Gamma + 2t\Delta^T\omega(I_d - 2tA\Omega)^{-1}A\omega\Delta)}{\Phi_m(\boldsymbol{\gamma}; \Gamma)}, \;\; \Omega^{-1} - 2tA > 0, \;\; t \in \Re.$$

(ii) *Suppose* $\boldsymbol{\mu} = \mathbf{0}$*, and* $A\omega\Delta = 0_{d\times m}$*, then the m.g.f of* Q *becomes*

$$M_Q(t) = |I_d - 2tA\Omega|^{-\frac{1}{2}}, \quad \Omega^{-1} - 2tA > 0, \;\; t \in \Re,$$

which is the m.g.f of $\boldsymbol{y}^T A\boldsymbol{y}$ *where* $\boldsymbol{y} \sim N_d(\mathbf{0}, \Omega)$ *and* A *is a* $d \times d$ *symmetric matrix. Consequently, properties of* Q *can be showed by using known results of the multivariate normal distribution.*

(iii) *Suppose* $\boldsymbol{\mu} = \mathbf{0}$*,* $A\omega\Delta = 0_{d\times m}$ *and* $A\Omega$*=diag(*$\tau_1, \ldots.., \tau_d$*), then the m.g.f of* Q *becomes*

$$M_Q(t) = \prod_{j=1}^{d}(1 - 2t\tau_j)^{-1/2}, \;\; t \in \Re.$$

Hence $\boldsymbol{y}^T A\boldsymbol{y} \sim \sum_{j=1}^{d} \tau_j X_j$*,where* $X_j \sim \chi^2_1, \;\; j = 1, \ldots, d$ *are independently and identically distributed.*

(iv) *Suppose* $\boldsymbol{\mu} = \mathbf{0}$*, and* $A\omega\Delta = 0_{d\times m}$*, and* A *=*Ω^{-1} *such that* $A\Omega = \Omega^{-1}\Omega = I$*, then the m.g.f of* $Q = \boldsymbol{y}^T\Omega^{-1}\boldsymbol{y}$ *becomes*

$$M_Q(t) = (1 - 2t)^{-d/2}, \;\; t \in \Re.$$

Hence,

$$\boldsymbol{y}^T\Omega^{-1}\boldsymbol{y} \sim \chi^2_d.$$

(v) *Suppose* $\boldsymbol{\mu} = \mathbf{0}$ *and* $\bar{\boldsymbol{\omega}} = \mathbf{1}_d$*, then the m.g.f of* Q *becomes*

$$M_Q(t) = \frac{|I_d - 2tA\bar{\Omega}|^{-\frac{1}{2}} \quad \Phi_m(\boldsymbol{\gamma}; \Gamma + 2t\Delta^T(I_d - 2tA\bar{\Omega})^{-1}A\Delta)}{\Phi_m(\boldsymbol{\gamma}; \Gamma)},$$

which is the m.g.f obtained by Arellano-Valle and Azzalini (2006). Taking $A\Delta = 0$ *and substituting* $A = \bar{\Omega}^{-1}$ *yields*

$$\boldsymbol{y}^T \bar{\Omega}^{-1} \boldsymbol{y} \sim \chi_d^2.$$

3.3 Independence of a Linear Form and a Quadratic Form

In this section we study the conditions under which a linear function of SUN random vector is independent of its quadratic form. We also give conditions under which the two quadratic forms are independent.

Theorem 3.3.1. *Suppose* $\boldsymbol{y} \sim SUN_{d,m}(\mathbf{0}, \boldsymbol{\gamma}, \bar{\boldsymbol{\omega}}, \Omega^*)$. *Then for* $\boldsymbol{h} \in \Re^d$, *the linear form* $\boldsymbol{h}^T\boldsymbol{y}$ *and the quadratic form* $\boldsymbol{y}^T A \boldsymbol{y}$ *are independent if and only if* $A\Omega\boldsymbol{h} = \mathbf{0}$ *and* $A\omega\Delta = 0$.

Proof. We first derive the joint m.g.f of $\boldsymbol{h}^T\boldsymbol{y}$ and $\boldsymbol{y}^T A\boldsymbol{y}$. For $t, s \in \Re$, the joint m.g.f of $\boldsymbol{h}^T\boldsymbol{y}$ and $\boldsymbol{y}^T A\boldsymbol{y}$ is

$$M(t,s) = \frac{1}{\Phi_m(\boldsymbol{\gamma};\Gamma)} \int_{\Re^d} exp\{t\ \boldsymbol{h}^T\boldsymbol{y} + s\ \boldsymbol{y}^T A\boldsymbol{y}\}\phi_d(\boldsymbol{y};\Omega)\Phi_m(\boldsymbol{\gamma} + \Delta^T\bar{\boldsymbol{\Omega}}^{-1}\omega^{-1}\boldsymbol{y}; \Gamma - \Delta^T\bar{\boldsymbol{\Omega}}^{-1}\Delta)\ \mathrm{d}\boldsymbol{y}$$

$$= K\ exp\{-\frac{1}{2}(\boldsymbol{y}^T\Omega^{-1}\boldsymbol{y} - 2t\ \boldsymbol{h}^T\boldsymbol{y} - 2s\ \boldsymbol{y}^T A\boldsymbol{y})\}\Phi_m(\boldsymbol{\gamma} + \Delta^T\bar{\Omega}^{-1}\omega^{-1}\boldsymbol{y}; \Gamma - \Delta^T\bar{\Omega}^{-1}\Delta)\ \mathrm{d}\boldsymbol{y}.$$

Now,

$$\begin{aligned}
& exp\{-\frac{1}{2}(\boldsymbol{y}^T\Omega^{-1}\boldsymbol{y} - 2t\ \boldsymbol{h}^T\boldsymbol{y} - 2s\ \boldsymbol{y}^T A\boldsymbol{y})\} \\
= \ & exp\{-\frac{1}{2}(\boldsymbol{y}^T(\Omega^{-1} - 2sA)^{-1}\boldsymbol{y} - 2t\ \boldsymbol{h}^T\boldsymbol{y})\} \\
= \ & exp\{-\frac{1}{2}(\boldsymbol{y} - t(\Omega^{-1} - 2sA)^{-1}\boldsymbol{h})^T(\Omega^{-1} - 2sA) \times \\
& (\boldsymbol{y} - t(\Omega^{-1} - 2sA)^{-1}\boldsymbol{h}) - t^2\ \boldsymbol{h}^T(\Omega^{-1} - 2sA)^{-1})\boldsymbol{h})\} \\
= \ & exp\{\frac{1}{2}t^2\ \boldsymbol{h}^T(\Omega^{-1} - 2sA)^{-1})\boldsymbol{h}\} \quad \times \\
& exp\{-\frac{1}{2}(\boldsymbol{y} - t(\Omega^{-1} - 2sA)^{-1}\boldsymbol{h})^T(\Omega^{-1} - 2sA)(\boldsymbol{y} - t(\Omega^{-1} - 2sA)^{-1}\boldsymbol{h})\}.
\end{aligned}$$

Therefore,

$$M(t,s) = \frac{exp\{\frac{1}{2}t^2\ \boldsymbol{h}^T(\Omega^{-1}-2sA)^{-1})\boldsymbol{h}\}|I_d-2sA\Omega|^{-\frac{1}{2}}}{\Phi_m(\boldsymbol{\gamma};\Gamma)} \times$$
$$\int_{\Re^d}\phi_d(\boldsymbol{y}-t(\Omega^{-1}-2sA)^{-1}\boldsymbol{h};(\Omega^{-1}-2sA)^{-1})\Phi_m(\boldsymbol{\gamma}+\Delta^T\bar{\Omega}^{-1}\omega^{-1}\boldsymbol{y};\Gamma-\Delta^T\bar{\Omega}^{-1}\Delta)\ \mathrm{d}\boldsymbol{y}$$
$$= \frac{exp\{\frac{1}{2}t^2\ \boldsymbol{h}^T(\Omega^{-1}-2sA)^{-1})\boldsymbol{h}\}|I_d-2sA\Omega|^{-\frac{1}{2}}}{\Phi_m(\boldsymbol{\gamma};\Gamma)} E_U[\Phi_m(\boldsymbol{\gamma}+\Delta^T\bar{\Omega}^{-1}\omega^{-1}U;\Gamma-\Delta^T\bar{\Omega}^{-1}\Delta)],$$

where $U \sim N_d(t(\Omega^{-1}-2sA)^{-1}\boldsymbol{h},(\Omega^{-1}-2sA)^{-1})$.

$$= \frac{exp\{\frac{1}{2}t^2\ \boldsymbol{h}^T(\Omega^{-1}-2sA)^{-1})\boldsymbol{h}\}|I_d-2sA\Omega|^{-\frac{1}{2}}}{\Phi_m(\boldsymbol{\gamma};\Gamma)} \times\ \Phi_m(\boldsymbol{\gamma}+t\Delta^T\bar{\Omega}^{-1}\omega^{-1}(\Omega^{-1}-2sA)^{-1}\boldsymbol{h};$$
$$\Gamma-\Delta^T\bar{\Omega}^{-1}\Delta+\Delta^T\bar{\Omega}^{-1}\omega^{-1}(\Omega^{-1}-2sA)^{-1}\omega^{-1}\bar{\Omega}^{-1}\Delta)$$

$$= \frac{exp\{\frac{1}{2}t^2\ \boldsymbol{h}^T(\Omega^{-1}-2sA)^{-1})\boldsymbol{h}\}}{|I_d-2sA\Omega|^{\frac{1}{2}}\ \Phi_m(\boldsymbol{\gamma};\Gamma)} \times\ \Phi_m(\boldsymbol{\gamma}+t\Delta^T\bar{\Omega}^{-1}\omega^{-1}(\Omega^{-1}-2sA)^{-1}\boldsymbol{h};$$
$$\Gamma+2s\Delta^T\omega(I_d-2sA\Omega)^{-1}A\omega\Delta). \tag{3.3}$$

Now, note that

$$(\Omega^{-1}-2sA)^{-1} = \Omega\sum_{j=0}^{\infty}(2s)^j(A\Omega)^j \tag{3.4}$$

for $||2sA\Omega|| < 1$, where $||.||$ is the matrix norm. Finally from (3.3) and (3.4) it follows that the necessary and sufficient conditions for the independence are $A\Omega\boldsymbol{h} = 0$ and $A\omega\Delta = 0$. □

Remark 3.3.1. *Taking $m = 1,\ \omega = I_d$ and defining $\Delta = \frac{\Omega\boldsymbol{\alpha}}{(1+\boldsymbol{\alpha}^T\Omega\boldsymbol{\alpha})^{\frac{1}{2}}}$ the condition $A\omega\Delta = \mathbf{0}$. becomes $\frac{A\Omega\boldsymbol{\alpha}}{(1+\boldsymbol{\alpha}^T\Omega\boldsymbol{\alpha})^{\frac{1}{2}}} = \mathbf{0}$ or $A\Omega\boldsymbol{\alpha} = \mathbf{0}$. Thus in this special case the conditions are $A\Omega\boldsymbol{h} = 0$ and $A\Omega\boldsymbol{\alpha} = \mathbf{0}$. These are the conditions for the independence obtained by Gupta and Huang (2002) for the Q.F of Azzalini's $SN_d(\Omega,\boldsymbol{\alpha})$ distribution.*

To study the independence between two quadratic forms we need the following lemma:

Lemma 3.3.1. *(Joint m.g.f of two quadratic forms of SUN density) Let $\boldsymbol{y} \sim SUN_{d,m}(\boldsymbol{\mu},\boldsymbol{\gamma},\bar{\boldsymbol{\omega}},\Omega^*)$,* *and A and B be $d \times d$ symmetric matrices. Consider the quadratic forms $Q_1(\boldsymbol{y}) = \boldsymbol{y}^T A \boldsymbol{y}$ and $Q_1(\boldsymbol{y}) = \boldsymbol{y}^T B \boldsymbol{y}$. Then the joint m.g.f of Q_1 and Q_2 is*

$$M_{\boldsymbol{y}^T A \boldsymbol{y}, \boldsymbol{y}^T B \boldsymbol{y}}(t,s)$$
$$= \frac{|I_d - 2(tA+sB)\Omega|^{-\frac{1}{2}} \quad \Phi_m(\boldsymbol{\gamma}; \Gamma + 2\Delta^T \omega (I_d - 2(tA+sB)\Omega)^{-1}(tA+sB)\omega\Delta)}{exp\ (t\boldsymbol{\mu}^T A\Omega(I_d - 2(tA+sB)\Omega)^{-1}\Omega^{-1}\boldsymbol{\mu}) \quad \Phi_m(\boldsymbol{\gamma};\Gamma)}, \quad t,s \in \Re. \tag{3.5}$$

Proof. By definition,

$$M_{\boldsymbol{y}^T A \boldsymbol{y}, \boldsymbol{y}^T B \boldsymbol{y}}(t,s) = E[exp(t\boldsymbol{y}^T A \boldsymbol{y} + s\boldsymbol{y}^T B \boldsymbol{y}]$$

$$= E[exp(\boldsymbol{y}^T (tA+sB)\boldsymbol{y}]$$

$$= M_{\boldsymbol{y}^T(tA+sB)\boldsymbol{y}}(1).$$

The result then follows from equation (3.1).

□

Theorem 3.3.2. *(Independence of two quadratic forms) Let $\boldsymbol{y} \sim SUN_{d,m}(\mathbf{0}, \boldsymbol{\gamma}, \bar{\boldsymbol{\omega}}, \Omega^*)$ and A and B be $d \times d$ symmetric matrices. Then the quadratic forms $Q_1 = \boldsymbol{y}^T A \boldsymbol{y}$ and $Q_2 = \boldsymbol{y}^T B \boldsymbol{y}$ are said to be independent if and only if $A\Omega B = 0_{d\times d}$, and $A\omega\Delta = 0_{d\times m} = B\omega\Delta$.*

Proof. From the equation (3.5), the joint m.g.f of Q_1 and Q_2 is

$$M_{Q_1,Q_2}(t,s)$$
$$= \frac{|I_d - 2(tA+sB)\Omega|^{-\frac{1}{2}} \quad \Phi_m(\boldsymbol{\gamma}; \Gamma + 2\Delta^T \omega (I_d - 2(tA+sB)\Omega)^{-1}(tA+sB)\omega\Delta)}{\Phi_m(\boldsymbol{\gamma};\Gamma)},$$
$$I_d - 2(tA+sB)\Omega > 0, \ t,s \in \Re.$$

Hence the m.g.f of Q_1, and Q_2 are $M_{Q_1,Q_2}(t,0)$, $M_{Q_1,Q_2}(0,s)$ and are obtained respectively as

$$M_{Q_1}(t) = \frac{|I_d - 2tA\Omega|^{-\frac{1}{2}} \quad \Phi_m(\boldsymbol{\gamma}; \Gamma + 2t\Delta^T \omega (I_d - 2tA\Omega)^{-1} A\omega\Delta)}{\Phi_m(\boldsymbol{\gamma};\Gamma)}, \ t \in \Re,$$

$$\text{and} \quad M_{Q_2}(s) = \frac{|I_d - 2sB\Omega|^{-\frac{1}{2}} \quad \Phi_m(\boldsymbol{\gamma}; \Gamma + 2s\Delta^T\omega(I_d - 2sB\Omega)^{-1}B\omega\Delta)}{\Phi_m(\boldsymbol{\gamma}; \Gamma)}, \ s \in \Re.$$

Now Q_1 and Q_2 are independent if and only if

$$M_{Q_1,Q_2}(t,s) = M_{Q_1}(t) \ M_{Q_2}(s).$$

That is, if

$$|I_d - 2(tA + sB)\Omega| \quad \frac{\Phi_m(\gamma; \Gamma + 2\Delta^T\omega(I_d - 2(tA + sB)\Omega)^{-1}(tA + sB)\omega\Delta)}{\Phi_m(\gamma; \Gamma)}$$

$$= |I_d - 2tA\Omega - 2sB\Omega + 4tsA\Omega B\Omega| \quad \frac{\Phi_m(\gamma; \Gamma + 2t\Delta^T\omega(I_d - 2tA\Omega)^{-1}A\omega\Delta)}{\Phi_m(\gamma; \Gamma)} \quad \times$$

$$\frac{\Phi_m(\gamma; \Gamma + 2s\Delta^T\omega(I_d - 2sB\Omega)^{-1}B\omega\Delta)}{\Phi_m(\gamma; \Gamma)}.$$

By imposing $A\omega\Delta = 0_{d\times m} = B\omega\Delta$, we observe that the denominator and numerator of the fraction in both sides of the above equation cancel out. Then it is seen that for the remaining parts to disappear one needs condition $A\Omega B = 0_{d\times d}$

Therefore, the conditions for the independence are

$(i) A\Omega B = \mathbf{0}$, and $(ii) A\omega\Delta = \mathbf{0} = B\omega\Delta$.

□

Remark 3.3.2. *The following results could be derived from Theorem 3.3.2.*

(i) Recall that we retain the multivariate normal distribution taking $\Delta = 0$ in the SUN density. When $\Delta = 0$, the only condition left in the above theorem is $A\Omega B = 0$, which is the condition required for the independence of two quadratic forms for mltivariate normal random vector.

(ii) Taking $m = 1$, $\omega = I_d$ and defining $\Delta = \frac{\Omega\boldsymbol{\alpha}}{(1+\boldsymbol{\alpha}^T\Omega\boldsymbol{\alpha})^{\frac{1}{2}}}$, the conditions $A\omega\Delta = \mathbf{0} = B\omega\Delta$ become $A\Omega\boldsymbol{\alpha} = \mathbf{0}$ and $B\Omega\boldsymbol{\alpha} = \mathbf{0}$. Thus in this special case conditions are $A\Omega\boldsymbol{h} = 0$ and $A\Omega\boldsymbol{\alpha} = \mathbf{0} = B\Omega\boldsymbol{\alpha} = \mathbf{0}$. The first condition was obtained by Gupta and Huang (2002) for

the independence of two Q.F of Azzalini's $SN_d(\Omega, \boldsymbol{\alpha})$ distribution. The latter two conditions are not necessary because the joint m.g.f for the two quadratic forms of Azzalini's $SN_d(\Omega, \boldsymbol{\alpha})$ distribution does not depend on $\boldsymbol{\alpha}$.

