A CLASS OF MULTIVARIATE SKEW DISTRIBUTIONS: PROPERTIES
AND INFERENTIAL ISSUES

Deniz Akdemir

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

May 2009

Committee:
Arjun K. Gupta, Advisor
James M. McFillen,
Graduate Faculty Representative
John T. Chen
Hanfeng Chen
Flexible parametric distribution models that can represent both skewed and symmetric distributions, namely skew symmetric distributions, can be constructed by skewing symmetric kernel densities by using weighting distributions. In this dissertation, we study a multivariate skew family that have either centrally symmetric or spherically symmetric kernel. Specifically, we define multivariate skew symmetric forms of uniform, normal, Laplace, and Logistic distributions by using the cdf’s of the same distributions as weighting distributions. Matrix variate extensions of these distributions are also introduced herein. To bypass the unbounded likelihood problem related to the inference about this model, we propose an estimation procedure based on the maximum product of spacings method. This idea also leads to bounded model selection criteria that can be considered as alternatives to Akaike’s and other likelihood based criteria when the unbounded likelihood may be a problem. Applications of skew symmetric distributions to data are also considered.

**AMS 2000 subject classification:** Primary 60E05, Secondary 62E10

**Key words and phrases:** Multivariate Skew-Symmetric Distribution, Matrix Variate Skew-Symmetric Distribution, Inference, Maximum Product of Spacings, Unbounded Likelihood, Model Selection Criterion
To little Pelin
ACKNOWLEDGMENTS

This is a great opportunity to express my respect and thanks to Prof. Arjun K. Gupta for his continuous support and invaluable input for this work. It was a pleasure to be his student. I am also grateful to my wife, my parents and my children.
# Table of Contents

## CHAPTER 1: Notation, Introduction and Preliminaries

1.1 Notation .................................................. 1  
1.2 Preliminaries ........................................... 3  
   1.2.1 Vectors and Matrices ................................. 3  
   1.2.2 Random Vectors and Matrices ......................... 5  
   1.2.3 Univariate, Multivariate, Matrix Variate Normal Distribution .... 9  
   1.2.4 Univariate Normal Distribution ....................... 9  
   1.2.5 Multivariate Normal Distribution .................... 10  
   1.2.6 Matrix Variate Normal Distribution ................. 12  
   1.2.7 Symmetric Distributions ............................... 13  
   1.2.8 Centrally Symmetric Distributions .................... 13  
   1.2.9 Spherically Symmetric Distributions ................. 14  
   1.2.10 Location-Scale Families from Spherically Symmetric Distributions .. 15  

## CHAPTER 2: A Class of Multivariate Skew Symmetric Distributions

2.1 Background: Multivariate Skew Normal Distributions ............... 18  
2.2 Skew-Centrally Symmetric Densities .......................... 22  
   2.2.1 Independent Observations, Location-Scale Family ............... 31  
2.3 Multivariate Skew-Normal Distribution .......................... 36  
   2.3.1 Motivation: Skew-Spherically Symmetric Densities ............... 36
# List of Figures

1.1 Centrally symmetric t distribution ........................................................................ 16

2.1 Surface of skew uniform density in two dimensions ............................................. 26
2.2 Surface of skew uniform density in two dimensions ............................................. 27
2.3 Surface of skew uniform density in two dimensions ............................................. 28
2.4 The contours for 2-dimensional skew normal pdf ............................................... 29
2.5 The surface for 2 dimensional pdf for skew normal variable ............................... 30
2.6 The surface for 2 dimensional pdf for skew Laplace variable ............................... 32
2.7 The surface for 2 dimensional pdf for skew-normal variable ............................... 63
2.8 The surface for 2 dimensional pdf for skew-normal variable ............................... 64
2.9 The surface for 2 dimensional pdf for skew-normal variable ............................... 65
2.10 The contours for 2 dimensional pdf for skew-normal variable ............................. 66
2.11 The surface for 2 dimensional pdf for skew-normal variable ............................... 67
2.12 The contours for 2 dimensional pdf for skew-normal variable ............................. 68

4.1 Frontier data ........................................................................................................ 93
4.2 Histogram of 5000 simulated values for $\hat{\alpha}$ by MPS ....................................... 95
4.3 1150 height measurements .................................................................................. 97
4.4 The scatter plot for 100 generated observations .................................................. 99
4.5 The scatter plot for LBM and BMI in Australian Athletes Data .............................. 100
4.6 The scatter plot for LBM and SSF in Australian Athletes Data .............................. 100
# List of Tables