3.4 Expected Value of the Quadratic Form and their Functions

Theorem 3.4.1. *Let $\boldsymbol{y}$ be a random vector with a unified skew normal distribution, $SUN_{d,m}(\boldsymbol{\mu}, \boldsymbol{\gamma}, \bar{\boldsymbol{\omega}}, \Omega^*)$, where*

$$\Omega^* = \begin{pmatrix} \Gamma & \Delta^T \\ \Delta & \bar{\Omega} \end{pmatrix}.$$

Let $\Gamma = diag(\tau_1^2, ..., \tau_m^2)$, A and B be two symmetric $d \times d$ matrices. Then

$$E(\boldsymbol{y}^T A \boldsymbol{y}) = tr(A\Omega) + \sum_{j=1}^{m} \zeta_2(\tau_j^{-1}\gamma_j)\tau_j^{-2}\boldsymbol{\delta}_{.j}^T \omega A \omega \boldsymbol{\delta}_{.j} + (\boldsymbol{\mu} + \sum_{j=1}^{m} \zeta_1(\tau_j^{-1}\gamma_j)\tau_j^{-1}\omega\boldsymbol{\delta}_{.j})^T A$$
$$(\boldsymbol{\mu} + \sum_{j=1}^{m} \zeta_1(\tau_j^{-1}\gamma_j)\tau_j^{-1}\omega\boldsymbol{\delta}_{.j}),$$

where $\boldsymbol{\delta}_{.1}, ..., \boldsymbol{\delta}_{.m}$ are the columns of Δ and $\zeta_r(x)$ is the rth derivative of $\zeta_0(x) = log\{\Phi(x)\}$.

Proof. We have the following relation from Li (1987).

$$E(\boldsymbol{y}^T A \boldsymbol{y}) = tr(AM_2),$$

where M_2 is the second raw moment of $\boldsymbol{y}$ obtained from Theorem 2.4.3

$$= tr(A(\Omega + \sum_{j=1}^{m} \zeta_2(\tau_j^{-1}\gamma_j)\tau_j^{-2}\omega\boldsymbol{\delta}_{.j}\boldsymbol{\delta}_{.j}^T\omega + (\boldsymbol{\mu} + \sum_{j=1}^{m} \zeta_1(\tau_j^{-1}\gamma_j)\tau_j^{-1}\omega\boldsymbol{\delta}_{.j})$$
$$(\boldsymbol{\mu} + \sum_{j=1}^{m} \zeta_1(\tau_j^{-1}\gamma_j)\tau_j^{-1}\omega\boldsymbol{\delta}_{.j})^T))$$
$$= \quad tr(A\Omega) + \sum_{j=1}^{m} \zeta_2(\tau_j^{-1}\gamma_j)\tau_j^{-2}\boldsymbol{\delta}_{.j}^T\omega A \omega\boldsymbol{\delta}_{.j} + (\boldsymbol{\mu} + \sum_{j=1}^{m} \zeta_1(\tau_j^{-1}\gamma_j)\tau_j^{-1}\omega\boldsymbol{\delta}_{.j})^T$$

$$A(\boldsymbol{\mu} + \sum_{j=1}^{m} \zeta_1(\tau_j^{-1}\gamma_j)\tau_j^{-1}\omega\boldsymbol{\delta}_{.j}).$$

$\square$

Corollary 3.4.1. *Let* $\boldsymbol{y}$ *be a random vector with a unified skew normal distribution,* $SUN_{d,m}(\boldsymbol{\mu}, \mathbf{0}, \bar{\boldsymbol{\omega}}, \Omega^*)$*, where*

$$\Omega^* = \begin{pmatrix} I_m & \Delta^T \\ \Delta & \bar{\Omega} \end{pmatrix}.$$

Let A *be a symmetric* $d \times d$ *matrix. Then*

$$\mathbb{E}(\boldsymbol{y}^T A\boldsymbol{y}) = tr[A\Omega] + \boldsymbol{\mu}^T A\boldsymbol{\mu} + 2\sqrt{\frac{2}{\pi}}\boldsymbol{\mu}^T A\omega\Delta\mathbf{1}_m.$$

Proof. We have,

$$\begin{aligned}
& E(\boldsymbol{y}^T A\boldsymbol{y}) = tr(AM_2), \\
& \text{where } M_2 \text{ is the second raw moment of } \boldsymbol{y} \text{ obtained from Corollary 2.4.3} \\
= \; & tr\left(A\left(\Omega - \frac{2}{\pi}\omega\Delta\Delta^T\omega + \left(\boldsymbol{\mu} + \sqrt{\frac{2}{\pi}}\omega\Delta\mathbf{1}_m\right)\left(\boldsymbol{\mu} + \sqrt{\frac{2}{\pi}}\omega\Delta\mathbf{1}_m\right)^T\right)\right) \\
= \; & tr(A\Omega) - \frac{2}{\pi}(\mathbf{1}_m^T\Delta^T\omega)A(\omega\Delta\mathbf{1}_m) + \boldsymbol{\mu}^T A\boldsymbol{\mu} + \\
& \sqrt{\frac{2}{\pi}}\boldsymbol{\mu}^T A\omega\Delta\mathbf{1}_m + \sqrt{\frac{2}{\pi}}\mathbf{1}_m^T\Delta^T\omega A\boldsymbol{\mu} + \frac{2}{\pi}(\mathbf{1}_m^T\Delta^T\omega)A(\omega\Delta\mathbf{1}_m) \\
= \; & tr[A\Omega] + \boldsymbol{\mu}^T A\boldsymbol{\mu} + 2\sqrt{\frac{2}{\pi}}\boldsymbol{\mu}^T A\omega\Delta\mathbf{1}_m.
\end{aligned}$$

$\square$

When $m = 1$, the result in the above corollary reduces to the one obtained by Genton et al (2001). When $\Delta = 0_{d\times m}$,the result reduces to the one obtianed in case of MND.

3.4.1 Expected Value of the Ratio of Two Quadratic Forms

The ratio of two quadratic forms is used as an estimator in economic and time series data. The expected value of the ratio of two quadratic forms for a multivariate normal distribution was studied by Magnus (1986) and Gupta and Kabe (1998). The expected value of the ratio of quadratic forms for skew normal distribution has not been studied extensively. In this section we will obtain the expected value of the ratio of two quadratic forms in SUN density. Before presenting the main result, we present some lemmas which are needed to evaluate the expected value.

Lemma 3.4.1. *Let* $\boldsymbol{y}$ *be a random vector with a unified skew normal distribution,* $SUN_{d,m}(\boldsymbol{\mu}, \boldsymbol{\gamma}, \bar{\boldsymbol{\omega}}, \Omega^*)$, *where* $\Omega^* = \begin{pmatrix} \Gamma & \Delta^T \\ \Delta & \bar{\Omega} \end{pmatrix}$ *and assume* $\boldsymbol{\mu} = \mathbf{0}$ *. Then the m.g.f of* $Q = \boldsymbol{y}^T A \boldsymbol{y}$ *is obtained as*

$$M_Q(t) = \frac{\Phi_m(\boldsymbol{\gamma}; \Gamma + 2t\Delta^T \omega A \omega \Delta)}{|I_d - 2tA\Omega|^{\frac{1}{2}} \quad \Phi_m(\boldsymbol{\gamma}; \Gamma)}, \quad I_d - 2tA\Omega > 0. \tag{3.6}$$

Proof. From the equation (3.2) we have,

$$M_Q(t) = \frac{|I_d - 2tA\Omega|^{-\frac{1}{2}} \Phi_m(\boldsymbol{\gamma}; \Gamma - \Delta^T \bar{\Omega}^{-1} \Delta + \Delta^T \bar{\Omega}^{-1} \omega^{-1} (\Omega^{-1} - 2tA)^{-1} \omega^{-1} \bar{\Omega}^{-1} \Delta)}{exp\ (t\boldsymbol{\mu}^T A\Omega(I_d - 2tA\Omega)^{-1} \Omega^{-1} \boldsymbol{\mu}) \quad \Phi_m(\boldsymbol{\gamma}; \Gamma)}. \tag{3.7}$$

Using $\boldsymbol{\mu} = \mathbf{0}$ and noting that

$$(\Omega^{-1} - 2tA)^{-1} = \Omega \sum_{j=0}^{\infty} (2t)^j (A\Omega)^j$$

for $||2tA\Omega|| < 1$, where $||.||$ is the matrix norm. By ignoring 2nd and higher terms, the multinormal c.d.f in the numerator of (3.7) becomes,

$$\Phi_m(\boldsymbol{\gamma}; \Gamma - \Delta^T \bar{\Omega}^{-1} \Delta + \Delta^T \bar{\Omega}^{-1} \omega^{-1} (\Omega^{-1} - 2tA)^{-1} \omega^{-1} \bar{\Omega}^{-1} \Delta) = \Phi_m(\boldsymbol{\gamma}; \Gamma + 2t\Delta^T \omega A \omega \Delta).$$

The last identity was obtained from the relation $\Omega = \omega\bar{\Omega}\omega$ or equivalently $\bar{\Omega} = \omega^{-1}\Omega\omega^{-1}$. Hence from (3.7) we obtained the desired result. $\square$

Lemma 3.4.2. *Let $\boldsymbol{y}$ be a random vector with a unified skew normal distribution, $SUN_{d,m}(\boldsymbol{\mu}, \boldsymbol{\gamma}, \bar{\boldsymbol{\omega}}, \Omega^*)$, where $\Omega^* = \begin{pmatrix} \Gamma & \Delta^T \\ \Delta & \bar{\Omega} \end{pmatrix}$ and assume $\boldsymbol{\mu} = \mathbf{0}$. Then the joint m.g.f of $Q_1 = \boldsymbol{y}^T A \boldsymbol{y}$ and $Q_2 = \boldsymbol{y}^T B \boldsymbol{y}$ is obtained as*

$$M_{Q_1,Q_2}(t) = \frac{\Phi_m(\boldsymbol{\gamma}; \Gamma + 2\Delta^T \omega(tA + sB)\omega\Delta)}{|I_d - 2(tA + sB)\Omega|^{\frac{1}{2}} \quad \Phi_m(\boldsymbol{\gamma}; \Gamma)}, \quad I_d - 2(tA + sB)\Omega > 0. \tag{3.8}$$

Proof. The proof is similar to the proof of Lemma 3.3.1. $\square$

Lemma 3.4.3. *(Sawa, T. (1978))*

Let $\boldsymbol{y}$ be a random vector. Consider the quadratic forms $Q_1 = \boldsymbol{y}^T A \boldsymbol{y}$ and $Q_2 = \boldsymbol{y}^T B \boldsymbol{y}$, where A and B are $d \times d$ symmetric matrices. Define $R = \frac{Q_1}{Q_2} = \frac{\boldsymbol{y}^T A \boldsymbol{y}}{\boldsymbol{y}^T B \boldsymbol{y}}$. If $M(t, s)$ is the joint m.g.f of Q_1 and Q_2, then the kth order moment of R is given by

$$\mathbb{E}(R^k) = \frac{1}{\Gamma(k)} \int_0^\infty s^{k-1} \Big[\frac{\partial^k}{\partial^k t} M(t, -s)\Big]\Big|_{t=0}.$$

Lemma 3.4.4. *(Gupta and Nagar (1999))*

Let $t \in \Re$ and A be a symmetric matrix. Let the elements of A be differentiable functions of t. Then

$$\frac{\partial |A(t)|}{\partial t} = |A| \cdot tr\Big[A^{-1}\Big(\frac{\partial A}{\partial t}\Big)\Big].$$

Lemma 3.4.5. *(Gupta and Nagar (1999))*

Let A be a symmetric matrix and the elements of A be functions of t, where $t \in \Re$. Then

$$\frac{\partial A^{-1}}{\partial t} = -A^{-1}\Big(\frac{\partial A}{\partial t}\Big)A^{-1}$$

Main result

Let $\boldsymbol{y}$ be a random vector distributed as $SUN_{d,m}(\boldsymbol{\mu}, \boldsymbol{\gamma}, \bar{\boldsymbol{\omega}}, \Omega^*)$, where $\Omega^* = \begin{pmatrix} \Gamma & \Delta^T \\ \Delta & \bar{\Omega} \end{pmatrix}$ and assume $\boldsymbol{\mu} = \mathbf{0}$. Let $Q_1 = \boldsymbol{y}^T A \boldsymbol{y}$ and $Q_2 = \boldsymbol{y}^T B \boldsymbol{y}$ are quadratic forms in $\boldsymbol{y}$, where A and B are $d \times d$ symmetric matrices. Define $R = Q_1/Q_2 = \boldsymbol{y}^T A \boldsymbol{y}/\boldsymbol{y}^T B \boldsymbol{y}$. Then from Lemma 3.4.3, the kth order moment of R is given by,

$$\mathbb{E}(R^k) = \frac{1}{\Gamma(k)} \int_0^\infty s^{k-1} \big[\frac{\partial^k}{\partial^k t} M(t, -s)\big]\Big|_{t=0} \, \mathrm{d}s. \tag{3.9}$$

Using Lemma 3.4.2 in (3.9) we obtain,

$$\mathbb{E}(R^k) = \frac{1}{\Gamma(k)} \int_0^\infty s^{k-1} \left\{ \frac{\partial^k}{\partial^k t} \left[\frac{\Phi_m(\boldsymbol{\gamma}; \Gamma + 2\Delta^T \omega (tA - sB) \omega \Delta)}{|I_d - 2(tA - sB)\Omega|^{\frac{1}{2}} \quad \Phi_m(\boldsymbol{\gamma}; \Gamma)} \right] \Bigg|_{t=0} \right\} \mathrm{d}s$$

$$= \frac{1}{\Gamma(k) \ \Phi_m(\boldsymbol{\gamma}; \Gamma)} \int_0^\infty s^{k-1} \left\{ \frac{\partial^k}{\partial^k t} \left[\frac{\Phi_m(\boldsymbol{\gamma}; \Gamma + 2\Delta^T \omega (tA - sB) \omega \Delta)}{|I_d - 2(tA - sB)\Omega|^{\frac{1}{2}}} \right] \Bigg|_{t=0} \right\} \mathrm{d}s. \tag{3.10}$$

Theorem 3.4.2. *Suppose $k = 1$ in (3.10). Then the first moment of the ratio of two quadratic forms is obtained as*

$$\begin{aligned} \mathbb{E}(R) = \int_0^\infty \frac{|I_d + 2sB\Omega|^{-1/2}}{\Phi_m(\boldsymbol{\gamma}; \Gamma)} \Bigg\{ & tr\Big((I_d + 2sB\Omega)^{-1}\Omega A\Big) \Phi_m(\boldsymbol{\gamma}; \Gamma - 2\Delta^T \omega s B \omega \Delta) + \\ & \Big[\mathbb{E}^*_{\boldsymbol{x}}(2\boldsymbol{x}^T(\Gamma - 2\Delta^T \omega s B \omega \Delta)^{-1} \Delta^T \omega A \omega \Delta (\Gamma - 2\Delta^T \omega s B \omega \Delta)^{-1} \boldsymbol{x}) + \\ & \mathbb{E}^*_{\boldsymbol{x}}(2tr((\Gamma - 2\Delta^T \omega s B \omega \Delta)^{-1} \Delta^T \omega A \omega \Delta)) \Big] \Bigg\} \, \mathrm{d}s, \end{aligned}$$

where $\boldsymbol{x} \sim N_m(\mathbf{0}, \Gamma - 2\Delta^T \omega s B \omega \Delta)$.

Proof. Suppose $k = 1$ in (3.10). Then the first moment of R is given by,

$$\mathbb{E}(R) = \frac{1}{\Phi_m(\boldsymbol{\gamma}; \Gamma)} \int_0^\infty \left\{ \frac{\partial}{\partial t} \left[\frac{\Phi_m(\boldsymbol{\gamma}; \Gamma + 2\Delta^T \omega (tA + sB) \omega \Delta)}{|I_d - 2(tA - sB)\Omega|^{\frac{1}{2}}} \right] \Bigg|_{t=0} \right\} \mathrm{d}s. \tag{3.11}$$

Let

$$R = I_d - 2(tA - sB)\Omega = R(t,s),$$

$$V = \Gamma + 2\Delta^T \omega(tA - sB)\omega\Delta = V(t,s).$$

Then

$$\frac{\partial R}{\partial t} = -2\Omega A, \quad \frac{\partial V}{\partial t} = 2\Delta^T \omega A \omega \Delta.$$

With the notations defined above, the problem reduces to evaluate

$$\frac{\partial}{\partial t}\Big[|R|^{-\frac{1}{2}}\Phi_m(\boldsymbol{\gamma};V)\Big]\Big|_{t=0} = \Big[\Phi_m(\boldsymbol{\gamma};V)\frac{\partial}{\partial t}|R|^{-\frac{1}{2}} + |R|^{-\frac{1}{2}}\frac{\partial}{\partial t}\Phi_m(\boldsymbol{\gamma};V)\Big]\Big|_{t=0}. \tag{3.12}$$

Now using lemma (3.4.4) we get,

$$\begin{aligned}
\frac{\partial}{\partial t}|R|^{-\frac{1}{2}} &= -\frac{1}{2}|R|^{-\frac{3}{2}}\ \frac{\partial |R|}{\partial t} \\
&= -\frac{1}{2}|R|^{-\frac{3}{2}}|R|\ tr\left(R^{-1}\Big(\frac{\partial |R|}{\partial t}\Big)\right) \\
&= -\frac{1}{2}|R|^{-\frac{1}{2}}\ tr(R^{-1}(-2\Omega A)) \\
&= |R|^{-\frac{1}{2}}\ tr(R^{-1}\Omega A) \\
&= |I_d - 2(tA - sB)\Omega|^{-\frac{1}{2}}\ tr((I_d - 2(tA - sB)\Omega)^{-1}\Omega A).
\end{aligned}$$

The last identity was obtained plugging back the value of R.

Thus

$$\frac{\partial}{\partial t}|R|^{-\frac{1}{2}}\Big|_{t=0} = |I_d + 2(tA - sB)\Omega|^{-\frac{1}{2}}\ tr((I_d + 2sB)\Omega)^{-1}\Omega A).$$

Next

$$\begin{aligned}
\frac{\partial}{\partial t}\Phi_m(\boldsymbol{\gamma};V) &= \frac{\partial}{\partial t}\int_{-\infty}^{\gamma_1}\cdots\int_{-\infty}^{\gamma_m}\frac{|V|^{-\frac{1}{2}}}{(2\pi)^{m/2}}\ e^{-\frac{1}{2}\boldsymbol{x}^T V^{-1}\boldsymbol{x}}\mathrm{d}\boldsymbol{x} \\
&= C\int_{-\infty}^{\gamma_1}\cdots\int_{-\infty}^{\gamma_m}\left[|V|^{-\frac{1}{2}}\frac{\partial}{\partial t}e^{-\frac{1}{2}\boldsymbol{x}^T V^{-1}\boldsymbol{x}} + e^{-\frac{1}{2}\boldsymbol{x}^T V^{-1}\boldsymbol{x}}\frac{\partial}{\partial t}|V|^{-\frac{1}{2}}\right]\mathrm{d}\boldsymbol{x}
\end{aligned}$$

$$= \quad C\int_{-\infty}^{\gamma_1}\cdots\int_{-\infty}^{\gamma_m}\Bigg\{|V|^{-\frac{1}{2}}\ e^{-\frac{1}{2}\boldsymbol{x}^TV^{-1}\boldsymbol{x}}\Big[-\frac{1}{2}\boldsymbol{x}^T\Big(-V^{-1}\Big(\frac{\partial V}{\partial t}\Big)V^{-1}\Big)\boldsymbol{x}\Big]+$$

$$e^{-\frac{1}{2}\boldsymbol{x}^TV^{-1}\boldsymbol{x}}|V|^{-\frac{1}{2}}\ tr(V^{-1}\frac{\partial V}{\partial t})\Bigg\}\mathrm{d}\boldsymbol{x},$$

where lemma 3.4.5 was used to find the derivative of V^{-1}

$$= \quad \int_{-\infty}^{\gamma_1}\cdots\int_{-\infty}^{\gamma_m}\frac{|V|^{-\frac{1}{2}}\ e^{-\frac{1}{2}\boldsymbol{x}^TV^{-1}\boldsymbol{x}}}{(2\pi)^{m/2}}\Big[\boldsymbol{x}^T\Big(-V^{-1}\Big(\frac{\partial V}{\partial t}\Big)V^{-1}\Big)\boldsymbol{x}+tr(V^{-1}\frac{\partial V}{\partial t})\Big]\mathrm{d}\boldsymbol{x}.$$

Now plugging back the value of V and $\frac{\partial V}{\partial t}$ in the integrand, we obtain,

$$\frac{\partial}{\partial t}\Phi_m(\boldsymbol{\gamma};V)$$

$$= \int_{-\infty}^{\boldsymbol{\gamma}}\frac{|\Gamma+2\Delta^T\omega(tA-sB)\omega\Delta|^{-\frac{1}{2}}\ e^{-\frac{1}{2}\boldsymbol{x}^T(\Gamma+2\Delta^T\omega(tA-sB)\omega\Delta)^{-1}\boldsymbol{x}}}{(2\pi)^{m/2}}$$

$$\Big[\boldsymbol{x}^T\Big(-(\Gamma+2\Delta^T\omega(tA-sB)\omega\Delta)^{-1}\Big(2\Delta^T\omega A\omega\Delta\Big)(\Gamma+2\Delta^T\omega(tA-sB)\omega\Delta)^{-1}\Big)\boldsymbol{x}+$$

$$tr((\Gamma+2\Delta^T\omega(tA-sB)\omega\Delta)^{-1}(2\Delta^T\omega A\omega\Delta))\Big]\mathrm{d}\boldsymbol{x}.$$

Therefore

$$\frac{\partial}{\partial t}\Phi_m(\boldsymbol{\gamma};V)\Big|_{t=0}$$

$$= \int_{-\infty}^{\boldsymbol{\gamma}}\frac{|\Gamma-2\Delta^T\omega sB\omega\Delta|^{-\frac{1}{2}}\ e^{-\frac{1}{2}\boldsymbol{x}^T(\Gamma-2\Delta^T\omega sB\omega\Delta)^{-1}\boldsymbol{x}}}{(2\pi)^{m/2}}$$

$$\Big[\boldsymbol{x}^T\Big((\Gamma-2\Delta^T\omega sB\omega\Delta)^{-1}\Big(2\Delta^T\omega A\omega\Delta\Big)(\Gamma-2\Delta^T\omega sB\omega\Delta)^{-1}\Big)\boldsymbol{x}+$$

$$tr((\Gamma-2\Delta^T\omega sB\omega\Delta)^{-1}(2\Delta^T\omega A\omega\Delta))\Big]\mathrm{d}\boldsymbol{x}$$

$$= 2\mathbb{E}^*_{\boldsymbol{x}}\Big[\boldsymbol{x}^T\Big((\Gamma-2\Delta^T\omega sB\omega\Delta)^{-1}\Big(2\Delta^T\omega A\omega\Delta\Big)(\Gamma-2\Delta^T\omega sB\omega\Delta)^{-1}\Big)\boldsymbol{x}\Big]+$$

$$2\mathbb{E}^*_{\boldsymbol{x}}\Big[tr((\Gamma-2\Delta^T\omega sB\omega\Delta)^{-1}(\Delta^T\omega A\omega\Delta))\Big],$$

where $\boldsymbol{x}\sim N_m(\mathbf{0},\Gamma-2\Delta^T\omega sB\omega\Delta)$ and the notation $\mathbb{E}^*_{\boldsymbol{x}}(.)$ is introduced to denote incomplete expectation of of (.).

Therefore from (3.12)

$$
\begin{aligned}
&\frac{\partial}{\partial t}\Big[|R|^{-\frac{1}{2}}\Phi_m(\boldsymbol{\gamma};V)\Big]\Big|_{t=0} \\
= &\Big[\Phi_m(\boldsymbol{\gamma};V)\frac{\partial}{\partial t}|R|^{-\frac{1}{2}} + |R|^{-\frac{1}{2}}\frac{\partial}{\partial t}\Phi_m(\boldsymbol{\gamma};V)\Big]\Big|_{t=0} \\
= &\Big[\Phi_m(\boldsymbol{\gamma};V)|I_d + 2(tA - sB)\Omega|^{-\frac{1}{2}}\ tr((I_d + 2sB)\Omega)^{-1}\Omega A) + |I_d + 2sB)\Omega|^{-\frac{1}{2}} \\
&\Big\{2\mathbb{E}^*_{\boldsymbol{x}}\Big[\boldsymbol{x}^T\Big((\Gamma - 2\Delta^T\omega sB\omega\Delta)^{-1}\Big(2\Delta^T\omega A\omega\Delta\Big)(\Gamma - 2\Delta^T\omega sB\omega\Delta)^{-1}\Big)\boldsymbol{x}\Big] + \\
&2\mathbb{E}^*_{\boldsymbol{x}}\Big[tr((\Gamma - 2\Delta^T\omega sB\omega\Delta)^{-1}(\Delta^T\omega A\omega\Delta))\Big]\Big\}.
\end{aligned}
$$

Finally from (3.11) we obtain,

$$
\begin{aligned}
\mathbb{E}(R) = \int_0^\infty \frac{|I_d + 2sB\Omega|^{-1/2}}{\Phi_m(\boldsymbol{\gamma};\Gamma)}&\Big\{tr\Big((I_d + 2sB\Omega)^{-1}\Omega A\Big)\Phi_m(\boldsymbol{\gamma};\Gamma - 2\Delta^T\omega sB\omega\Delta) + \\
&\Big[\mathbb{E}^*_{\boldsymbol{x}}(2\boldsymbol{x}^T(\Gamma - 2\Delta^T\omega sB\omega\Delta)^{-1}\Delta^T\omega A\omega\Delta(\Gamma - 2\Delta^T\omega sB\omega\Delta)^{-1}\boldsymbol{x}) + \\
&\mathbb{E}^*_{\boldsymbol{x}}(2tr((\Gamma - 2\Delta^T\omega sB\omega\Delta)^{-1}\Delta^T\omega A\omega\Delta))\Big]\Big\}\ \mathrm{d}s.
\end{aligned}
$$

□

CHAPTER 4

ESTIMATION OF PARAMETERS OF SUN DENSITY: METHOD OF WEIGHTED MOMENTS

4.1 Introduction

Concerning the inference of the SUN density, the first problem is how to estimate its parameters based on a sample of observations. Although the method of maximum likelihood estimation is very useful in various estimation problems, use of this method to estimate SUN parameters could become cumbersome because of the large number of parameters involved with this density. On the other hand, in order to apply the SUN density to real data, one needs to be able to implement an easy estimation method that provides reliable estimators for SUN parameters. Among other existing alternatives "the method of moments" (MOM) is often used because it leads to very simple computations and provides consistent estimators although not as efficient as the maximum likelihood estimators. In this chapter, we use the method of moments (weighted) to estimate the parameters of SUN density. We will observe that even the MOM estimation to the general form of SUN density may not achieve the goal of estimating all its parameters. The aim could be achieved in some specific cases as would be discussed in the subsequent sections.

4.2 Estimation Method

To estimate the SUN density parameters by MOM, we need explicit expressions for the moments of the density in terms of unknown parameters. As pointed out in chapter 2, the mathematical expressions for the higher moments of SUN density are very complex and have no practical use. Moreover, estimations of the third and the fourth SUN moments are contaminated by large variances. In addition, in the univariate case for the values of the skewness parameter near zero, the third moment gets closer to zero and optimization becomes complicated and sometimes even impossible. Thus use of third or fourth moments may not produce accurate estimates. Following Flecher (2009) we will use weighted moments method (WMOM) to estimate the unknown parameters. In order to obtain the weighted moments we need the following results.

Lemma 4.2.1. *Consider two multivariate normal cumulative distribution functions (cdf)* $\Phi_d(Ab;\boldsymbol{\mu},\Sigma)$ *and* $\Phi_m(Bb;\nu,\Gamma)$*, where* A *and* B *are* $d\times d$ *and* $m\times d$ *matrices respectively,* Σ *and* Γ *are* $d\times d$ *and* $m\times m$ *matrices,* $\boldsymbol{b}$ *and* $\boldsymbol{\mu}$ *are* $d\times 1$ *vectors, and* $\boldsymbol{\nu}$ *is an* $m\times 1$ *vector. Then*

$$\Phi_d(A\boldsymbol{b};\boldsymbol{\mu},\Sigma)\ \Phi_m(B\boldsymbol{b};\boldsymbol{\nu},\Gamma)=\Phi_{d+m}(Cb;\boldsymbol{\gamma},V), \tag{4.1}$$

where $C_{(d+m)\times d}=\begin{bmatrix}A\\ B\end{bmatrix}$, $\boldsymbol{\gamma}_{(d+m)\times 1}=\begin{pmatrix}\boldsymbol{\mu}\\ \boldsymbol{\nu}\end{pmatrix}$ *and* $V_{(d+m)\times(d+m)}=\begin{pmatrix}\Sigma & 0\\ 0 & \Gamma\end{pmatrix}$.

Proof.

$$\begin{aligned}&R.H.S\\ =\ &\Phi_{d+m}(C\boldsymbol{b};\boldsymbol{\gamma},V)\end{aligned}$$

$$= \quad P(\boldsymbol{y} \leq C\boldsymbol{b}),$$

where

$$y = \begin{pmatrix} \boldsymbol{y}_1 \\ \boldsymbol{y}_2 \end{pmatrix} \sim N_{d+m}\left[\begin{pmatrix} \boldsymbol{\mu} \\ \boldsymbol{\nu} \end{pmatrix}, \quad \begin{pmatrix} \Sigma & 0 \\ 0 & \Gamma \end{pmatrix}\right]$$

and $\boldsymbol{y}_1$ and $\boldsymbol{y}_2$ are independently distributed as $\boldsymbol{y}_1 \sim N_d(\boldsymbol{\mu}, \Sigma)$ and $\boldsymbol{y}_2 \sim N_m(\nu, \Gamma)$. Therefore

$$\begin{aligned} P(\boldsymbol{y} \leq C\boldsymbol{b}) = P\left(\begin{pmatrix} \boldsymbol{y}_1 \\ \boldsymbol{y}_2 \end{pmatrix} \leq \begin{pmatrix} A\boldsymbol{b} \\ B\boldsymbol{b} \end{pmatrix}\right) &= P(\boldsymbol{y}_1 \leq A\boldsymbol{b}, \boldsymbol{y}_2 \leq B\boldsymbol{b}) \\ &= \Phi_d(A\boldsymbol{b}; \boldsymbol{\mu}, \Sigma) \cdot \Phi_m(B\boldsymbol{b}; \boldsymbol{\nu}, \Gamma). \end{aligned}$$

□

Theorem 4.2.1. *Let* $\boldsymbol{y}$ *be a SUN random vector distributed as* $\boldsymbol{y} \sim SUN_{d,m}(\boldsymbol{\mu}, \boldsymbol{\gamma}, \bar{\boldsymbol{\omega}}, \Omega^*)$ *with the density*

$$f(\boldsymbol{y}) = \phi_d(\boldsymbol{y} - \boldsymbol{\mu}; \Omega)\frac{\Phi_m(\boldsymbol{\gamma} + \Delta^T\bar{\Omega}^{-1}\omega^{-1}(\boldsymbol{y} - \boldsymbol{\mu}); \Gamma - \Delta^T\bar{\Omega}^{-1}\Delta)}{\Phi_m(\boldsymbol{\gamma}; \Gamma)},$$

where $\Omega = \omega\bar{\Omega}\omega$, $\bar{\boldsymbol{\omega}} = \omega\mathbf{1}_{\boldsymbol{d}}$ *and* $\Omega^* = \begin{pmatrix} \Gamma & \Delta^T \\ \Delta & \bar{\Omega} \end{pmatrix}$. *Also let* $h(\boldsymbol{y}) = h(y_1, \dots y_d)$ *be any real valued function such that* $\mathbb{E}(h(\boldsymbol{y}))$ *is finite, then*

$$\mathbb{E}(h(\boldsymbol{y})\ \Phi_d^r(\boldsymbol{y}; \mathbf{0}, \boldsymbol{I_d})) \quad = \quad \frac{\Phi_{rd+m}(\boldsymbol{\gamma}^\dagger; \Gamma^\dagger)}{\Phi_m(\boldsymbol{\gamma}; \Gamma)}\ \mathbb{E}(h(\boldsymbol{y}^\dagger)), \tag{4.2}$$

where $\boldsymbol{y}^\dagger$ *follows a* $SUN_{d,rd+m}(\boldsymbol{\mu}, \boldsymbol{\gamma}^\dagger, \bar{\boldsymbol{\omega}}, \Omega^{*\dagger})$ *with* $\boldsymbol{\gamma}^{\dagger T}_{(rd+m)\times 1} = \begin{pmatrix} \Delta_*^T\boldsymbol{\mu} & \boldsymbol{\gamma} \end{pmatrix}$; Δ_* *is a* $d \times rd$ *matrix defined by* $\Delta_* = (I_d, \dots I_d)$, $\Delta^{*T}_{d\times(rd+m)} = \begin{pmatrix} \Delta_*^T\omega\bar{\Omega} \\ \Delta^T \end{pmatrix}$, $\Gamma^\dagger = \begin{pmatrix} I_{rd} + \Delta_*^T\Omega\Delta_* & \Delta_*^T\omega\Delta \\ \Delta^T\omega\Delta_* & \Gamma \end{pmatrix}$

and

$$\Omega^{*\dagger} = \begin{pmatrix} \Gamma^\dagger & \Delta^{\dagger T} \\ \\ \Delta^\dagger & \bar{\Omega} \end{pmatrix}.$$

Proof.

$$\begin{aligned}
& \mathbb{E}(h(\boldsymbol{y})\ \Phi_d^r(\boldsymbol{y};\mathbf{0},I_d)) \\
= \ & K\int_{\Re^d} h(\boldsymbol{y})\Phi_d^r(\boldsymbol{y};\mathbf{0},\boldsymbol{I_d})\phi_d(\boldsymbol{y}-\boldsymbol{\mu};\Omega)\Phi_m(\boldsymbol{\gamma}+\Delta^T\bar{\Omega}^{-1}\omega^{-1}(\boldsymbol{y}-\boldsymbol{\mu});\Gamma-\Delta^T\bar{\Omega}^{-1}\Delta)\ \mathrm{d}\boldsymbol{y}, \\
& \text{where } K^{-1}=\Phi_m(\boldsymbol{\gamma};\Gamma) \text{ is a constant} \\
= \ & K\int_{\Re^d} h(\boldsymbol{y})\phi_d(\boldsymbol{y}-\boldsymbol{\mu};\Omega)\Phi_{rd}(\Delta_*^T(\boldsymbol{y}-\boldsymbol{\mu});-\Delta_*^T\boldsymbol{\mu},I_{rd}) \quad \times \\
& \Phi_m(\Delta^T\bar{\Omega}^{-1}\omega^{-1}(\boldsymbol{y}-\boldsymbol{\mu});-\boldsymbol{\gamma},\Gamma-\Delta^T\bar{\Omega}^{-1}\Delta)\ \mathrm{d}\boldsymbol{y}, \\
& \text{where } \Delta_* \text{ is defined as above.}
\end{aligned}$$

Now using lemma 4.2.1 we have,

$$\begin{aligned}
& \Phi_{rd}(\Delta_*^T(\boldsymbol{y}-\boldsymbol{\mu});-\Delta_*^T\boldsymbol{\mu},\boldsymbol{I_{rd}})\Phi_m(\Delta^T\bar{\Omega}^{-1}\omega^{-1}(\boldsymbol{y}-\boldsymbol{\mu});-\boldsymbol{\gamma},\Gamma-\Delta^T\bar{\Omega}^{-1}\Delta) \\
= \ & \Phi_{rd+m}\left[\begin{pmatrix} \Delta_*^T \\ \Delta^T\bar{\Omega}^{-1}\omega^{-1} \end{pmatrix}(\boldsymbol{y}-\boldsymbol{\mu});\begin{pmatrix} -\Delta_*^T\boldsymbol{\mu} \\ -\boldsymbol{\gamma} \end{pmatrix}\begin{pmatrix} \boldsymbol{I_{rd}} & \mathbf{0} \\ \mathbf{0} & \Gamma-\Delta^T\bar{\Omega}^{-1}\Delta) \end{pmatrix}\right] \\
= \ & \Phi_{rd+m}\left[\begin{pmatrix} \Delta_*^T\omega\bar{\Omega} \\ \Delta^T \end{pmatrix}\bar{\Omega}^{-1}\omega^{-1}(\boldsymbol{y}-\boldsymbol{\mu});\begin{pmatrix} -\Delta_*^T\boldsymbol{\mu} \\ -\boldsymbol{\gamma} \end{pmatrix}\begin{pmatrix} \boldsymbol{I_{rd}} & \mathbf{0} \\ \mathbf{0} & \Gamma-\Delta^T\bar{\Omega}^{-1}\Delta) \end{pmatrix}\right] \\
= \ & \Phi_{rd+m}(\Delta^{\dagger T}\bar{\Omega}^{-1}\omega^{-1}(\boldsymbol{y}-\boldsymbol{\mu});-\boldsymbol{\gamma}^\dagger,\Gamma^\dagger-\Delta^{\dagger T}\bar{\Omega}^{-1}\Delta^\dagger) \\
= \ & \Phi_{rd+m}(\boldsymbol{\gamma}^\dagger+\Delta^{\dagger T}\bar{\Omega}^{-1}\omega^{-1}(\boldsymbol{y}-\boldsymbol{\mu});\Gamma^\dagger-\Delta^{\dagger T}\bar{\Omega}^{-1}\Delta^\dagger),
\end{aligned}$$

where $\boldsymbol{\gamma}^\dagger,\ \Delta^\dagger,\ \Gamma^\dagger$ are defined in the statement of the theorem. Therefore,

$$\begin{aligned}
& \mathbb{E}(h(\boldsymbol{y})\ \Phi_d^r(\boldsymbol{y};\mathbf{0},I_d)) \\
= \;& K\int_{\Re^d} h(\boldsymbol{y})\phi_d(\boldsymbol{y}-\boldsymbol{\mu};\Omega)\Phi_{rd+m}(\boldsymbol{\gamma}^\dagger+\Delta^{\dagger T}\bar{\Omega}^{-1}\omega^{-1}(\boldsymbol{y}-\boldsymbol{\mu});\Gamma^\dagger-\Delta^{\dagger T}\bar{\Omega}^{-1}\Delta^\dagger)\ \mathrm{d}\boldsymbol{y} \\
= \;& \frac{\Phi_{rd+m}(\boldsymbol{\gamma}^\dagger;\Gamma^\dagger)}{\Phi_m(\boldsymbol{\gamma};\Gamma)}\int_{\Re^d} h(\boldsymbol{y})\phi_d(\boldsymbol{y}-\boldsymbol{\mu};\Omega)\frac{\Phi_{rd+m}(\boldsymbol{\gamma}^\dagger+\Delta^{\dagger T}\bar{\Omega}^{-1}\omega^{-1}(\boldsymbol{y}-\boldsymbol{\mu});\Gamma^\dagger-\Delta^{\dagger T}\bar{\Omega}^{-1}\Delta^\dagger)}{\Phi_{rd+m}(\boldsymbol{\gamma}^\dagger;\Gamma^\dagger)}\ \mathrm{d}y \\
= \;& \frac{\Phi_{rd+m}(\boldsymbol{\gamma}^\dagger;\Gamma^\dagger)}{\Phi_m(\boldsymbol{\gamma};\Gamma)}\mathbb{E}(h(\boldsymbol{y}^\dagger)),
\end{aligned}$$

where $\boldsymbol{y}^\dagger \sim SUN_{d,rd+m}(\boldsymbol{\mu},\boldsymbol{\gamma}^\dagger,\bar{\boldsymbol{\omega}},\Omega^{*\dagger})$. □