<table>
<thead>
<tr>
<th>Table</th>
<th>Description</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>4.1</td>
<td>MPSIC, MPSIC(_2), and MPSIC(_3) are compared</td>
<td>91</td>
</tr>
<tr>
<td>4.2</td>
<td>MPS estimation of (\alpha)</td>
<td>93</td>
</tr>
<tr>
<td>4.3</td>
<td>ML estimation of (\alpha)</td>
<td>94</td>
</tr>
<tr>
<td>4.4</td>
<td>Estimators of (\alpha, \mu) and (\sigma)</td>
<td>98</td>
</tr>
<tr>
<td>4.5</td>
<td>The simulation results about (\hat{\mu}) and (\hat{\alpha})</td>
<td>101</td>
</tr>
<tr>
<td>4.6</td>
<td>The simulation results about (\hat{\Sigma}^{1/2})</td>
<td>102</td>
</tr>
</tbody>
</table>
CHAPTER 1

Notation, Introduction and Preliminaries

1.1 Notation

We denote matrices by capital letters, vectors by small bold letters, scalars by small letters. We usually use letters from the beginning of the alphabet to denote the constants, sometimes we use Greek letters. For random variables we choose letters from the end of the alphabet. We do not follow the convention of making a distinction between a random variable and its values. Also the following notation is used in this dissertation:

\( \mathbb{R}^k \): \( k \)-dimensional real space.

\((A)_i\): the \( i \)th row vector of the matrix \( A \).

\((A)_j\): the \( j \)th row vector of the matrix \( A \).

\((A)_{ij}\): the element located at the \( i \)th row and \( j \)th column of matrix \( A \).

\( A' \): transpose of matrix \( A \).

\( A^+ \): Moore-Penrose inverse of matrix \( A \).
$|A|$: determinant of square matrix $A$.

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

$etr(A) = e^{tr(A)}$: for square matrix $A$.

$A^{1/2}$: symmetric square root of symmetric matrix $A$.

$A_{c}^{1/2}$: square root of positive definite matrix $A$ by Cholesky decomposition.

$I_{k}$: $k$-dimensional identity matrix.

$S^{k}$: the surface of the unit sphere in $\mathbb{R}^{k}$.

$0_{k \times n}$: $k \times n$ dimensional matrix of zeros.

$e_{j}$ (or sometimes $c_{j}$): a vector that has zero elements except for one 1 at its $j$th row.

$1_{k}$: $k$-dimensional vector of ones.

$0_{k}$: $k$-dimensional vector of zeros.

$x \sim F$: $x$ is distributed as $F$.

$x \overset{d}{=} y$: $x$ and $y$ have the same distribution.

$N_{k}(\mu, \Sigma^{1/2})$: $k$-dimensional normal distribution with mean $\mu$ and covariance $\Sigma$.

$\phi_{k}(x, \mu, \Sigma)$: density of $N_{k}(\mu, \Sigma^{1/2})$ distribution evaluated at $x$.

$\Phi(x, \mu, \sigma)$: cumulative distribution function of $N_{1}(\mu, \sigma)$ distribution evaluated at $x$.

$\phi_{k}(x)$: density of $N_{k}(0, I_{k})$ distribution evaluated at $x$.

$\Phi(x)$: cumulative distribution function of $N_{1}(0, 1)$ distribution evaluated at $x$.

$\Phi_{k}(x, \mu, \Sigma)$: cumulative distribution function of $N_{k}(\mu, \Sigma^{1/2})$ distribution evaluated at $x$.

$\Phi_{k}(x)$: cumulative distribution function of $N_{k}(0, I_{k})$ distribution evaluated at $x$. 

\( \chi^2_k \): central \( \chi^2 \) distribution with \( k \) degrees of freedom.

\( \chi^2_k(\delta) \): non-central \( \chi^2 \) distribution of \( k \) degrees of freedom and non centrality parameter \( \delta \).

\( W_k(n, \Sigma) \): Wishart distribution with parameters \( n \), and \( \Sigma \).

1.2 Preliminaries

In this section, we provide certain definitions and results involving vectors and matrices as well as random vectors and matrices. Most of these results are stated here without proof; proofs can be obtained from ([4],[25]), or in most other textbooks on matrix algebra and multivariate statistics.

1.2.1 Vectors and Matrices

A \( k \)-dimensional real vector \( a \) is an ordered array of real numbers \( a_i, i = 1, 2, \ldots, k \), organized in a single column, written as \( a = (a_i) \). The set of all \( k \)-dimensional real vectors is denoted by \( \mathbb{R}^k \). A real matrix \( A \) of dimension \( k \times n \) is an ordered rectangular array of real numbers \( a_{ij} \) arranged in rows \( i = 1, 2, \ldots, k \) and columns \( j = 1, 2, \ldots, n \), written as \( A = (a_{ij}) \).

Two matrices, \( A \) and \( B \) of same dimension, are equal if \( a_{ij} = b_{ij} \) for \( i = 1, 2, \ldots, k \) and \( j = 1, 2, \ldots, n \). If all \( a_{ij} = 0 \), we write \( A = 0 \). A vector of dimension \( k \) that has all zero components except for a single component equal to 1 at its \( i \)th row is called the \( i \)th elementary vector of \( \mathbb{R}^k \) and is denoted by \( e_i^{(k)} \); we drop the superscript \( k \) when no confusion is possible.

For two \( k \times n \) matrices \( A \) and \( B \), the sum is defined as \( A + B = (a_{ij} + b_{ij}) \). For a real number \( c \), scalar multiplication is defined as \( cA = (ca_{ij}) \). Product of two matrices \( A \) of dimension \( k \times n \) and \( B \) of dimension \( n \times m \) is defined as \( AB = (c_{ij}) \) where \( c_{ij} = \sum_{\ell=1}^{k} a_{i\ell}b_{\ell j} \).

A matrix is called square matrix of order \( k \) if the number of columns or rows it has are
equal to \( k \). The transpose of a matrix \( A \) is obtained by interchanging the rows and columns of \( A \) and represented by \( A' \). When \( A = A' \), we say \( A \) is symmetric. A \( k \times k \) symmetric matrix \( A \) for which \( a_{ij} = 0, i \neq j \) for \( i, j = 1, 2, \ldots, k \) is called diagonal matrix of order \( k \), and represented by \( A = \text{diag}(a_{11}, a_{22}, \ldots, a_{kk}) \); in this case if \( a_{ii} = 1 \) for all \( i \), then we denote this by \( I_k \) and call it identity matrix. A square matrix of order \( k \) is orthogonal if \( A'A = AA' = I_k \).

A symmetric matrix \( A \) of order \( k \) is positive definite and is denoted by \( A > 0 \) if for all vectors \( a \in \mathbb{R}^k \), \( a'Aa > 0 \). A symmetric matrix \( A \) of order \( k \) is positive semi definite and is denoted by \( A \geq 0 \) if for all vectors \( a \in \mathbb{R}^k \), \( a'Aa \geq 0 \).

A symmetric matrix \( A \) of order \( k \) is positive definite and is denoted by \( A > 0 \) if for all vectors \( a \in \mathbb{R}^k \), \( a'Aa > 0 \). A symmetric matrix \( A \) of order \( k \) is positive semi definite and is denoted by \( A \geq 0 \) if for all vectors \( a \in \mathbb{R}^k \), \( a'Aa \geq 0 \).

Given a square matrix \( A \) of order \( k \), the equation \( A\mathbf{x} = \lambda \mathbf{x} \) can be solved for \( k \) pairs \((\lambda, \mathbf{x}) \ (\lambda \in \mathbb{R} \text{ and } \mathbf{x} \in \mathbb{R}^k) \), the first of a pair of solutions is called an eigenvalue, then second of the pair is called a corresponding eigenvector. Eigenvalues and eigenvectors of a matrix are usually complex, but a real symmetric matrix always have real eigenvalues and eigenvectors. Additional assumption of positive definiteness requires positive eigenvalues, assumption of nonnegative definiteness requires nonnegative eigenvalues. The number of nonzero eigenvalues of a square matrix is the rank of \( A \) and written as \( \text{rk}(A) \). We can define the rank of a \( k \times n \) matrix as \( \text{rk}(A) = \text{rk}(AA') \) or equivalently as \( \text{rk}(A) = \text{rk}(A'A) \), if \( \text{rk}(A) = \text{min}(k, n) \) then \( A \) is called a full rank matrix.

The determinant of a square matrix \( A \) of order \( k \) is defined as the product of its eigenvalues and written as \( |A| \). If \( |A| \neq 0 \), \( A \) is called nonsingular, and we can calculate the unique matrix \( B \) such that \( AB = BA = I_k \), \( B \) is called the inverse of \( A \) and written as \( A^{-1} \). A positive definite matrix is nonsingular and has positive determinant. The trace of a square matrix \( A \) of order \( k \) is defined as the sum of its eigenvalues and written as \( \text{tr}(A) \).