Remark 4.2.1. *The following calculation was done in Theorem 4.2.1.*

- *Note that*

$$\begin{aligned}
& \Gamma^\dagger-\Delta^{\dagger T}\bar{\Omega}^{-1}\Delta^\dagger \\
= \;& \begin{pmatrix} I_{rd}+\Delta_*^T\Omega\Delta_* & \Delta_*^T\omega\Delta \\ \Delta^T\omega\Delta_* & \Gamma \end{pmatrix} - \begin{pmatrix} \Delta_*^T\omega\bar{\Omega} \\ \Delta^T \end{pmatrix}\bar{\Omega}^{-1}\begin{pmatrix} \bar{\Omega}\omega\Delta_* & \Delta \end{pmatrix} \\
= \;& \begin{pmatrix} I_{rd}+\Delta_*^T\Omega\Delta_* & \Delta_*^T\omega\Delta \\ \Delta^T\omega\Delta_* & \Gamma \end{pmatrix} - \begin{pmatrix} \Delta_*^T\Omega\Delta_* & \Delta_*^T\omega\Delta \\ \Delta^T\omega\Delta_* & \Delta^T\bar{\Omega}^{-1}\Delta \end{pmatrix} \\
= \;& \begin{pmatrix} I_{rd} & \mathbf{0} \\ \mathbf{0} & \Gamma-\Delta^T\bar{\Omega}^{-1}\Delta \end{pmatrix}.
\end{aligned}$$

- *The independence condition required by Lemma 4.2.1 is satisfied here as shown below*

$$\Phi_{rd+m}\left[\begin{pmatrix} \Delta_*^T\omega\bar{\Omega} \\ \Delta^T \end{pmatrix}\bar{\Omega}^{-1}\omega^{-1}(\boldsymbol{y}-\boldsymbol{\mu}); \begin{pmatrix} -\Delta_*^T\boldsymbol{\mu} \\ -\boldsymbol{\gamma} \end{pmatrix}\begin{pmatrix} I_{rd} & \mathbf{0} \\ \mathbf{0} & \Gamma-\Delta^T\bar{\Omega}^{-1}\Delta) \end{pmatrix}\right]$$

$$
\begin{aligned}
&= \Phi_{rd+m}\left[\begin{pmatrix}\Delta_*^T\\ \Delta^T\bar{\Omega}^{-1}\omega^{-1}\end{pmatrix}(\boldsymbol{y}-\boldsymbol{\mu}); \begin{pmatrix}-\Delta_*^T\boldsymbol{\mu}\\ -\boldsymbol{\gamma}\end{pmatrix}\begin{pmatrix}I_{rd} & \mathbf{0}\\ \mathbf{0} & \Gamma-\Delta^T\bar{\Omega}^{-1}\Delta)\end{pmatrix}\right]\\
&= P\left[X=\begin{pmatrix}\boldsymbol{x}_1\\ \boldsymbol{x}_2\end{pmatrix}\le\begin{pmatrix}\Delta_*^T\\ \Delta^T\bar{\Omega}^{-1}\omega^{-1}\end{pmatrix}(\boldsymbol{y}-\boldsymbol{\mu})\right]\\
&= \Phi_{rd}(\Delta_*^T(\boldsymbol{y}-\boldsymbol{\mu}); -\Delta_*^T\boldsymbol{\mu}, I_{rd})\Phi_m(\Delta^T\bar{\Omega}^{-1}\omega^{-1}(\boldsymbol{y}-\boldsymbol{\mu}); -\boldsymbol{\gamma}, \Gamma-\Delta^T\bar{\Omega}^{-1}\Delta).
\end{aligned}
$$

The last two equalities are obtained by observing that

$$
\boldsymbol{x}=\begin{pmatrix}\boldsymbol{x}_1\\ \boldsymbol{x}_2\end{pmatrix}\sim N_{rd+m}\left[\begin{pmatrix}-\Delta_*^T\boldsymbol{\mu}\\ -\boldsymbol{\gamma}\end{pmatrix}\begin{pmatrix}I_{rd} & \mathbf{0}\\ \mathbf{0} & \Gamma-\Delta^T\bar{\Omega}^{-1}\Delta\end{pmatrix}\right]
$$

and $\boldsymbol{x}_1$ *and* $\boldsymbol{x}_2$ *are independently distributed as* $\boldsymbol{x}_1 \sim N_{rd}(-\Delta_*^T\boldsymbol{\mu}, I_{rd})$ *and* $\boldsymbol{x}_2 \sim N_m(-\boldsymbol{\gamma}, \Gamma-\Delta^T\bar{\Omega}^{-1}\Delta)$.

Corollary 4.2.1. *Let* $y_k, k=1,2,\dots,n$ *be univariate independent and identical (iid) random variables distributed as* $SUN_{1,1}(\mu, 0, w, \Omega^*)$, *where* $\Omega^*=\begin{pmatrix}1 & \delta\\ \delta & v\end{pmatrix}$ *is the correlation matrix. Then*

- *Defining* $h(y)=y_i, i=1,2,\dots,n$ *we obtain*

$$
\mathbb{E}(y_i\Phi(y_i)] = 2\mu\Phi_2\left[\begin{pmatrix}\mu\\ 0\end{pmatrix}; \begin{pmatrix}1+\sigma^2 & w\delta\\ w\delta & 1\end{pmatrix}\right]+\Phi_2^*\left[\begin{pmatrix}\mu\\ 0\end{pmatrix}; \begin{pmatrix}1+\sigma^2 & w\delta\\ w\delta & 1\end{pmatrix}\right], \sigma^2=\omega v\omega.
$$

- *Defining* $h(y)=y_i^2, i=1,2,\dots,n$ *we obtain*

$$
\mathbb{E}(y_i^2\Phi(y_i)] = 2\Phi_2\left[\begin{pmatrix}\mu\\ 0\end{pmatrix}; \begin{pmatrix}1+\sigma^2 & w\delta\\ w\delta & 1\end{pmatrix}\right]\Big(\mu^2+\sigma^2\Big)+2\mu\Phi_2^*\left[\begin{pmatrix}\mu\\ 0\end{pmatrix}; \begin{pmatrix}1+\sigma^2 & w\delta\\ w\delta & 1\end{pmatrix}\right],
$$

where

$$\Phi_2^*\left[\begin{pmatrix}\mu\\ 0\end{pmatrix};\begin{pmatrix}1+\sigma^2 & w\delta\\ w\delta & 1\end{pmatrix}\right. = 2\sigma^2\phi\Big(\mu;1+\sigma^2\Big)\Phi\Big(\frac{-\mu w^2\delta}{1+\sigma^2};1+\sigma^2-w^2\delta^2\Big)+$$

$$2\delta\frac{1}{\sqrt{2\pi}}\Phi\Big(\mu;1+\sigma^2-w^2\delta^2\Big).$$

Proof. Taking $r=m=d=1$, in Theorem 4.2.1 we have,

- For the first part,

$$\begin{aligned} & \mathbb{E}(\Phi_1(y)) \\ = \; & \frac{\Phi_2(\gamma^\dagger;\Gamma^\dagger)}{\Phi_1(0;1)}\mathbb{E}(y^\dagger); \\ & \text{where } y^\dagger \sim SUN_{1,2}(\mu,\boldsymbol{\gamma}^\dagger,\bar{\omega},\Omega^{*\dagger}) \; with \; \boldsymbol{\gamma}^\dagger=(\mu,0)^T,\Gamma^\dagger=\begin{pmatrix}1+\sigma^2 & w\delta\\ w\delta & 1\end{pmatrix}. \end{aligned}$$

  Now plugging in the value of $\mathbb{E}(y^\dagger)$ and using (2.9), we obtain

$$\mathbb{E}(\Phi_1(y)) = 2\mu\Phi_2(\gamma^\dagger;\Gamma^\dagger)+2\Phi_2^*(\gamma^\dagger;\Gamma^\dagger).$$

  The value of $\Phi_2^*(\gamma^\dagger;\Gamma^\dagger)$ is calculated following Example 2.4.1.

- For the second part,

$$\begin{aligned} & \mathbb{E}(y^2\Phi_1(y)) \\ = \; & \frac{\Phi_2(\gamma^\dagger;\Gamma^\dagger)}{\Phi_1(0;1)}\mathbb{E}(y^{\dagger 2}); \quad \text{where } \boldsymbol{\gamma}^\dagger=(\mu,0)^T,\Gamma^\dagger=\begin{pmatrix}1+\sigma^2 & w\delta\\ w\delta & 1\end{pmatrix}. \end{aligned}$$

  Now plugging in the value of $\mathbb{E}(y^{\dagger 2})$ and using (2.10) and $\Phi_2^*(\gamma^\dagger;\Gamma^\dagger)$, we obtain the desired result.

$\square$

Corollary 4.2.2. *Suppose* $\boldsymbol{y} \sim SUN_{d,d}(\boldsymbol{\mu}, \mathbf{0}, \bar{\boldsymbol{\omega}}, \Omega^*)$ *with* $\Omega^* = \begin{pmatrix} I_d & \Delta^T \\ \Delta & \bar{\Omega} \end{pmatrix}$ *and* $h(\boldsymbol{y}) = 1$. *Then*

$$\mathbb{E}(\Phi_d(\boldsymbol{y}, \mathbf{0}, I_d)) = 2^d \Phi_{2d}\left[\begin{pmatrix} \boldsymbol{\mu} \\ 0 \end{pmatrix}; \begin{pmatrix} I_d + \omega\bar{\Omega}\omega & \omega\Delta^T \\ \omega\Delta & I_d \end{pmatrix}\right].$$

Proof. Taking $r = 1$ and $m = d$, in Theorem 4.2.1 we have,

$$\begin{aligned}
& \mathbb{E}(\Phi_d(\boldsymbol{y}, \mathbf{0}, I_d)) \\
= \; & \frac{\Phi_{d+d}(\boldsymbol{\gamma}^\dagger; \Gamma^\dagger)}{\Phi_d(\boldsymbol{\gamma}; \Gamma)} \mathbb{E}(h(\boldsymbol{y}^\dagger)); \quad \text{where } \boldsymbol{\gamma}^\dagger, \Gamma^\dagger, \text{and } \boldsymbol{y}^\dagger \text{ are defined in Theorem 4.2.1.} \\
& \text{With } h(\boldsymbol{y}) = 1 \text{ and } \boldsymbol{y} \sim SUN_{d,d}(\boldsymbol{\mu}, \mathbf{0}, \bar{\boldsymbol{\omega}}, \Omega^*), \\
& \mathbb{E}(\Phi_d(\boldsymbol{y}, \mathbf{0}, I_d)) \\
= \; & \frac{\Phi_{2d}(\boldsymbol{\gamma}^\dagger; \Gamma^\dagger)}{\Phi_d(\mathbf{0}; I_d)} \\
= \; & 2^d \Phi_{2d}\left[\begin{pmatrix} \boldsymbol{\mu} \\ 0 \end{pmatrix}; \begin{pmatrix} I_d + \omega\bar{\Omega}\omega & \omega\Delta^T \\ \omega\Delta & I_d \end{pmatrix}\right].
\end{aligned}$$

□

4.2.1 Estimation of Univariate SUN Density

Let $y_k, k = 1, 2, \ldots, n$ be univariate independent and identical (iid) random variables distributed as $SUN_{1,1}(\mu, 0, w, \Omega^*)$ with the correlation matrix $\Omega^* = \begin{pmatrix} 1 & \delta \\ \delta & v \end{pmatrix}$. Thus we have unknown parameters μ, w, v and δ. To estimate these parameters we need four equations. From Theorem 2.4.3 the first two moments are given by

$$\mathbb{E}(y_i) = \mu + \delta\omega\sqrt{\frac{2}{\pi}}$$

and

$$Var(y_i) = \sigma^2 - \delta^2\omega^2\frac{2}{\pi}, \text{where } \sigma^2 = wvw.$$

For the remaining two equations we could use third and fourth moment accordingly. However, the third and fourth moments do not have simple form even in the univariate case. Flecher (2009) also pointed out that estimation of higher order moments are classically contaminated by large variances. Therefore we use weighted moments instead of the higher order moments. The remaining two weighted moments are obtained from Corollary 4.2.1.

By the method of moment approach we need to equate these moments to the sample moments. The sample mean and the sample variance which have usual expressions of $\bar{y}_n = \sum_{i=1}^n y_i/n$ and $S_n^2 = \frac{1}{n-1}\sum_{i=1}^n (y_i - \bar{y}_n)^2$ respectively can be plugged in the above first two moment expressions. For the weighted moments $\mathbb{E}(y_i\Phi(y_i)]$ and $\mathbb{E}(y_i^2\Phi(y_i)]$, we will use the unbiased statistics $m_{1n} = \frac{1}{n}\sum_{i=1}^n y_i\Phi(y_i)$ and $m_{2n} = \frac{1}{n}\sum_{i=1}^n y_i^2\Phi(y_i)$. Thus in terms of estimation the following system of four equations with four unknown parameters follows

$$\begin{cases} \bar{y_n} = \hat{\mu} + \hat{\delta}\hat{w}\sqrt{\frac{2}{\pi}} \\ S_n^2 = \hat{\sigma}^2 - \hat{\delta}^2\hat{w}^2\frac{2}{\pi} \\ m_{1n} = 2\hat{\mu}\Phi_2\left[\begin{pmatrix}\hat{\mu} \\ 0\end{pmatrix}; \begin{pmatrix} 1+\hat{\sigma}^2 & \hat{w}\hat{\delta} \\ \hat{w}\hat{\delta} & 1\end{pmatrix}\right] + 2\hat{\sigma}^2\phi\Big(\hat{\mu}; 1+\hat{\sigma}^2\Big)\Phi\Big(\frac{-\hat{\mu}\hat{w}^2\hat{\delta}}{1+\hat{\sigma}^2}; 1+\hat{\sigma}^2-\hat{w}^2\hat{\delta}^2\Big) \\ \qquad + 2\hat{\delta}\frac{1}{\sqrt{2\pi}}\Phi\Big(\hat{\mu}; 1+\hat{\sigma}^2-\hat{w}^2\hat{\delta}^2\Big) \\ m_{2n} = 2\Phi_2\left[\begin{pmatrix}\hat{\mu} \\ 0\end{pmatrix}; \begin{pmatrix} 1+\hat{\sigma}^2 & \hat{w}\hat{\delta} \\ \hat{w}\hat{\delta} & 1\end{pmatrix}\right]\Big(\hat{\mu}^2+\hat{\sigma}^2\Big) + 4\mu\hat{\sigma}^2\phi\Big(\hat{\mu}; 1+\hat{\sigma}^2\Big) \times \\ \qquad \Phi\Big(\frac{-\hat{\mu}\hat{w}^2\hat{\delta}}{1+\hat{\sigma}^2}; 1+\hat{\sigma}^2-\hat{w}^2\hat{\delta}^2\Big) + 2\hat{\delta}\frac{1}{\sqrt{2\pi}}\Phi\Big(\hat{\mu}; 1+\hat{\sigma}^2-\hat{w}^2\hat{\delta}^2\Big). \end{cases} \tag{4.3}$$

4.2.2 Estimation of Parameters of Multivariate SUN Distribution

Following our univariate procedure we estimate the parameters of the density $SUN_{d,d}(\boldsymbol{\mu}, \mathbf{0}, \bar{\boldsymbol{\omega}}, \Omega^*)$ with $\Omega^* = \begin{pmatrix} I_d & \Delta^T \\ \Delta & \bar{\Omega} \end{pmatrix}$. According to the discussions in section 4.1, for simplicity we assume $m = d, \boldsymbol{\gamma} = \mathbf{0}$ and take Γ to be the identity matrix. These assumptions ease the estimation procedure and make the optimization faster. In this case the moments are given by

$$\begin{aligned}
&\mathbb{E}\boldsymbol{y} = \boldsymbol{\mu} + \omega\Delta\mathbf{1}_d\sqrt{\frac{2}{\pi}} \\
&Var(\boldsymbol{y}) = \Omega - \frac{2}{\pi}\omega\Delta\Delta^T\omega \\
&\mathbb{E}(\Phi_d(\boldsymbol{y}, \mathbf{0}, I_d)) = 2^d\Phi_{2d}\left[\begin{pmatrix} \boldsymbol{\mu} \\ 0 \end{pmatrix}; \begin{pmatrix} I_d + \Omega & \omega\Delta^T \\ \omega\Delta & I_d \end{pmatrix}\right].
\end{aligned}$$

4.3 Numerical Results

In this section we perform numerical analysis for the theoretical results obtained in sections 4.2.1 and 4.2.2.

4.3.1 Simulation Study in the Univariate Case

The specifications of the simulation design are described as follows. For simplicity we take, $w = 1$, so $v = \sigma^2$ and we just need first three equations in (4.3). The sample sizes n are set at $50, 100$, and 500. We choose the values of δ to be $0.80, 0.89$ and 0.97. From figure 2.4.1, we notice that the skewness is not evident for the δ values less than 0.80, so the delta values only near one are considered. The true values for μ and σ are set to 0 and 1 respectively. The following table gives estimated values of the parameters of univariate SUN density. In parentheses are the mean squared error (MSE).