A symmetric matrix \( A \) of order \( k \) can be written as \( A = GDG' \) where \( G \) is an orthogonal matrix of order \( k \) and \( D \) is a diagonal matrix of order \( k \). The symmetric square root matrix \( A^{1/2} \) is defined as \( A^{1/2} = GD^{1/2}G' \) where \( D^{1/2} \) is the diagonal matrix with diagonal elements equal to the square roots of the elements of \( D \). If \( A \) is also positive definite, there exists a
unique lower triangular matrix $L$ with positive diagonal elements such that $A = LL'$. This is called the Cholesky decomposition of $A$. In the following, we denote the square root of $A$ obtained through Cholesky decomposition as $A^{1/2}$.

Suppose that $A$ is a symmetric matrix of order $k$ and rank $r \leq k$. In this case, the unique pseudo inverse of $A$ can be written as $A^{-} = GD^{-}G'$, where $D^{-}$ is the diagonal matrix of order $r$ with diagonal elements equal to the inverses of nonzero eigenvalues of $A$, $G$. A matrix of dimensions $k \times r$ is constructed such that $j$th column of it is the normalized eigenvector corresponding to the inverse of the $j$th diagonal element of $D^{-}; A$ and $A^{-}$ satisfy $AA^{-}A = A$, and $A^{-}AA^{-} = A^{-}$. Let $B$ be a $k \times n$ matrix of rank $r, r \leq k \leq n$. The unique Moore-Penrose inverse of $B$ is defined as $B^{+} = B'(B'B)^{-}$. A symmetric matrix $A$ of order $k$ is idempotent of order $k$ if it satisfies $AA' = A'A = A$. $BB^{+}$ and $B^{+}B$ are idempotent matrices.

If $A$ is idempotent of order $k$ and rank $q$, then it can be written as $A = G \begin{pmatrix} I_{q} & 0 \\ 0 & 0 \end{pmatrix} G'$ where $G$ is an orthogonal matrix of order $k$, 0’s denote zero matrices of appropriate dimensions. If $A_{1}, A_{2}, \ldots, A_{m}$ are each idempotent of order $k$ with ranks $r_{1}, r_{2}, \ldots, r_{m}$ such that $\sum_{j=1}^{m} r_{j} = k$ and if $A_{i}A_{j} = 0$ for $i \neq j$, then we can find an orthogonal matrix $G$ of order $k$

such that $A_{1} = G \begin{pmatrix} I_{r_{1}} & 0 \\ 0 & 0 \end{pmatrix} G'$, $A_{2} = G \begin{pmatrix} 0 & 0 & 0 \\ 0 & I_{r_{2}} & 0 \\ 0 & 0 & 0 \end{pmatrix} G'$, $A_{m} = G \begin{pmatrix} 0 & 0 & 0 \\ 0 & 0 & 0 \\ 0 & I_{r_{m}} \end{pmatrix} G'$ where 0’s denote zero matrices of appropriate dimensions.

1.2.2 Random Vectors and Matrices

$k$ random variables $x_{1}, x_{2}, \ldots, x_{k}$ observable on a sampling unit are represented by a $k$ dimensional column vector $x$, and is called a multivariate random variable. The matrix variable random variable obtained by considering $n$ sampling units selected to a random sample observed on $k$ variables is represented by a $k \times n$ matrix, say $X$.

The distribution of a random variable $x$ can be characterized by its cumulative distribution function (cdf) $F_{x}(.)$. It is defined as $F_{x}(c) = P(x \leq c)$. Given a cdf $F(.)$, if there exists a
nonnegative function \( f(\cdot) \) such that \( \int_{-\infty}^{\infty} f(t)dt \) for every \( x \in \mathbb{R} \), we say that the random variable \( x \) with cdf \( F_x(.) \) has probability density function (pdf), \( dF(x) = f(x) \). We can also represent a distribution using the moment generating function (mgf) of the random variable defined as \( M_x(t) = E(e^{tx}) \), if it exists for all \( t \) in some neighborhood of 0. The mgf of a random variable does not always exist, but the characteristic function (cf) always exists and it is defined as \( \Psi_x(t) = E(e^{itx}) \). More generally, the distribution of a random variable \( x \) in \( \mathbb{R}^k \) can be characterized by its joint cumulative distribution function (jcdf) \( F_x(.) \). For \( c \) in \( \mathbb{R}^k \), it is defined as \( F_x(c) = P(x_1 \leq c_1, x_2 \leq c_2, \ldots, x_k \leq c_k) \). Given a jcdf \( F(.) \), if there exists a nonnegative function \( f(.) \) such that \( \int_{-\infty}^{c_1} \int_{-\infty}^{c_2} \ldots \int_{-\infty}^{c_k} f(t_1, t_2, \ldots, t_k)dt_1dt_2\ldots dt_k \) for every \( x \in \mathbb{R}^k \), we say that the random vector \( x \) with jcdf \( F_x(.) \) has joint probability density function (jpdf), \( dF(x) = f(x) \). The joint moment generating function (jmgf) of the random variable defined as \( M_x(t) = E(e^{tx}) \), if it exists for all \( t \) in some neighborhood of 0 in \( \mathbb{R}^k \). The joint characteristic function (jcf) is defined as \( \Psi_x(t) = E(e^{itx}) \) for \( t \in \mathbb{R}^k \). We write \( x \sim F(x), x \sim f(x), x \sim M_x(t), \) or \( x \sim \Psi_x(t) \) equivalently to say that a random vector has a certain distribution. When we say that two random vectors \( x \) and \( y \) are have the same distribution, we refer to the equivalence of either jcdf’s, jpdf’s, jmgf’s, or jcf’s for these variables, in this case we write \( x \overset{d}{=} y \).