From the table 4.4.1 we see that our estimation method accurately estimates the parameter especially for the large sample sizes. Figure 4.3.1 shows the box plot of the estimates from 1000 samples of size 500 with the δ value 0.97. The Boxplot also illustrates that our estimation method works well especially for the large sample size.

Table 4.3.1: Estimated values (mean square error) of the parameters of univariate SUN density

Sample size	δ	$\hat{\mu}$	$\hat{\sigma}$	$\hat{\delta}$
n=50	0.80	0.1203(0.0942)	0.9331(0.0378)	0.7606(0.1028)
	0.89	0.1108(0.0579)	0.9318(0.0311)	0.8312(0.0512)
	0.97	0.0934(0.0401)	0.9201(0.0267)	0.8718(0.0394)
n=100	0.80	0.0938(0.0585)	0.9500(0.0241)	0.7977(0.0687)
	0.89	0.0605(0.0242)	0.9507(0.0147)	0.8906(0.0223)
	0.97	0.0632(0.0159)	0.9500(0.0145)	0.9175(0.0163)
n=500	0.80	0.0130(0.0088)	0.9903(0.0054)	0.8929(0.0194)
	0.89	0.0251(0.0040)	0.9834(0.0034)	0.9455(0.0073)
	0.97	0.0271(0.0032)	0.9796(0.0028)	0.9651(0.0026)

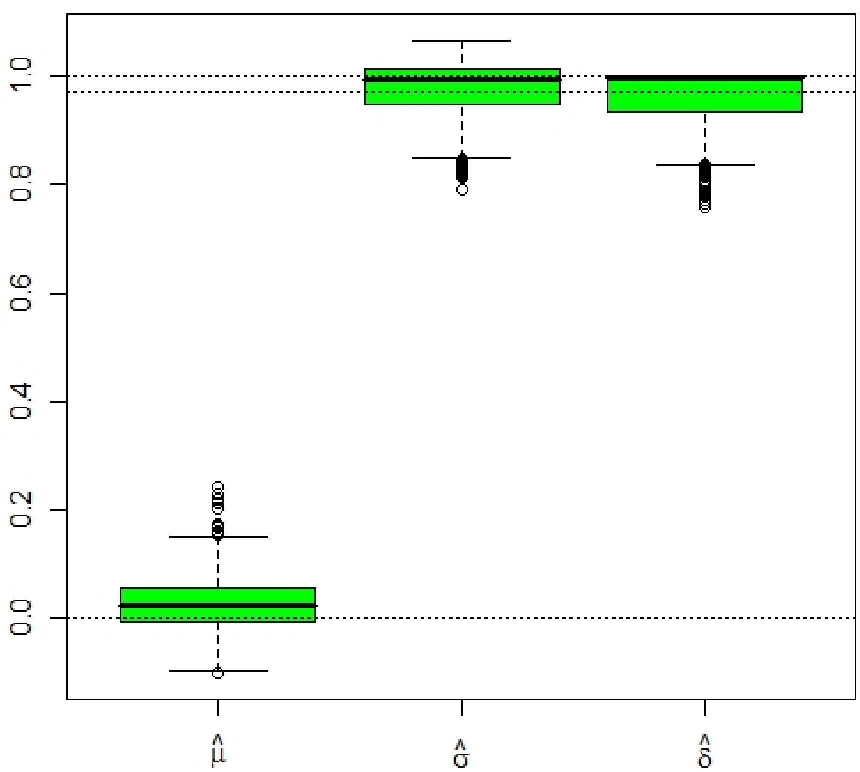

Figure 4.3.1: Boxplot of estimated values of μ, σ and δ obtained from 1000 replicates of size 500 with the true values 0,1,and 0.97 respectively. The dot lines represent the true values.

4.3.2 Simulation Study in the Bivariate Case

Following our univariate procedure, we assume $\omega = I_d$ and $\Delta = \delta\Omega^{\frac{1}{2}}$ for some $0 \leq \delta \leq 1$ for simplicity. Since $\omega = I_d$, we also have $\Omega = \bar{\Omega}$. Thus for WMOM approach we use the bivariate SUN density $SUN_{2,2}(\boldsymbol{\mu}, \mathbf{0}, \mathbf{1}_2, \Omega^*)$ with $\Omega^* = \begin{pmatrix} I_d & \delta\Omega^{\frac{1}{2}} \\ \delta\Omega^{\frac{1}{2}} & \Omega \end{pmatrix}$. In this setting, the parameters involved in the bivariate study are mean vector $\boldsymbol{\mu}$, covarince matrix Ω and shape parameter δ. The true values for the parameters $\boldsymbol{\mu}$ and Ω are taken as $\boldsymbol{\mu} = (0,0)^T, \Omega = \begin{pmatrix} 1 & 0.90 \\ 0.90 & 1 \end{pmatrix}$. Following the univariate case the δ values are selected to be 0.80, 0.89 and 0.97. The sample sizes n are set at 100 and 500. Table 4.3.2 gives estimated values of the parameters of bivariate SUN density. From the table we observe that estimated values and true values are almost same especially for large sample sizes. Thus our WMOM approach accurately estimates the parameter of bivariate SUN density. However, from the tabulated values we also see that there is some tendency of underestimation for δ when $\delta = 0.80$ and the sample size is 100. Figure 4.3.2 to 4.3.5 present a series of histograms of the parameters obtained from 1000 samples of varying sample sizes and varying delta values. In Figure 4.3.5 with $\delta = 0.97$ and the sample size 500, we observe a slight departure of some estimates from the true values. Overall, the histogram plots reveal that WMOM approach works well especially for the large sample sizes. In figure 4.3.6 we also present the boxplot of the estimates obtained from 1000 replicates with the size 500 and δ value 0.89. From the boxplot we do not observe too many outlying observations.

Table 4.3.2: Estimated values of parameters of bivariate SUN density

Sample size	δ	$\hat{\boldsymbol{\mu}}$	$\hat{\Omega}$	$\hat{\delta}$
n=100	0.80	$(0.0285, 0.0281)^T$	$\begin{bmatrix} 0.9968 & 0.8975 \\ 0.8975 & 0.9984 \end{bmatrix}$	0.7566
	0.89	$(0.0251, 0.0217)^T$	$\begin{bmatrix} 0.9892 & 0.8999 \\ 0.8999 & 0.9920 \end{bmatrix}$	0.8769
	0.97	$(0.0388, 0.0378)^T$	$\begin{bmatrix} 0.9760 & 0.8991 \\ 0.8991 & 0.9775 \end{bmatrix}$	0.9556
n=500	0.80	$(-0.0147, -0.0155)^T$	$\begin{bmatrix} 1.0064 & 0.8994 \\ 0.8994 & 1.0068 \end{bmatrix}$	0.8090
	0.89	$(0.0010, 0.0015)^T$	$\begin{bmatrix} 0.9989 & 0.8999 \\ 0.8999 & 0.9983 \end{bmatrix}$	0.8931
	0.97	$(0.0387, 0.0379)^T$	$\begin{bmatrix} 0.9778 & 0.9784 \\ 0.8996 & 0.8996 \end{bmatrix}$	0.9784

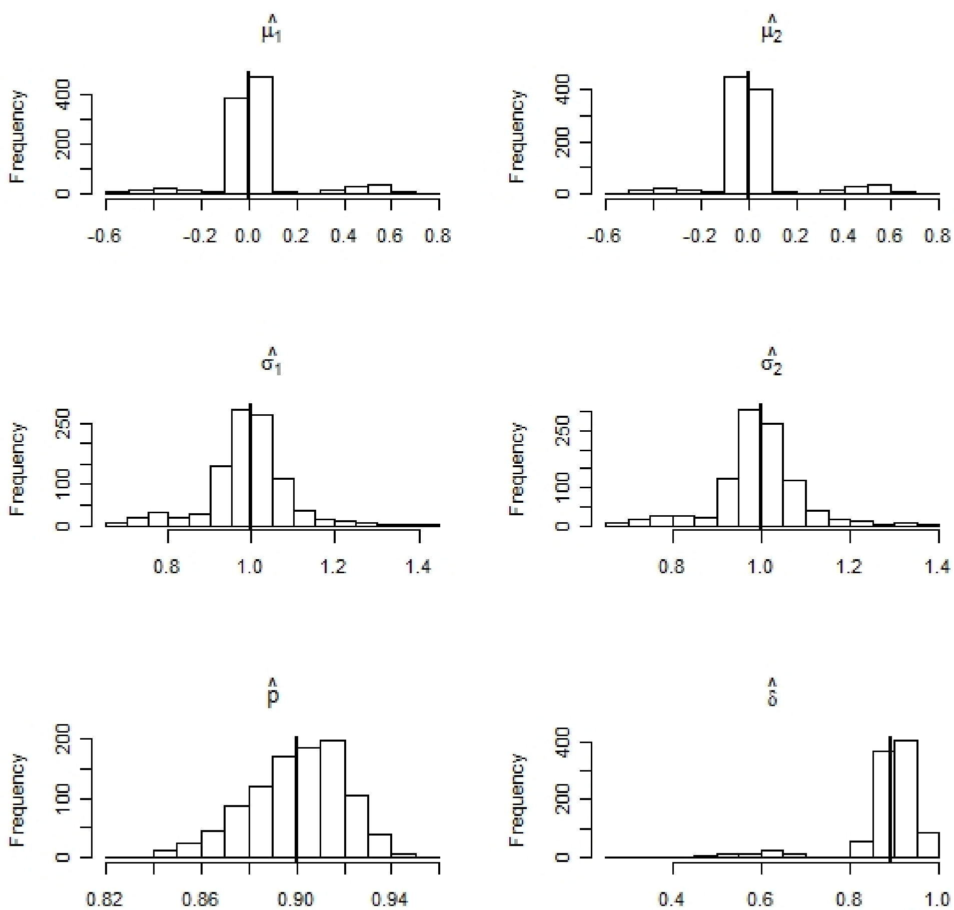

Figure 4.3.2: Histogram of estimated values of $\mu_1, \mu_2, \sigma_1, \sigma_2, \rho$ and δ obtained from 1000 replicates of size **100** with the true values 0,0,1,1, 0.90 and **0.89** respectively. The dot lines represent the true values.

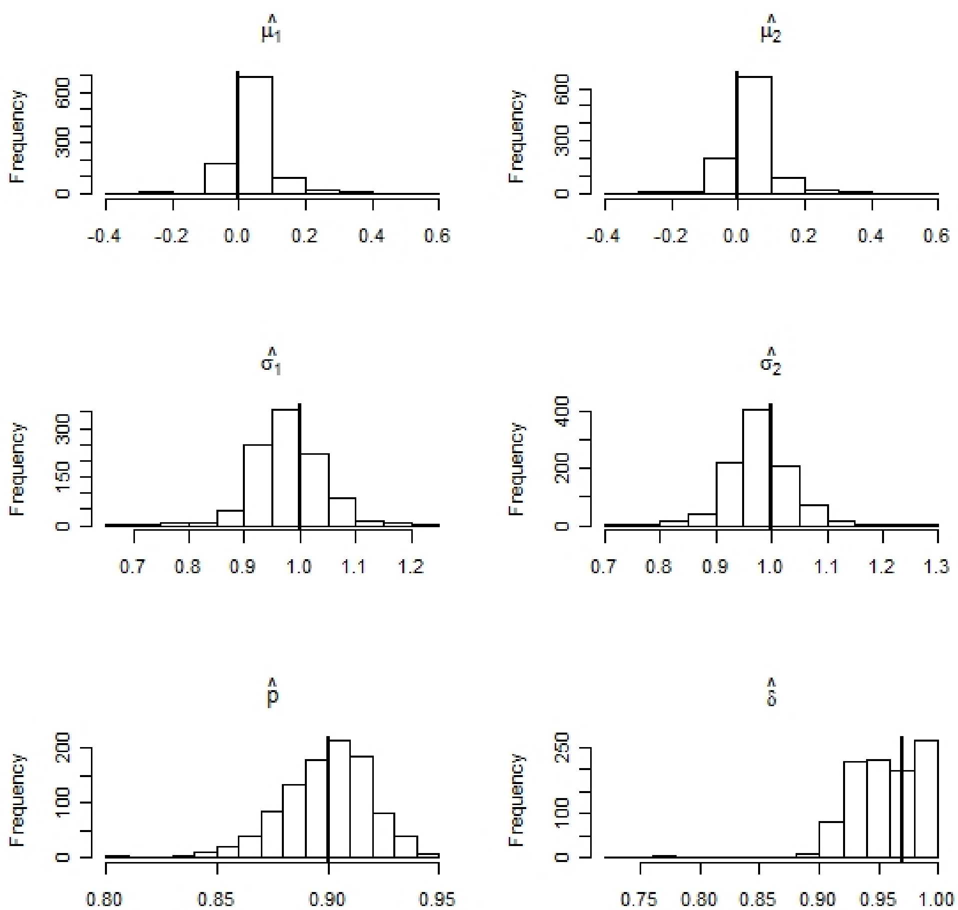

Figure 4.3.3: Histogram of estimated values of $\mu_1, \mu_2, \sigma_1, \sigma_2, \rho$ and δ obtained from 1000 replicates of size **100** with the true values 0,0,1,1, 0.90 and **0.97** respectively. The dot lines represent the true values.

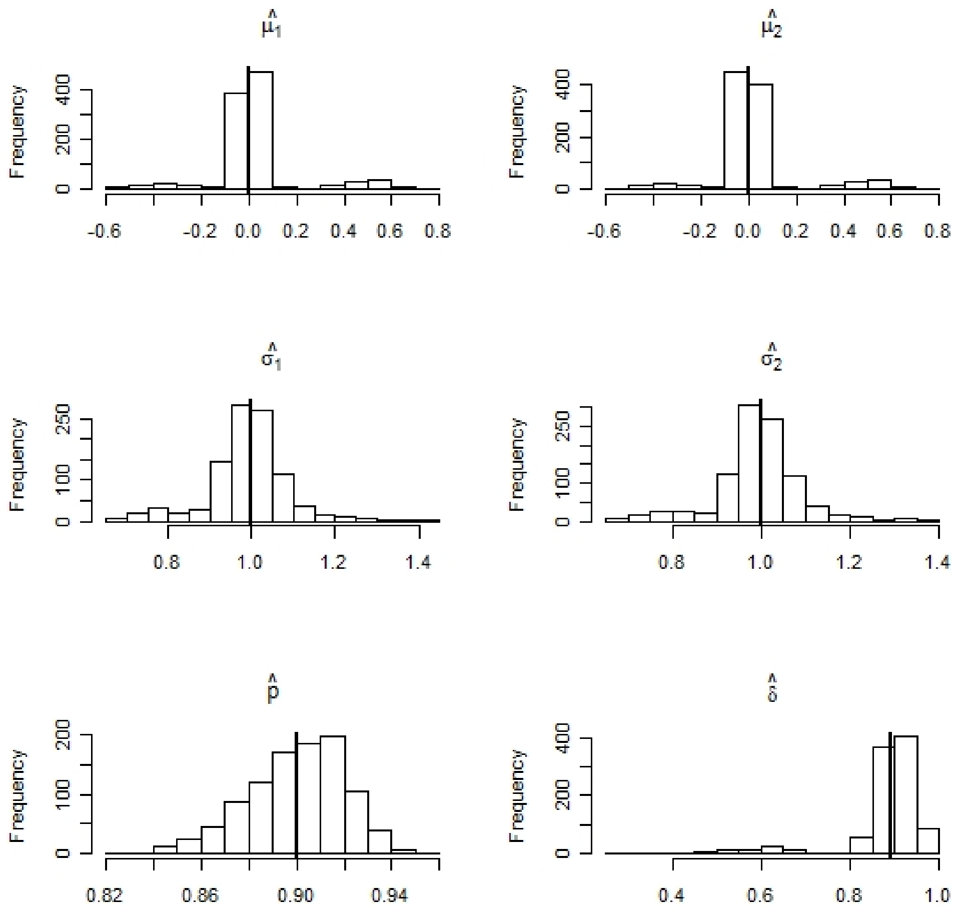

Figure 4.3.4: Histogram of estimated values of $\mu_1, \mu_2, \sigma_1, \sigma_2, \rho$ and δ obtained from 1000 replicates of size **500** with the true values 0,0,1,1, 0.90 and **0.89** respectively. The dot lines represent the true values.

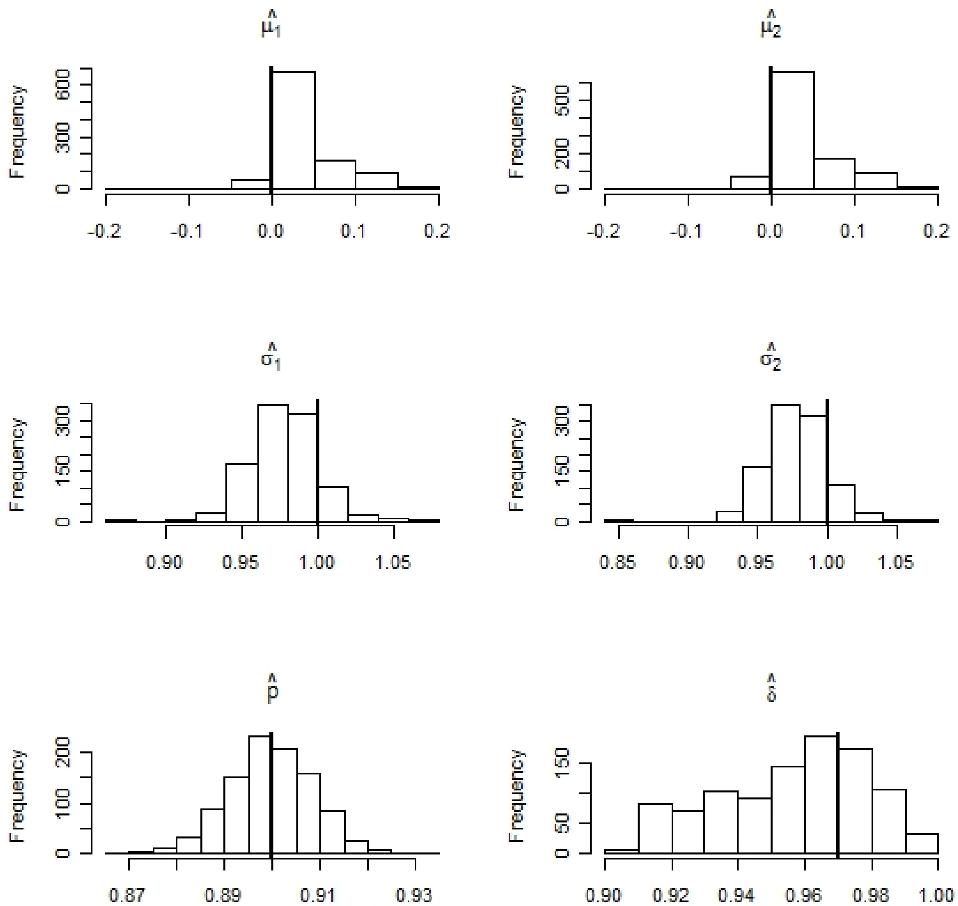

Figure 4.3.5: Histogram of estimated values of $\mu_1, \mu_2, \sigma_1, \sigma_2, \rho$ and δ obtained from 1000 replicates of size **500** with the true values 0,0,1,1, 0.90 and **0.97** respectively. The dot lines represent the true values.