Let \( x \) be a \( k \) dimensional random variable. Suppose we know either one of the following: \( x \sim F(x), x \sim f(x), x \sim M_x(t), \) or \( x \sim \Psi_x(t) \). Let \( k \) denote the set of indices for random variables in \( x \). Let \( m \) be a subset of \( k \). The marginal distribution of \( m \)-dimensional vector of random variables with indices in \( m \), say \( x^{(m)} \), can be found in one of the following ways:

\[ x^{(m)} \sim F(x)|_{x^{(m^c)}=\infty}, \quad x^{(m)} \sim \int \ldots \int_{\mathbb{R}^{k-m}} f(x)dx^{(m^c)}, \quad x^{(m)} \sim M_x(t)_{t^{(m^c)}=0}, \quad \text{or} \]

\[ x^{(m)} \sim \Psi_x(t)_{t^{(m^c)}=0} \]  

where \( m^c \) denotes the complement of \( m \), \( x^{(m^c)} \) is the corresponding sub vector of \( x \), and similarly \( t^{(m^c)} \) denotes the sub vector of \( t \) corresponding to indices in \( m^c \). If the random variable \( x \) has a jpdf, the conditional jpdf of \( x^{(m)} \) given \( x^{(m^c)} \) is \( f(x^{(m)}|x^{(m^c)}) = f(x) f(x^{(m^c)}) \). When \( x \) does not have a density, the distribution of \( x^{(m)}|x^{(m^c)} \) is still defined from definition of conditional probability. Two random vectors, \( x \) and \( y \), are
said to be independent if \( F(x, y) = G(x)H(y) \) where \( F(x, y) \), \( G(x) \) and \( H(y) \) are the jcdf's of \( (x, y), x, \text{ and } y \) correspondingly. If they have densities \( f(x, y), g(x), \text{ and } h(y) \), then \( x \) and \( y \) are independent if and only if \( f(x, y) = g(x)h(y) \).

The expectation of a random variable \( x \) with cdf \( F(x) \) is defined by \( E(x) = \int_{\mathbb{R}} x dF(x) \). For \( r = 1, 2, \ldots, \mu^r = \int_{\mathbb{R}} (x - a)^r dF(x) \) is called the moment of order \( r \) about \( a \). In general, the expected value of an arbitrary function of \( x, g(X) \) with respect to the probability density function \( dF(x) \) is given by \( E(g(x)) = \int_{\mathbb{R}} g(x) dF(x) \). We can write \( E(x) = \mu_0 \) for the mean.

The moment of order 2 about \( E(x) \) is called the variance of the random variable \( x \), and write it as \( \text{var}(x) \). Let \( x = (x_1, x_2, \ldots, x_k)' \) be a \( k \)-dimensional random vector. If every element \( x_j \) of \( x \) has finite expectation \( \mu_j \), we write \( E(x) = (E(x_1), E(x_2), \ldots, E(x_k))' = \mu = (\mu_1, \mu_2, \ldots, \mu_k)' \) for the mean vector. A \( k \) variable moment \( \mu^{r_1 \ldots r_k}_a \) about an origin \( a = (a_1, a_2, \ldots, a_k) \) is defined as \( \mu^{r_1 \ldots r_k}_a = E((x - a_1)^{r_1} (x_2 - a_2)^{r_2} \ldots (x_k - a_k)^{r_k}) \) when it exists.

Expectation of an arbitrary function of \( x, g(x) \) is defined as \( E(g(x)) = \int_{\mathbb{R}^k} g(x) dF(x) \).

Covariance matrix \( \Sigma \) for a \( k \)-dimensional random vector \( x \) is the expectation of the \( k \times k \) matrix \( (x - \mu)(x - \mu)' \). Obviously, \( \Sigma \) is symmetric, that is, the class of covariance matrices coincides with the class of positive semi definite matrices. The following is very useful: If \( x \) is a \( k \)-dimensional random vector with mean vector \( \mu \) and covariance matrix \( \Sigma \), then the variable \( y = Ax + \xi \) for \( A \) a \( m \times k \) matrix and \( \xi \in \mathbb{R}^m \) has mean vector \( A\mu + \xi \), and covariance matrix \( A\Sigma A' \).

Assume \( x \in \chi \) be a \( k \)-dimensional random vector described by its jcdf, \( F(x) \). Let \( h(x) \) be a measurable function from \( \chi \) to \( \Upsilon \subset \mathbb{R}^m \). Then \( y = h(x) \) is a \( m \)-dimensional random vector with with its distribution described by \( H(y) = \int_{\Theta} dF(x) \), where \( \Theta = \{ x : h(x) \leq y \} \). The jcf of \( y \) can be obtained as \( \psi_y(t) = \int_{\mathbb{R}^k} e^{it'h(x)} dF(x) \) for \( t \in \mathbb{R}^m \). Similarly, the when it exists jmgf of \( y \) is given by \( M_y(t) = \int_{\mathbb{R}^k} e^{t'h(x)} dF(x) \) for \( t \in \mathbb{R}^m \).

Let \( x \in \chi \) be a \( k \)-dimensional random vector having density function \( f(x) \). Let \( h(x) \) be an invertible function from \( \chi \) to \( \Upsilon \subset \mathbb{R}^m \) such that \( h^{-1} \) has continuous partial derivatives in \( \Upsilon \). Then \( y = h(x) \) is a \( m \)-dimensional random vector with density function \( f(h^{-1}(y)) | J(y) | \),
here $|J(y)|$ is called the Jacobian of the transformation $h^{-1}$, and it is the determinant of the $k \times k$ matrix whose $ij$th component is the partial derivative of the $i$th component of $h^{-1}(y)$ with respect to $j$th component of $y$. For example, the Jacobian of the linear transformation $y = Ax + \mu$ for $A$ a nonsingular matrix of order $k$ and constant vector $\mu \in \mathbb{R}^k$ is given by $|A|^{-1}$. Also, for the $k \times n$ matrix variate random variable $X$ the transformation $Y = AXB + M$ for nonsingular matrices $A$ and $B$ of orders $k$ and $n$ and a $k \times n$ constant matrix $M$ is given by $|A|^{-n}|B|^{-k}$. We also have the following useful results for transformation of variables: If $x \sim MX(t)$, then $y = Ax + \mu \sim e^{\mu^t Mx(A't)}$ for any $m \times k$ matrix and $\mu \in \mathbb{R}^m$, if $X \sim MX(T)$ then $Y = AYB + M \sim etr(M'T)MX(A'TB')$ for any $m \times k$ matrix $A$, $n \times l$ matrix $B$ and $m \times l$ matrix $M$.

We define a model to be a $k$-dimensional random vector $x$ taking values in a set $\chi \subset \mathbb{R}^k$ and having jcdf $F(x; \theta)$ where $\theta$ is a $p$-dimensional vector of parameters taking values in $\Theta \subset \mathbb{R}^p$. $\chi$ is called the sample space, $\Theta$ is called the parameter space. In parametric statistical inference, we assume that the sample space, the parameter space, and the jcdf $F(x; \theta)$ are all known. Usually we refer to a model family in one of the following ways: $F = [\chi, F_x(x; \theta), \Theta]$ or $F = [\chi, f_x(x; \theta), \Theta]$. We try to make some inference about $\theta$ after we observe a sample of observations, say $X$.

A group of transformations, $G$, is a collection of transformations from $\chi$ to $\chi$ that is closed under inversion, and closed under composition. The model $F = [\chi, f_x(x; \theta), \Theta]$ is invariant with respect to $G$ when $x \sim f_x(x; \theta)$, $g(x) \sim f_{g(x)}(g(x); \overline{\theta}(\theta))$ where $\overline{\theta} \in \overline{G}$ and $\overline{G}$ is a group of transformations from $\Theta$ to $\Theta$. The family of the normal distributions and more generally elliptical distributions are closed under nonsingular linear transformations.