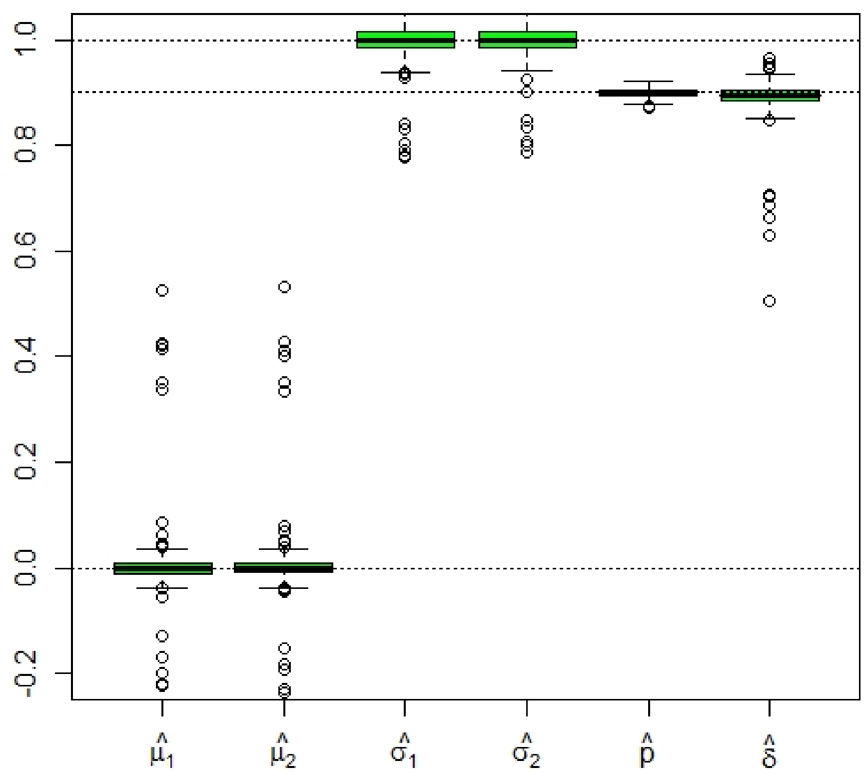

Figure 4.3.6: Box Plot of estimated values of $\mu_1, \mu_2, \sigma_1, \sigma_2, \rho$ and δ obtained from 1000 replicates of size **500** with the true values 0,0,1,1, 0.90 and **0.89** respectively. The dot lines represent the true values.

CHAPTER 5

LOWER CONVEX ORDER BOUND APPROXIMATIONS FOR SUMS OF LOG UNIFIED SKEW NORMAL RANDOM VARIABLES

5.1 Introduction

It is well known that in finance and actuarial science, the data usually have "fat tail" and in that case the normal distribution is not a good model to use. The skew normal distributions recently draw considerable attention as an alternative model. Unfortunately, the distribution of the sum of log-skew normal random variables does not have a closed form. In this work, we discuss the use of the lower convex order of random variables to approximate this distribution. Further, two applications of this approximate distribution are given : first is to describe the final wealth of a series of payments, and second is to describe the present value of a series of payments.

5.2 Basic Concepts and Definitions

Definition 5.2.1. *Consider two random variables X and Y such that $E[\phi(X)] \leq E[\phi(Y)]$, for all the convex functions ϕ, provided expectation exist.Then X is said to be smaller than Y in the convex order denoted as $X \leq_{cx} Y$.*

Definition 5.2.2. *(Convex order definition using stop-loss premium) Consider two random variables X and Y. Then X is said to precede Y in convex order sense if and only if*

$$E[X] = E[Y]$$

$$E[(X-d)_+] \leq E[(Y-d)_+], I_{(-\propto,\propto)}(d),$$

where

$$(X-d)_+ = max(X-d, 0).$$

An equivalent definition can be derived from the following relation

$$E[(X-d)_+] - E[(d-X)_+] = E(X) - d.$$

For the random variable Y the same relation is given by,

$$E[(Y-d)_+] - E[(d-Y)_+] = E(Y) - d.$$

Now assume $X \leq_{cx} Y$, which implies that

$$E[X] = E[Y],$$

and

$$E[(X-d)_+] \leq E[(Y-d)_+], I_{(-\propto,\propto)}(d).$$

Hence

$$E[(d-X)_+] \leq E[(d-Y)_+].$$

Therefore, a definition equivalent to the definition here is

$$E[X] = E[Y],$$

$$E[(d-X)_+] \leq E[(d-Y)_+].$$

Remark 5.2.1. *In economic term $\mathbb{E}[(X-d)_+]$ is the net premium for a stop-loss contract. It represents the expected loss over d, and $(X-d)_+$ is often called stop-loss premium. It is defined as follows:*

For a nonnegative loss X the payments equals

$$(X-d)_+ = max\{X-d, 0\} = \begin{cases} (X-d) & if\ X > d \\ 0 & if\ X \leq d, \end{cases}$$

The insurer retains a risk d (also called priority) and lets the reinsurer pay for the remainder. From the insurer point of view the loss stops at d and hence the name "stop-loss".

Properties of convex order of random variables

1. If X precedes Y in convex order sense i.e if $X \leq_{cx} Y$, then

   $E[X] = E[Y]$ and $Var[X] \leq Var[Y]$.

2. If $X \leq_{cx} Y$ and Z is independent of X and Y then $X + Z \leq_{cx} Y + Z$.

3. Let X and Y be two random varibales, then $X \leq_{cx} Y \Leftrightarrow -X \leq_{cx} -Y$.

4. Let X and Y be two random variables such that $E[X] = E[Y]$. Then $X \leq_{cx} Y$ if and only if $E|X-a| \leq_{cx} E|Y-a|, \forall$ a $\in \Re$.

5. The convex order is closed under mixtures:

Let X, Y and Θ be random variables such that $[X|\Theta = \theta] \leq_{cx} [Y|\Theta = \theta]$ $\forall$ θ in the support of Θ. Then $X \leq_{cx} Y$.

6. The convex order is closed under convolution:

   Let $X_1, X_2, ..., X_m$ be a set of independent random variables and $Y_1, Y_2, ..., Y_n$ be another set of independent random variables. If $X_i \leq_{cx} Y_i$, for $i = 1, ..., m$, then $\sum_{j=1}^{m} X_j \leq_{cx} \sum_{j=1}^{m} Y_j$.

7. Let X be a random variable with finite mean. Then $X + E[X] \leq_{cx} 2X$.

8. Let $X_1, X_2, ..., X_m$ and Y be (n+1) random variables. If $X_i \leq_{cx} Y$, $i = 1, ..., n$, then $\sum_{i=1}^{n} a_i X_i \leq_{cx} Y$, whenever $a_i \geq 0$,$i = 1, ..., n$ and $\sum_{i=1}^{n} a_i = 1$.

9. Let X and Y be independent random variables. Then $X_i \leq_{cx} Y_i$ if and only if

   $E[\phi(X, Y)] \leq E[\phi(Y, X)]$ $\forall$ $\phi \in \wp_{cx}$, where $\wp_{cx} = \{\phi : \Re^2 \longrightarrow \Re : \phi(X, Y) - \phi(Y, X)$ is convex for all x in y$\}$.

10. Let X_1 and X_2 be a pair of independent random variables and let Y_1 and Y_2 be another pair of independent random variables. If $X_i \leq_{cx} Y_i$, $i = 1, 2$ then $X_1 X_2 \leq_{cx} Y_1 Y_2$.

5.3 Risk Measures and Comonotonicity

5.3.1 Risk Measures

A risk measure provides the information contained in the distribution function of a random variable in one single real number. Risk measures are useful to evaluate and monitor the risk exposures of investors. One of the most commonly used risk measures in the field of actuarial science and financial economics is the p-quantile risk measure, based on a percentile concept. It is also called value-at-risk (VaR). Roughly speaking, VaR at level p, is the amount of capital required to ensure that the enterprise does not become technically insolvent. In probabilistic terms, the VaR at level p is defined as the $100p\%$ quantile of the distribution of the terminal wealth. More precisely, for any $p \in (0,1)$, the p-quantile measure or VaR for a random variable X, denoted by $Q_p[X]$, is defined as

$$Q_p[X] = inf\{x \in \Re | F_X(x) \geq p\} \tag{5.1}$$

It is a non-decreasing function and left continuous function of p. Other risk measures concerning the upper tail of the distributions are conditional tail expectation, tail-value-at-risk etc.

5.3.2 Comonotonicity

Comonotonicity is a well-studied and attractive property with diverse applications in the financial and actuarial field. It describes a very special dependence structure of a random vector in the sense that a random vector is comonotonic if all its components move in the same direction. To put it in a simple way, a random vector is said to be comonotonic if all its components are non-decreasing (or non-increasing) functions of the same random variable. One of the most important uses of comonotonic property is found in calculating the risk measures of sums of random variables. When the marginal risks posses the comonotonic dependence structure, the global value-at-risk can be obtained by summing up the marginal VaR measures (Roach and Valdez, 2008). Thus for a comonotonic random vector $\boldsymbol{X} = (X_1, X_2, \ldots, X_n)$ and the sum $W = \sum_{i=1}^{n} X_i$, the value at risk (VaR) is

$$Q_P(W) = \sum_{i=1}^{n} Q_P[X_i]$$

5.3.3 Main Result on Convex Order: Convex Lower Bound for Sums of Random Variables

For the evaluation of lower bound for the distribution of sum of log unified skew normal density we need the following result:

Lemma 5.3.1. *For any random vector* $\boldsymbol{X} = (X_1, X_2, \ldots, X_n)$ *and any radom variable* Λ*, which is assumed to be a function of* $\boldsymbol{X}$*, we have,*

$$\sum_{i=1}^{n} \mathbb{E}[X_i|\Lambda] \leq_{cx} \sum_{i=1}^{n} X_i. \tag{5.2}$$

Proof. From the convex order definition we have, $X \leq_{cx} Y$ if and only if $\mathbb{E}[\varphi(X)] \leq_{cx} \mathbb{E}[\varphi(y)]$. In accordance with this definition we need to show that

$$\mathbb{E}_\Lambda\Big[\varphi\Big(\sum_{i=1}^{n} \mathbb{E}[X_i|\Lambda]\Big)\Big] \leq \mathbb{E}\Big[\varphi(\sum_{i=1}^{n} X_i)\Big].$$

Now

$$\begin{aligned} \mathbb{E}\Big[\varphi(\sum_{i=1}^{n} X_i)\Big] &= \mathbb{E}_\Lambda\mathbb{E}\Big[\varphi(\sum_{i=1}^{n} X_i|\Lambda)\Big] \geq \mathbb{E}_\Lambda\Big[\varphi\Big(\mathbb{E}[\sum_{i=1}^{n} X_i|\Lambda]\Big)\Big] \\ &= \mathbb{E}_\Lambda\Big[\varphi\Big(\sum_{i=1}^{n} \mathbb{E}[X_i|\Lambda]\Big)\Big]. \end{aligned}$$

The last inequality was obtained by using Generalized Jenson inequality. Therefore,

$$\mathbb{E}_\Lambda\Big[\varphi\Big(\sum_{i=1}^{n} \mathbb{E}[X_i|\Lambda]\Big)\Big] \leq \mathbb{E}\Big[\varphi(\sum_{i=1}^{n} X_i)\Big].$$

□

5.3.4 Examples

*These two examples follow from Dhaene et al (2002)

Example 1: Convex lower bounds for the distribution of sum of independent normal random variables

Let X and Y be independent $N(0,1)$ random variables. We want to derive lower bounds for $S = X + Y$. In this case we know the exact distribution of S which is $N(0,2)$. Let us demonstrate how lower bound approximation works. Let $Z = X + aY$ for some real a . Then $Z \sim N(0, 1 + a^2)$. The conditional distribution of $S|Z$ is

$$N\Big[\boldsymbol{\mu}_S + \frac{\rho_{s,z}\sigma_s}{\sigma_z}(z - \boldsymbol{\mu}_z), \sigma_s^2(1 - \rho_{s,z}^2)\Big] = N\Big[z\frac{1+a}{1+a^2}, \frac{(1-a)^2}{1+a^2}\Big],$$

where $Cov(X + Y, X + aY) = Cov(X, X) + a.Cov(Y, Y) = 1 + a$ and $\rho_{s,z} = \frac{1+a}{\sqrt{2}\sqrt{1+a^2}}$. Then $\mathbb{E}(S|Z) = \frac{1+a}{1+a^2}Z$ is a random variable and has the distribution $N\Big[0, \frac{(1+a)^2}{1+a^2}\Big]$. Now for some choices of a we obtain the following distributions for the lower bounds of S:

$$\begin{aligned} a = 0 \;\; \text{gives} \;\; N(0,1) &\leq_{cx} S = X + Y \sim N(0,2) \\ a = 1 \;\; \text{gives} \;\; N(0,2) &\leq_{cx} S = X + Y \sim N(0,2) \\ a = -1 \;\; \text{gives} \;\; N(0,0) &\leq_{cx} S = X + Y \sim N(0,2). \end{aligned}$$

Thus in this case best lower bound is obtained when a = 1, which is same as exact distribution. The variance of the lower bound is seen to have a maximum at a = 1 and a minimum at a = -1.

Example 2: Convex lower bounds for the distribution of sums of independent log normal random variables

Suppose Y_1 and Y_2 are independent $N(0,1)$. Define $X_1 = e^{Y_1}$ which implies that $X_1 \sim lognormal(0,1)$, and $X_2 = e^{Y_1+Y_2}$ that is $X_2 \sim lognormal(0,2)$. We want to find the lower bound for the distribution of $S = X_1 + X_2$. Let $Z = Y_1 + Y_2$. As shown in Example (5.1) the conditional distribution of $Y_1|Z$ is given by,

$$Y_1|Y_1 + Y_2 = z \sim N\Big(\frac{1}{2}z, \frac{1}{2}\Big).$$

Therefore, $\mathbb{E}\Big[X_1 = e^{Y_1}|Y_1 + Y_2 = z\Big] = M_Y\Big(1; \frac{1}{2}z, \frac{1}{2}\Big)$ with $Y \sim N(\boldsymbol{\mu}, \sigma^2)$

$$= exp(\frac{1}{2}z + \frac{1}{4})$$

We also observe that $\mathbb{E}\Big[X_2 = e^{Y_1}|Y_1 + Y_2 = z\Big] = e^Z$. Therefore the lower bound for approximating the distribution of $S = X_1 + X_2$ is $S_l = \mathbb{E}\Big[X_1 + X_2|Z\Big] = exp(\frac{1}{2}z + \frac{1}{4})$. It can be easily verified that $\mathbb{E}(S_l) = \mathbb{E}\Big[exp(\frac{1}{2}z + \frac{1}{4})\Big] = e^{\frac{1}{2}} + e$ and $\mathbb{E}(S_l^2) = e^{\frac{3}{2}} + 2e^{\frac{5}{2}} + e^4$. Thus variance of the lower bound is 64.374 and is close to the variance of S =67.281. The idea is to obtain lower convex bound in such a way that the variance of the lower bound gets as close as possible to the variance of the sum. With this view in mind considering more general form of the conditioning variable as $Z = Y_1 + aY_2$, , it could be shown that optimal lower bound is reached for a = 1.27 and the variance in this case is 66.082. Thus the choice of the conditioning variable is crucial in determining the lower convex order bound.

5.4 Description of the Model

Let $\alpha_0, \alpha_1, \alpha_2, \ldots, \alpha_{n-1}$ be non-negative real numbers. Let $\boldsymbol{Y} = (Y_1, Y_2, \ldots, Y_n)^T$ be a multivariate skew normal random vector with the specified mean vector and variance-covariance matrix and satisfying additive properties. Define $Z_i = \sum_{k=i+1}^{n} Y_k, i = 0, 1, \ldots, n-1$, that is, Z_i's are linear combinations of the components $(Y_1, Y_2, \ldots, Y_n)$. With the components so defined, consider the sum

$$S \;=\; \sum_{i=0}^{n-1} \alpha_i e^{Z_i} = \sum_{i=0}^{n-1} \alpha_i e^{Y_{i+1}+\cdots+Y_n}. \tag{5.3}$$

From economic or actuarial point of view, the sum S could be interpreted as the final wealth or the terminal wealth or the accumulated value of a series of deterministic saving amounts or alternatively the accumulated value of a series of payments. In this situation, α_i $(i = 0, \ldots, n-1)$ represents yearly saving in period i or amount invested in period i, Y_{i+1} refers to the random rate of return in period i for i = 0,. . . ,n-1. The term $Y_k = log\frac{P_k}{P_{k-1}} = logP_k - logP_{k-1}$ i.e $e^{Y_k} = \frac{P_k}{P_{k-1}}$, where P_k is the price of the asset at the period k = 0,. . . ,n ; is called the random log-return in period k and Z_i denote the sum of stochastic or random returns in the period $i = 0, \ldots, n-1$. With some suitable adjustment, S could also be referred as the present value of a series of payments. More precisely, if $-Z_i$ denotes the stochastic log-return over the period [0, i], then e^{Z_i} represents the stochastic discount factor over the period [0, i]. In this situation, the sum S is the present value of α_i (Vanduffel et al 2008).

The sum defined in (5.3) plays a central role in the actuarial and financial theory because it allows computation of risk measures such as value at risk or stop-loss premium. To calculate the risk measures we need to evaluate the distribution function of S. Unfortunately, the distribution of the sum S (of log-normally or log-skew normally distributed random variables) is not available in the closed-form. It is possible to use Monte Carlo simulation method to approximate the distribution function. However, Monte Carlo simulation of the distribution is often time-consuming. Thus one has to find alternative way to approximate the distribution of the sum. Among the proposed solutions, moment matching methods and inverse gamma approximations are commonly used. Both methods approximate the unknown distribution function by a given one such that the first two moments coincide.