Invariance under a group of transformations, say $G$, can be useful for reducing the dimension of the data since inference about $\theta$ should not depend on whether the simple random sample $X = (x_1, x_2, \ldots, x_n)$ is observed or $g(X) = (g(x_1), g(x_2), \ldots, g(x_n))$ is observed for $g \in G$. Suppose that $F$ is invariant with respect to $G$ and that the statistic $T = T(X)$ is an estimator of $\theta$. That is, $T$ is a function that maps $\chi^n$ to $\Theta$. Then $T$ is equivariant estimator.
if \( T(g(X)) = \overline{g}(T(X)) \) for all \( g \in G, \overline{g} \in \overline{G} \), and all \( X \in \chi^n \). It can also be shown that if \( \mathbb{F} \) is invariant with respect to \( \mathbb{G} \) and the mle is unique then the mle of \( \theta \) is equivariant, if the mle is not unique then the mle can be chosen to be equivariant. A statistic \( T(X) \) is invariant under a group of transformations \( \mathbb{G} \) if \( T(g(X)) = T(X) \) for all \( g \in \mathbb{G} \) and all \( X \in \chi^n \). A function \( M(X) \) is maximal invariant under a group of transformations \( \mathbb{G} \) if \( M(g(X)) = M(X) \) for all \( g \in \mathbb{G} \) and all \( X \in \chi^n \), and if \( M(X_1) = M(X_2) \) for some \( X_1 \in \chi^n \) and \( X_2 \in \chi^n \) then \( X_1 = g(X_2) \). Suppose that \( T \) is a statistic and that \( M \) is a maximal invariant. Then \( T \) is invariant if and only if \( T \) is a function of \( M \). Invariance property is usually utilized for obtaining suitable test statistics for hypothesis testing problems that stay invariant under a group of transformations.

### 1.2.3 Univariate, Multivariate, Matrix Variate Normal Distribution

The normal distribution arises in many areas of theoretical and applied statistics. The use of the normal model is usually justified by assuming many small, independent effects additively contributing to each observation. Also, in probability theory, normal distributions arise as the limiting distributions of several continuous and discrete families of distributions. In addition, the normal distribution maximizes information entropy among all distributions with known mean and variance, which makes it the natural choice of underlying distribution for data summarized in terms of sample mean and variance. In the following, we give the definition and review some important results about the normal distribution for univariate, multivariate and matrix variate cases.

### 1.2.4 Univariate Normal Distribution

Each member of the univariate normal distribution family is defined by two parameters location and scale: the mean ("average", \( \mu \)) and variance (standard deviation squared) \( \sigma^2 \),
respectively. To indicate that a real-valued random variable \( x \) is normally distributed with mean \( \mu \) and variance \( \sigma^2 \), we write \( x \sim N(\mu, \sigma) \). The pdf of the normal distribution is the Gaussian function \( \phi(x; \mu, \sigma) = \frac{1}{\sqrt{2\pi}\sigma}e^{-\frac{1}{2\sigma^2}(x-\mu)^2} \) where \( \sigma > 0 \) is the standard deviation, the real parameter \( \mu \) is the expected value. \( \phi(z; 0, 1) = \phi(z) \) is the density function of the "standard" normal random variable. The cumulative distribution function of the normal distribution is expressed in terms of the density function \( \Phi(x; \mu, \sigma) = \int_{-\infty}^{x} \phi(u; \mu, \sigma) du \). The standard normal cdf, \( \Phi(z) \), is just the general cdf evaluated with \( \mu = 0 \) and \( \sigma = 1 \). The values \( \Phi(z) \) may be approximated very accurately by a variety of methods, such as numerical integration, Taylor series, asymptotic series and continued fractions. Mgf exists, for \( x \sim N(\mu, \sigma) \) and is given by \( M_x(t) = e^{\mu t + \frac{1}{2}t^2\sigma^2} \). Characteristic function is \( h_x(t) = e^{i\mu t - \frac{1}{2}t^2\sigma^2} \).

1.2.5 Multivariate Normal Distribution

In classical multivariate analysis the central distribution is the multivariate normal distribution. There seems to be at least two reasons for this: First due to the effect of multivariate central limit theorem, in most cases multivariate observations are at least approximately normal. Second, multivariate normal distribution and its sampling distributions are usually tractable.