Kaas et al (2001) and Dhaene et al (2002a, 2002b) propose to approximate the distribution function of S by so called "convex lower bound". The underlying idea of convex lower order bound is to replace an unknown or too complex distribution (for which no explicit form is found) by another one which is easier to determine. In this approach, the real distribution is known to be bounded in terms of convex ordering to the approximated distribution. To be more precise, the distribution function of $S = \sum_{i=0}^{n-1} \alpha_i e^{Z_i}$ is approximated by the distribution function of S_l, where S_l is defined by,

$$S_l = \sum_{i=0}^{n-1} \alpha_i E(e^{Z_i}|\Lambda). \tag{5.4}$$

An appropriate choice of the conditioning random variable Λ is required. This approach has two-fold advantages. Firstly, use of this approach transforms the multidimensionality problem caused by $(Z_0, Z_2, \ldots, Z_{n-1})$ to a single dimension caused by Λ. Secondly, an appropriate choice of Λ (that makes the expectation in (5.4) non-decreasing or non-increasing function of the conditioning random variable Λ) will make a comonotonic sum, i.e, the elements of the sum in (5.4) posses the so called comonotonic dependence structure. Using additivity properties of sum of comonotonic random variables risk measures related to the distribution function of S is then approximated by the corresponding risk measures of S_l. According to Kaas et al. (2001), comonotonic upper bound for the sum in convex order sense can also be derived using the result

$$\sum_{i=0}^{n-1} X_i \leq_{cx} \sum_{i=0}^{n-1} F_{X_i}(U) \ ,$$

where U is the uniform random variable over $(0, 1)$. However, the comonotonic upper bounds generally provide too conservative estimates of the cumulative distribution function (Roach and Valdez 2008). Thus we only discuss convex lower bound here.

Remark 5.4.1. *In general, the random vector* $(E(X_0|\Lambda), E(X_1|\Lambda), \ldots, E(X_{n-1}|\Lambda))$ *does not have the same marginal distribution as* $(X_0, X_1, \ldots, X_{n-1})$*. However, if the conditioning random variable* Λ *is chosen in such a way that all random variables* $E(X_i|\Lambda)$*,* $(i = 0, 1, 2, \ldots, n-1)$ *are non-decreasing functions of* Λ *(or non-increasing functions of* Λ*), then the sum* $\sum_{i=0}^{n-1} E[X_i|\Lambda]$ *is a sum of n comonotonius random variables and can be referred to as comonotonic lower bound. Hence the risk measures for the sum could easily be obtained by summing the corresponding risk measures for the marginals involved.*

5.5 Bounds for the Sum of Log Unified Skew Normal Random Variables

In this section we derive the bounds to approximate the distribution of sums of log unified skew normal variables. The derivation of this bound requires some results that are presented in the following lemmas. Recall that Y_k denote the random log-return in the period k, for $k = 1, 2, \ldots, n$ and Z_i denote the accumulated returns from th time i to the final time $t = n$.

Lemma 5.5.1. *(Joint distribution of $\boldsymbol{Y} = (Y_1, \ldots, Y_n)^T$) Let $Y_k, k = 1, \ldots, n$ be univariate iid random variables distributed as*

$$SUN_{1,m}(\mu, \gamma, \bar{\omega}, \Omega^*), \text{ where } \ \Omega^* = \begin{pmatrix} \Gamma & \Delta^T \\ \Delta & \bar{\Omega} \end{pmatrix}.$$

Then the distribution of $\boldsymbol{Y} = (Y_1, \ldots, Y_n)^T$ is

$$SUN_{n,mn}(\boldsymbol{\mu}_Y, \boldsymbol{\gamma}_Y, \bar{\boldsymbol{\omega}}_Y, \Omega^*_Y),$$

where

$$\boldsymbol{\mu}_Y = \mathbf{1}_n \otimes \mu, \quad \boldsymbol{\gamma}_Y = \mathbf{1}_n \otimes \boldsymbol{\gamma}, \omega_Y = I_n \otimes \omega \quad \bar{\boldsymbol{\omega}}_Y = \omega_Y \otimes \mathbf{1}_n,$$

and

$$\Omega_Y = I_n \otimes \Omega, \quad \bar{\Omega}_Y = I_n \otimes \bar{\Omega}, \quad \Delta_Y = I_n \otimes \Delta, \quad \Gamma_Y = I_n \otimes \Gamma, \quad \Omega^*_Y = \begin{pmatrix} \Gamma_Y & \Delta_Y^T \\ \Delta_Y & \bar{\Omega}_Y \end{pmatrix}$$

Proof. The proof follows from Corollary 2.4.5 by noting that $\oplus_1^n = I_n \otimes A$ for any matrix A where I_n is an $n \times n$ identity matrix and $\mathbf{1}_n$ is a unit vector of dimension n. $\square$

Lemma 5.5.2. *(Joint distribution of $\boldsymbol{Z} = (Z_0, \ldots, Z_{n-1})^T$) Let $Z_i, i = 0, \ldots, n-1$ be the sum of returns of one unit of capital invested from time $t = i$ to the final time $t = n$, that* *is,* $Z_i = \sum_{k=i+1}^{n} Y_k$*. Let* $T \in \Re^{n\times n}$ *be an upper unit triangular matrix. Then the distribution of* $\boldsymbol{Z} = (Z_0, \ldots, Z_{n-1})^T$ *is*

$$SUN_{n,mn}(\boldsymbol{\mu}_Z, \boldsymbol{\gamma}_Z, \bar{\boldsymbol{\omega}}_Z, \Omega_Z^*)$$

where

$$\boldsymbol{\mu}_Z = T\boldsymbol{\mu}_Y, \quad \boldsymbol{\gamma}_Z = \boldsymbol{\gamma}_Y, \quad \omega_Z = T\omega_Y T^T, \quad \bar{\boldsymbol{\omega}}_Z = \omega_Z \mathbf{1}_n,$$

and

$$\Omega_Z = T\Omega_Y T^T, \;\; \bar{\Omega}_Z = \omega_Z^{-1}\Omega_Z\omega_Z^{-1}, \;\; \Delta_Z = (T\Omega_Y T^T)^{-1}T\omega_Y\Delta_Y, \;\; \Gamma_Z = \Gamma_Y, \;\; \Omega_Z^* = \begin{pmatrix} \Gamma_Z & \Delta_Z^T \\ \Delta_Z & \bar{\Omega}_Z \end{pmatrix}$$

Proof. The proof follows from Theorem 2.4.4 with T being the matrix of coefficients □

As mentioned in section 5.5, the comonotonicity of the convex lower bound strongly depends on the special choice of the conditioning random variable Λ. Therefore, it is required to choose a functional form of this random variable. Since a good choice of Λ is important in determining the accurate approximations for the final wealth, different choices of Λ have been proposed in the literature. Following Dhaene et al (2002a), we will choose Λ in such a way that it becomes a linear transformation of a first order approximation to S_n. This is known as "Taylor-based" approach. In this approach, Λ is defined as,

$$\Lambda = \sum_{i=0}^{n-1} \nu_i Z_i,$$

with the following choice of the coefficents ν_i,

$$\nu_i = \alpha_i e^{\mathbb{E}[Z_i]}.$$

If the random variables $Y_k, k = 1, \ldots, n$ are iid then the coefficients ν_i is given by

$$\nu_i = \alpha_i e^{\mathbb{E}[Z_i]} = \alpha_i e^{\mathbb{E}[Y]}.$$

Lemma 5.5.3. *(Distribution of Λ) Let the random variable Λ be defined by $\Lambda = \sum_{i=0}^{n-1} \nu_i Z_i$ and $\boldsymbol{V} = (\nu_0, \dots \nu_{n-1})$ be a row vector. Then the distribution of Λ is*

$$SUN_{1,mn}(\mu_\Lambda, \boldsymbol{\gamma}_\Lambda, \bar{\omega}_\Lambda, \Omega^*_\Lambda)$$

where

$$\mu_\Lambda = \boldsymbol{V}\boldsymbol{\mu}_Z, \quad \boldsymbol{\gamma}_\Lambda = \boldsymbol{\gamma}_Z, \quad \omega_\Lambda = \boldsymbol{V}\omega_Z\boldsymbol{V}^t, \quad \bar{\omega}_\Lambda = \omega_\Lambda 1,$$

and

$$\Omega_\Lambda = \boldsymbol{V}\Omega_Z\boldsymbol{V}^t, \;\; \bar{\Omega}_\Lambda = \omega_\Lambda^{-1}\Omega_\Lambda\omega_\Lambda^{-1}, \;\; \Delta_\Lambda = (\boldsymbol{V}\Omega_Z\boldsymbol{V}^t)^{-1}\boldsymbol{V}\omega_Z\Delta_Z, \;\; \Gamma_\Lambda = \Gamma_Z, \;\; \Omega^*_\Lambda = \begin{pmatrix} \Gamma_\Lambda & \Delta_\Lambda^T \\ \Delta_\Lambda & \bar{\Omega}_\Lambda \end{pmatrix}.$$

Proof. The proof follows from the Theorem 2.4.4 with $\boldsymbol{V}$ being the vector of coefficients. □

Lemma 5.5.4. *(Joint distribution of Λ and each of the elements of vector $\boldsymbol{Z}$) Let $S_i \in \Re^{2\times n}$ be a matrix with the first row as $\boldsymbol{V}$ and second row of 0's except in column $i+1$ where the 0 is replaced by 1. That is*

$$S_i = \begin{pmatrix} \nu_0 & \nu_1 & \dots & \nu_i \dots & \nu_{n-1} \\ 0 & 0 & \dots & 1 \dots & 0 \end{pmatrix}. \tag{5.5}$$

Then the distribution of $X_i = \big(\Lambda, Z_i\big)^T$ is

$$SUN_{2,mn}(\boldsymbol{\mu}_{X_i}, \boldsymbol{\gamma}_{X_i}, \bar{\omega}_{X_i}, \Omega^*_{X_i}),$$

where

$$\boldsymbol{\mu}_{X_i} = S_i\boldsymbol{\mu}_Z, \quad \boldsymbol{\gamma}_{X_i} = \boldsymbol{\gamma}_Z, \quad \omega_{X_i} = S_i\omega_Z S_i^T, \quad \bar{\boldsymbol{\omega}}_{X_i} = \omega_{X_i}\mathbf{1}_2,$$

$$\Omega_{X_i} = S_i\Omega_Z S_i^T, \quad \bar{\Omega}_{X_i} = \omega_{X_i}^{-1}\Omega_{X_i}\omega_{X_i}^{-1}, \quad \Delta_{X_i} = (S_i\omega_{X_i}S_i^T)^{-1}S_i\omega_{X_i}\Delta_Z, \quad \Gamma_{X_i} = \Gamma_Z,$$

and

$$\Omega^*_{X_i} = \begin{pmatrix} \Gamma_{X_i} & \Delta^T_{X_i} \\ \Delta_{X_i} & \bar{\Omega}_{X_i} \end{pmatrix}.$$

Proof. The proof follows from Theorem 2.4.4 with S_i being the matrix of coefficients. □

Lemma 5.5.5. *(Conditional distribution of $Z_i|\Lambda = \lambda$) Let $\boldsymbol{\mu_{X_i}}, \bar{\boldsymbol{\Omega}}_{X_i}, \bar{\omega}_{X_i}$ and Δ_{X_i} be partitioned as in Theorem 2.4.7. Then the distribution of $H_i = (Z_i|\Lambda = \lambda)$ is given by*

$$SUN_{1,mn}(\mu_{H_i}, \boldsymbol{\gamma}_{H_i}, \bar{\omega}_{H_i}, \Omega^*_{H_i})$$

where

$$\mu_{H_i} = \mu_2 + \Omega_{21}\Omega_{21}^{-1}(\lambda - \mu_1), \quad \boldsymbol{\gamma}_{H_i} = \boldsymbol{\gamma}_{X_i} + \Delta_1^T\bar{\Omega}_{11}^{-1}\omega_2^{-1}(\lambda - \mu_1), \;\; \bar{\omega}_{H_i} = \bar{\omega}_1,$$

$$\Gamma_{H_i} = \Gamma_{X_i} - \Delta_1^T\bar{\Omega}_{11}^{-1}\Delta_1, \Delta_{H_i} = \Delta_2 - \bar{\Omega}21\bar{\Omega}_{11}^{-1}\Delta1, \;\; \bar{\Omega}_{H_i} = \bar{\Omega}_{22} - \bar{\Omega}_{21}\bar{\Omega}_{11}^{-1}\bar{\Omega}_{12}, \;\; \Omega_{H_i} = \omega_{H_i}\bar{\Omega}_{H_i}\omega_{H_i},$$

and

$$\Omega^*_{H_i} = \begin{pmatrix} \Gamma_{H_i} & \Delta^T_{H_i} \\ \Delta_{H_i} & \bar{\Omega}_{H_i} \end{pmatrix}.$$

Theorem 5.5.1. *(The lower convex order bound) The lower convex order bound S_n^l which is used to approximate the distribution function of the sum $S_n = \sum_{i=0}^{n-1} \alpha_i e^{Z_i}$ is given by*

$$S_n^l = \sum_{i=0}^{n-1} \alpha_i exp\big(\mu_{H_i} + \frac{1}{2}\Omega_{H_i}\big)\frac{\Phi_{mn}\big(\boldsymbol{\gamma}_{H_i} + \Delta^T_{H_i}\omega_{H_i}; \Gamma_{H_i}\big)}{\Phi_{mn}(\boldsymbol{\gamma}_{H_i}; \Gamma_{H_i})} \tag{5.6}$$

with μ_{H_i}, Ω_{H_i}, $\boldsymbol{\gamma}_{H_i}$, Δ_{H_i}, ω_{H_i}, and Γ_{H_i} defined as in Lemma 5.5.5.

Proof. By Lemma 5.3.1 the distribution of $\sum_{i=0}^{n-1} \alpha_i e^{Z_i}$ is approximated by the distribution of the sum $\sum_{i=0}^{n-1} \mathbb{E}\big[\alpha_i e^{Z_i}|\Lambda = \sum_{i=0}^{n-1} \nu_i Z_i\big]$.

Thus the lower convex order bound S_n^l is given by

$$
\begin{aligned}
S_n^l &= \sum_{i=0}^{n-1} \mathbb{E}\Big[\alpha_i e^{Z_i} | \Lambda = \sum_{i=0}^{n-1} \nu_i Z_i\Big] \\
&= \sum_{i=0}^{n-1} \alpha_i \mathbb{E}\Big[e^{Z_i} | \Lambda = \sum_{i=0}^{n-1} \nu_i Z_i\Big]. \qquad (5.7)
\end{aligned}
$$

The expectation in (5.7) is the m.g.f of a random varible y evaluated at $t = 1$ where y is distributed as

$$SUN_{1,mn}(\boldsymbol{\mu}_{H_i}, \boldsymbol{\gamma}_{H_i}, \bar{\omega}_{H_i}, \Omega^*_{H_i}).$$

From (2.8), this m.g.f is obtained as

$$M_Y(1) = exp\big(\mu_{H_i} + \frac{1}{2}\Omega_{H_i}\big)\frac{\Phi_{nm}\big(\boldsymbol{\gamma}_{H_i} + \Delta_{H_i}^T \omega_{H_i}; \Gamma_{H_i}\big)}{\Phi_{mn}(\boldsymbol{\gamma}_{H_i}; \Gamma_{H_i})}.$$

Therefore from (5.7) the convex lower order bound is

$$S_n^l = \sum_{i=0}^{n-1} \alpha_i exp\big(\mu_{H_i} + \frac{1}{2}\Omega_{H_i}\big)\frac{\Phi_{mn}\big(\boldsymbol{\gamma}_{H_i} + \Delta_{H_i}^T \omega_{H_i}; \Gamma_{H_i}\big)}{\Phi_{mn}(\boldsymbol{\gamma}_{H_i}; \Gamma_{H_i})}.$$

$\square$

5.6 Lower Bound in Multi-Asset Case

In the previous section we consider only one asset while deriving the distribution of terminal wealth. In the same fashion it is also possible to find the lower bound when the portfolio consists of multiple assets including risk-free and risky assets. Throughout this section we will assume that the portfolio has one risk-free asset (e.g cash account) and multiple risky assets (e.g stock funds). Following the previous section we will derive the lower bound step by step. However, we will have to redefine some variables to accommodate in case of multiple assets.

Let Z_j^i be the sum of returns of one unit of capital invested at time $t = j$ to the final time $t = n$ of assest $i, i = 1, \ldots, q$, that is

$$Z_j^i = \sum_{k=j+1}^{n} Y_k^i,$$

and the terminal wealth $S_n(\pi)$ is given by

$$S_n(\boldsymbol{\pi}) = \sum_{i=1}^{q}\sum_{j=0}^{n-1} \pi_i \alpha_j \ exp(Z_j^i) + \sum_{j=0}^{n-1} \pi_0 \alpha_j \ exp((n-j)r),$$

where $\boldsymbol{\pi} = (\pi_1, \ldots, \pi_q)^T$ is the vector of proportions of savings amounts in the risky assets and π_0 is the weight in the risk-free asset.

Lemma 5.6.1. *(Joint distribution of $\boldsymbol{Y} = (\boldsymbol{Y}_1^T, \ldots, \boldsymbol{Y}_n^T)^T$) Let the joint returns random vector $\boldsymbol{Y}_k, k = 1, \ldots, n$ be iid distributed as*

$$SUN_{q,m}(\boldsymbol{\mu}, \boldsymbol{\gamma}, \bar{\boldsymbol{\omega}}, \Omega^*), \textit{where} \ \ \Omega^* = \begin{pmatrix} \Gamma & \Delta^T \\ \Delta & \bar{\Omega} \end{pmatrix}.$$

Then the vector of log returns $\boldsymbol{Y} = (\boldsymbol{Y}_1^T, \ldots, \boldsymbol{Y}_n^T)^T$ is distributed as

$$SUN_{nq,mn}(\boldsymbol{\mu}_Y, \boldsymbol{\gamma}_Y, \bar{\boldsymbol{\omega}}_Y, \Omega_Y^*),$$

where

$$\boldsymbol{\mu}_Y = \mathbf{1}_n \otimes \boldsymbol{\mu}, \quad \boldsymbol{\gamma}_Y = \mathbf{1}_n \otimes \boldsymbol{\gamma}, \quad \omega_Y = I_n \otimes \omega, \quad \bar{\boldsymbol{\omega}}_Y = \omega_Y \otimes \mathbf{1}_n$$

and

$$\Omega_Y = I_n \otimes \Omega, \quad \bar{\Omega}_Y = I_n \otimes \bar{\Omega}, \quad \Delta_Y = I_n \otimes \Delta, \quad \Gamma_Y = I_n \otimes \Gamma, \quad \Omega_Y^* = \begin{pmatrix} \Gamma_Y & \Delta_Y^T \\ \Delta_Y & \bar{\Omega}_Y \end{pmatrix}.$$

Proof. The proof follows from Corollary 2.4.5 $\square$

Lemma 5.6.2. *(Distribution of $\boldsymbol{Z} = (\boldsymbol{Z}_1^T, \ldots, \boldsymbol{Z}_q^T)^T$) Let $\boldsymbol{Z} = (\boldsymbol{Z}_1^T, \ldots, \boldsymbol{Z}_q^T)^T$ be the vector of accumulated returns, where $Z_i = (Z_0^i, \ldots, Z_{n-1}^i), i = 0, \ldots, q$. Let $\boldsymbol{T}_j^i$ be an $m.n$ dimensional row vector of $0's$ except in the $(i + q(j+k))th$ positions, $k = 0, 1, \ldots, n-(j+1)$ where they are $1's$ and let T be a matrix whose rows are defined by vectors $\boldsymbol{T}_j^i$. Then the distribution of $\boldsymbol{Z}$ is*

$$SUN_{nq,mn}(\boldsymbol{\mu}_Z, \boldsymbol{\gamma}_Z, \bar{\boldsymbol{\omega}}_Z, \Omega_Z^*),$$

where

$$\boldsymbol{\mu}_Z = T\boldsymbol{\mu}_Y, \quad \boldsymbol{\gamma}_Z = \boldsymbol{\gamma}_Y, \omega_Z = T\omega_Y T^T, \quad \bar{\boldsymbol{\omega}}_Z = \omega_Z \mathbf{1}_n,$$

and

$$\Omega_Z = T\Omega_Y T^T, \ \ \bar{\Omega}_Z = \omega_Z^{-1}\Omega_Z\omega_Z^{-1}, \ \ \Delta_Z = (T\Omega_Y T^T)^{-1} T\omega_Y\Delta_Y, \ \ \Gamma_Z = \Gamma_Y, \ \ \Omega_Z^* = \begin{pmatrix} \Gamma_Z & \Delta_Z^T \\ \Delta_Z & \bar{\Omega}_Z \end{pmatrix}.$$

Proof. The proof follows from Theorem (2.4.4) with T being the matrix of coefficients □

As in the single asset case, we use "Taylor-based" approach for choosing the random variable Λ. The random variable is accommodated to the multi asset case in the following way:

$$\Lambda(\pi) = \sum_{i=0}^{q}\sum_{j=0}^{n-1} \nu_j^i(\pi) Z_j^i$$

with the following choice of the coefficents ν_j^i,

$$\nu_j^i(\pi) = \pi_i \alpha_j e^{\mathbb{E}[Z_j^i]}.$$

If the random variables $Y_k, k = 1, \ldots, n$ are iid then the coefficients ν_j^i is given by

$$\nu_j^i(\pi) = \pi_i \alpha_j e^{\mathbb{E}[Z_j^i]} = \pi_i \alpha_j e^{(n-j)\mathbb{E}[Y^i]},$$

where $\mathbb{E}[Y^i]$ denotes the expectation of the ith marginal distribution of the random vector

$\boldsymbol{Y_j}$.

Lemma 5.6.3. *(Distribution of Λ) Let the random variable $\Lambda(\pi)$ be defined by $\Lambda = \sum_{i=0}^{q}\sum_{j=0}^{n-1}\nu_j^i(\pi)Z_j^i$ and let $\boldsymbol{V} = (V_1,\dots V_q)$, where $V_i = (\nu_0^i(\pi),\dots \nu_{n-1}^i(\pi)), i = 1,\dots,q$. Then the distribution of $\Lambda(\pi)$ is*

$$SUN_{1,mn}(\mu_\Lambda, \boldsymbol{\gamma}_\Lambda, \bar{\omega}_\Lambda, \Omega_\Lambda^*),$$

where

$$\mu_\Lambda = \boldsymbol{V}\boldsymbol{\mu}_Z, \quad \gamma_\Lambda = \boldsymbol{\gamma}_Z, \quad \omega_\Lambda = \boldsymbol{V}\omega_Z\boldsymbol{V}^t \quad \bar{\omega}_Z = \omega_\Lambda 1,$$

and

$$\Omega_\Lambda = \boldsymbol{V}\Omega_Z\boldsymbol{V}^T, \;\; \bar{\Omega}_\Lambda = \omega_\Lambda^{-1}\Omega_\Lambda\omega_\Lambda^{-1}, \;\; \Delta_\Lambda = (\boldsymbol{V}\Omega_Z\boldsymbol{V}^T)^{-1}\boldsymbol{V}\omega_Z\Delta_Z, \;\; \Gamma_\Lambda = \Gamma_Z, \;\; \Omega_\Lambda^* = \begin{pmatrix} \Gamma_\Lambda & \Delta_\Lambda^T \\ \Delta_\Lambda & \bar{\Omega}_\Lambda \end{pmatrix}.$$

Proof. The proof follows from the Theorem 2.4.4 with $\boldsymbol{V}$ being the vector of coefficients. □

Lemma 5.6.4. *(Joint distribution of $\Lambda(\pi)$ and $\boldsymbol{Z_j^i}$) Let $S_j^i \in \Re^{2\times q\cdot n}$ be a matrix with the first row as $\boldsymbol{V}$ and the second row of 0's except in column $i\cdot n-(n-j-1), i = 1,\dots,q, j = 0,\dots,n-1$ where the 0 is replaced by 1. Then the distribution of $\boldsymbol{X}_j^i = \left(\Lambda(\pi), Z_j^i\right)^T$ is*

$$SUN_{2,mn}(\boldsymbol{\mu}_{X_j^i}, \boldsymbol{\gamma}_{X_j^i}, \bar{\boldsymbol{\omega}}_{X_j^i}, \Omega_{X_j^i}^*),$$

where

$$\boldsymbol{\mu}_{X_j^i} = S_j^i\boldsymbol{\mu}_Z, \quad \boldsymbol{\gamma}_{X_j^i} = \boldsymbol{\gamma}_Z, \quad \omega_{X_j^i} = S_j^i\omega_Z S_{j^i}^T \quad \bar{\boldsymbol{\omega}}_{X_j^i} = \omega_{X_j^i}1_2,$$

$$\Omega_{X_j^i} = S_j^i\Omega_Z S_{j^i}^T, \quad \bar{\Omega}_{X_i} = \omega_{X_j^i}^{-1}\Omega_{X_j^i}\omega_{X_j^i}^{-1}, \quad \Delta_{X_j^i} = (S_j^i\omega_{X_j^i}S_{j^i}^T)^{-1}S_j^i\omega_{X_j^i}\Delta_Z, \quad \Gamma_{X_j^i} = \Gamma_Z$$

and

$$\Omega_{X_j^i}^* = \begin{pmatrix} \Gamma_{X_j^i} & \Delta_{X_j^i}^T \\ \Delta_{X_j^i} & \bar{\Omega}_{X_j^i} \end{pmatrix}.$$

Proof. The proof follows from Theorem 2.4.4 with S_j^i being the matrix of coefficients. $\square$

Lemma 5.6.5. *(Conditional distribution of Z_j^i given $\Lambda(\pi)$) Let $\boldsymbol{\mu}_{X_j^i}, \bar{\boldsymbol{\Omega}}_{X_j^i}, \bar{\omega}_{X_j^i}$ and $\Delta_{X_j^i}$ be partitioned as in theorem 2.4.7. Then the conditional distribution of $H_j^i = (Z_j^i | \Lambda = \lambda)$ is given by*

$$SUN_{1,mn}(\mu_{H_j^i}, \boldsymbol{\gamma}_{H_j^i}, \bar{\omega}_{H_j^i}, \Omega^*_{H_j^i}),$$

where

$$\mu_{H_j^i} = \mu_2 + \Omega_{21}\Omega_{21}^{-1}(\lambda - \mu_1), \quad \boldsymbol{\gamma}_{H_j^i} = \boldsymbol{\gamma}_{X_j^i} + \Delta_1^T \bar{\Omega}_{11}^{-1}\omega_2^{-1}(\lambda - \mu_1), \ \bar{\omega}_{H_j^i} = \bar{\omega}_1,$$

$$\Gamma_{H_j^i} = \Gamma_{X_j^i} - \Delta_1^T \bar{\Omega}_{11}^{-1}\Delta_1, \ \ \Delta_{H_j^i} = \Delta_2 - \bar{\Omega}21\bar{\Omega}_{11}^{-1}\Delta 1, \ \ \bar{\Omega}_{H_j^i} = \bar{\Omega}_{22} - \bar{\Omega}_{21}\bar{\Omega}_{11}^{-1}\bar{\Omega}_{12}, \ \ \Omega_{H_j^i} = \omega_{H_j^i}\bar{\Omega}_{H_j^i}\omega_{H_j^i},$$

and

$$\Omega^*_{H_j^i} = \begin{pmatrix} \Gamma_{H_j^i} & \Delta_{H_j^i}^T \\ \Delta_{H_j^i} & \bar{\Omega}_{H_j^i} \end{pmatrix}.$$

Theorem 5.6.1. *(The lower convex order bound)*

The lower convex order bound $S_n^l(\pi)$ to approximate the distribution function of the sum $S_n(\boldsymbol{\pi}) = \sum_{i=1}^q \sum_{j=0}^{n-1} \pi_i \alpha_j \ exp(Z_j^i) + \sum_{j=0}^{n-1} \pi_0 \alpha_j \ exp((n-j)r)$ is given by

$$S_n^l(\pi) = \sum_{i=1}^q \sum_{j=0}^{n-1} \pi_i \alpha_j exp\big(\mu_{H_j^i} + \frac{1}{2}\Omega_{H_j^i}\big) \frac{\Phi_{mn}\big(\boldsymbol{\gamma}_{H_j^i} + \Delta_{H_j^i}^T \omega_{H_j^i}; \Gamma_{H_j^i}\big)}{\Phi_{mn}(\boldsymbol{\gamma}_{H_j^i}; \Gamma_{H_j^i})} + \sum_{j=0}^{n-1} \pi_0 \alpha_j \ exp((n-j)r)$$

with $\mu_{H_j^i}$, $\Omega_{H_j^i}$, $\boldsymbol{\gamma}_{H_j^i}$, $\Delta_{H_j^i}$, $\omega_{H_j^i}$, and $\Gamma_{H_j^i}$ defined as in lemma 5.7.5.

Proof. The proof is same as the one given in Theorem 5.6.1. $\square$

5.7 Numerical Results

In this section we numerically illustrate the accuracy of the approximations obtained in the previous two sections by using two examples. We do not use a real data set to analyze the approximations since making inference on the unified skew normal distribution is difficult due to its large number of parameters, instead we use a hypothetical data to evaluate the accuracy.

In the first example, the final wealth of yearly savings distributed as log unified skew normal random variables is computed using Monte Carlo method and the result presented in section 5.6.

For $n = 20$, at every period $i, i = 0, \ldots, n-1$, consider the yearly savings amounts $\alpha_i = 1, i = 0, \ldots, 19$. That is, at the beginning of each year one unit of savings amount is invested in the considered asset. At time $i = n$ the invested amount $\alpha_n = 0$, i.e, no contribution is made at the final period. The returns are considered to be independently and indentically distributed SUN random variables with parameters $m = d = 1,\ \mu = 0.02,\ \gamma = 0,\ \omega = 1,\ \Omega = \bar{\Omega} = 0.03,\ \Gamma = 1$, and $\Delta = 0.97$.

The results for some selected quantiles of the distribution function of the terminal wealth obtained by the Monte Carlo simulation (denoted by MCB) and from the convex lower bound (CLB) are presented in table 5.8.1. The simulated results are obtained from 5000 random paths. The relative deviations of the approximated values from the Monte Carlo simulation are computed as follows:

$$\frac{Q_p[S_n^l] - Q_p[S_n^{MC}]}{Q_p[S_n^{MC}]} \times 100.$$

Table 5.7.1: Comparison of the selected quantiles of the distribution of the final wealth in single asset case

p	MCB	CLB	*Relative deviation*
0.01	22.5789	22.8575	1.23%
0.025	25.9758	26.1992	0.86%
0.05	29.7387	29.9730	0.78%
0.95	152.9541	151.5890	-0.89%
0.975	184.0342	182.4662	-0.85%
0.99	226.9628	222.1930	-2.10%

Comparing the results obtained with the Monte Carlo simulations, all the lower bound approximations seem to perform reasonably well, some of them are excellent. The approximations lose some precision in the tails of the distribution.

In the second example, we illustrate the approximations for a portfolio consisting of two risky assets and one risk free asset. We consider the same savings amount as in the first example (that is, $\alpha_i = 1, i = 1, \ldots, n$) and the weights are assigned as follows: 19% in the risk-free asset, 45% in the first risky asset and the remaining 36% will be invested in the second risky asset. In addition, the yearly return of the risk-free asset is considered to be 0.03. The parameters of the joint distribution of the risky assets are chosen to be $m = 1,\ d = 2,\ \mu = (0.06, 0.1)^T,\ \gamma = 0,\ \omega = I_2,\ \Omega = \bar{\Omega} = \begin{pmatrix} 0.01 & 0.01 \\ 0.01 & 0.04 \end{pmatrix},\ \Gamma = 1$, and $\Delta = (-0.95, -0.97)^T$.

The results for the distribution function of the terminal wealth obtained by the Monte Carlo simulation and from the convex lower bound are presented in table 5.8.2. Following single asset case, the simulated results are obtained from 5000 random paths.

Table 5.7.2: Comparison of the selected quantiles of the distribution of the final wealth in multi asset case

p	MCB	CLB	*Relative deviation*
0.01	19.4143	19.7961	1.97%
0.025	22.0037	22.3248	1.46%
0.05	24.5388	24.7480	0.85%
0.95	124.8726	124.3719	-0.40%
0.975	152.1901	150.5944	-1.05%
0.99	196.4655	191.0344	-2.76%

As observed from Table 5.8.2, the approximation is still reasonably good when we consider a portfolio. The approximations at the tails of the distribution lose more precisions compared to the single asset situation. One of the reasons might be that in multi asset case we include an extra risky asset thus making the number of log unified skew normal random variables double compared to the single asset case. However, the approximations will surely improve with a better choice of the conditioning random variable Λ.

BIBLIOGRAPHY

[1] Arellano-Valle, R. B. and Azzalini, A. (2006). On the unification of families of skew-normal distributions. *Scand. J. Statist.*, **33**, 561-574.

[2] Arellano-Valle, R. B. and Genton, M. G. (2005). On fundamental skew distributions. *J. Multivariate Anal.*, **96**, 93-116.

[3] Azzalini, A. (1985). A class of distributions which includes the normal ones. *Scand. J. Statist.*, **12**, 171-178.

[4] Azzalini, A. and Capitanio, A. (1999). Statistical applications of the multivariate skew normal distributions. *J. R. Stat. Soc., Ser. B*, **61**, 579-602.

[5] Azzalini, A. and Dalla Valle, A. (1996). The multivariate skew-normal distribution. *Biometrika*, **83**, 715-726.

[6] Cochran, W. G. (1934). The distributions of quadratic forms in a normal systems, with appplications to the analysis of variance. *Proc. Camb. Phil. Soc.*, **30**, 178-191.

[7] Craig, A. T. (1943). Note on the independence of certain quadratic forms. *The Annals of Mathematical Statistics*, **14**, 195-197.

[8] Dhaene, J., Denuit, M., Gooverts, M., Kass, R., and Vyncke, D. (2002a). The Concept of Comonotonicity in Actuarial Science and Finance: Theory. *Insurance: Mathematics and Economics*, **31**, 3-33.

[9] Dhaene, J., Denuit, M., Gooverts, M., Kass, R., and Vyncke, D. (2002b). The Concept of Comonotonicity in Actuarial Science and Finance: Applications. *Insurance: Mathematics and Economics*, **31**, 133-161.

[10] Domingnez-Molina, J. A., Gonzalez-Farias, G., and and Gupta, A. K. (2001). General multivariate skew normal distribution. Department of Mathematics and Statistics, Bowling Green State University, Tecnical Report No.01-09.

[11] Flecher, C., Naveau, P. and Allard, D. (2009). Estimating the closed skew-normal distribution parameters using weighted moments. *Statistics and Probability Letters*, **79(19)**, 1977-1984.

[12] Genton, M. G., editor (2004a). *Skew-elliptical distributions and their applications: a journey beyond normality.* Chapman and Hall/CRC.

[13] Genton, M. G., He, L., Liu, X. (2001). Moments of skew-normal random vectors and their quadratic forms. *Statistics and Probability Letters*, **51**, 319-325.

[14] Gonzalez-Farias, G., and Dominguez-Molina, J. A. and Gupta, A. K. (2004b). The closed skew-normal distribution. In Genton, M. G., editor, *Skew-elliptical distributions and their applications: a journey beyond normality*, chapter 2, pages 2542. Chapman and Hall/CRC.

[15] Gupta, A. K. and Chen, J. T. (2004). A class of multivariate skew-normal models. *Annals of the Institute of Statistical Mathematics*, **56(2)**, 305-315.

[16] Gupta, A. K., Gonzalez-Farias, G., and Dominguez-Molina, J. A. (2004a). A multivariate skew normal distribution. *J. Multivariate Anal*, **89**, 181-190.

[17] Gupta, A. K. and Huang, W.J. (2002). Quadratic forms in skew normal variates. *J. Math. Anal. Appl.*, **273**, 558-564.

[18] Gupta, A. K. and Kabe, D. G. (1998). Moments of ratios of quadratic forms. *Statistics and Probability Letters*, **38(1)**, 69-71.

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

[20] Horn, R. A. and Johnson, C. R. (1991). *Topics in Matrix Analysis.* Cambridge University Press.

[21] Huang, W. J. and Chen, Y. H. (2006). Quadratic forms of multivariate skew normal symmetric distributions. *Statist. Probab. Lett.* **76**, 871-879.

[22] Kaas, R., Goovaerts , M.J., Dhaene, J. and Denuit, M. (2001). *Modern Actuarial Risk Theory.* Kluwer Academic Publishers, Dordrecht.

[23] Li, G. (1987). Moments of random vector and its quadratic forms. *J. Statist Appl. Probab*, **2**, 219-229.

[24] Liseo, B. and Loperfido, N. (2003). A Bayesian interpretation of the multivariate skew-normal distribution. *Statist. Probab. Lett.*, **61**, 395-401.

[25] Loperfido, N. (2001). Quadratic forms of skew-normal random vectors. *Statist. Probab. Lett.*, **54**, 381-387.

[26] Magnus, J. R. (1986). The exact moments of a ratio of quadratic forms in normal random variables. *Ann. Econom. Statist.*, **4**, 95-109.

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

[28] Roberts, C. (1966). A correlation model useful in the study of twins. *Journal of the American Statistical Association*, **61**, 1184-1190.

[29] Roch, O. and Valdez, E. (2010). Lower convex order bound approximations for sums of log-skew normal random variables. *Applied Stochastic Models in Business and Industry* (In press)

[30] Sahu, K., Dey, D. K., and Branco, M. D. (2003). A new class of multivariate skew distributions with applications to Bayesian regression models. *Canad. J. Statist.*, **31**, 129-150.

[31] Sawa, T. (1978). The exact moments of the least squares estimator for the autoregressive model. *Journal of Econometrics*, **8**, 159-172.

[32] Vanduffel, S., Chen, X.; Dhaene, J.; Goovaerts, M.; Henrard, L.; and Kass, R. (2008). Optimal approximations for risk measures of sums of lognormals based on conditional expectations. *Journal of Computational and Applied Mathematics*, **221(1)**, 202-218.

[33] Wang, T., Li, B., Gupta, A. K. (2009). Distribution of quadratic forms under skew normal settings. *Journal of Multivariate Analysis*, **100**, 533-545