A \( k \)-dimensional random vector \( \mathbf{x} \) is said to have an \( k \)-dimensional normal distribution if the distribution of \( \alpha'\mathbf{x} \) is univariate normal for every \( \alpha \in \mathbb{R}^k \). A random vector \( \mathbf{x} \) with \( k \) components has normal distribution with parameters \( \mathbf{\mu} \in \mathbb{R}^k \) and \( k \times k \) positive definite matrix \( A \) has jpdf

\[
\phi_k(\mathbf{x}; \mathbf{\mu}, A) = \frac{1}{(2\pi)^{p/2}|A|}e^{-\frac{1}{2}(\mathbf{x}-\mathbf{\mu})'(A\mathbf{A}')^{-1}(\mathbf{x}-\mathbf{\mu})}.
\]  

(1.1)

We say that \( \mathbf{x} \) is distributed according to \( N_k(\mathbf{\mu}, A) \), and write \( \mathbf{x} \sim N_k(\mathbf{\mu}, A) \). The mgf corresponding to the density in (1.1) is

\[
M_{\mathbf{x}}(t) = e^{t^\prime\mathbf{\mu} + \frac{1}{2}t^\prime AA't}.
\]  

(1.2)
If we define a \( m \)-dimensional random variable \( y \) as \( y = Bx + \xi \) where \( B \) is any \( m \times k \) matrix and \( \xi \in \mathbb{R}^m \), then by the moment generating function, we can write \( y \sim N_m(B\mu + \xi, BA) \); note that a pdf representation does not exist when \( BB' \) is singular, i.e. when \( \text{rk}(B) = \text{rk} < m \).

If \( x \sim N_k(\mu, A) \) where \( A \) is nonsingular, then \( z = A^{-1}(x - \mu) \) has the distribution \( N_k(0_k, I_k) \) with density

\[
\phi_k(z) = \frac{1}{(2\pi)^{k/2}} e^{-\frac{1}{2}z'z}
\]

and moment generating function

\[
E(e^{t'z}) = e^{-\frac{1}{2}t't}.
\]

We denote the jpdf of \( z \) with \( \phi_k(z) \), and its jcdf with \( \Phi_k(x) \). \( v = z'z \) has chi-square distribution with \( k \) degrees of freedom and denoted by \( \chi^2_k \), the density of \( v \) can be written as

\[
dF(v) = \frac{1}{2^{k/2}\Gamma(k/2)} v^{k-1/2} e^{-\frac{1}{2}v}.
\]

Let \( u \) have uniform distribution on the \( k \)-dimensional sphere, and \( r \overset{d}{=} v^{1/2} \). Then, \( z \overset{d}{=} ru \) and \( x = Az + \mu \) satisfies \( x \overset{d}{=} \mu + rAu \); in general, if \( y \) has \( N_m(\xi, B) \) distribution for any \( m \times k \) matrix \( B \) and any \( k \)-dimensional vector \( \xi \), we have \( y \overset{d}{=} \xi + rBu \).

The parameters of the \( N_k(\mu, A) \) distribution have direct interpretation as the expectation and the covariance of the variable \( x \) with this distribution since \( E(x) = \mu, \ E((x - \mu)(x - \mu)') = AA' \). Any odd moments of \( x - \mu \) are zero. The fourth order moments are \( E[(x_i - mu_i)(x_j - mu_j)(x_k - mu_k)(x_l - mu_l)] = (\sigma_{ij}\sigma_{kl} + \sigma_{ik}\sigma_{jl} + \sigma_{il}\sigma_{jk}) \) where \( AA' = (\sigma_{ij}) \).

The family of the normal distributions is closed under linear transformations, marginalization and conditioning. A characterization of multivariate normal distribution in the elliptical family of distributions is due to fact that it is the only distribution in this class for which zero covariance matrix between two partitions of a random vector implies independence.

The following are important for \( x \sim N_k(\mu, A) \), where \( A \) is nonsingular:

1. \((x - \mu)'(AA')^{-1}(x - \mu) \sim \chi^2_k\)
2. \( \mathbf{x}'(AA')^{-1}\mathbf{x} \sim \chi^2_k(\delta) \), here \( \chi^2_k(\delta) \) denotes the non central \( \chi^2 \) distribution with non centrality parameter \( \delta = \frac{1}{2}\mathbf{\mu}'(AA')^{-1}\mathbf{\mu} \).

When \( \mathbf{x} \sim N_k(\mathbf{\mu}, A) \) where \( A \) is singular of rank \( r \) we can prove similar results:

1. \( (\mathbf{x} - \mathbf{\mu})'(AA')^+(\mathbf{x} - \mathbf{\mu}) \sim \chi^2_r \)
2. \( \mathbf{x}'(AA')^+\mathbf{x} \sim \chi^2_r(\delta) \), here \( \chi^2_r(\delta) \) denotes the non central \( \chi^2 \) distribution with non centrality parameter \( \delta = \frac{1}{2}\mathbf{\mu}'(AA')^+\mathbf{\mu} \).

Assume that \( \mathbf{x} \sim N_k(\mathbf{\mu}, A) \). Let \( \mathbf{y} = B\mathbf{x} + \mathbf{b}, \mathbf{z} = C\mathbf{x} + \mathbf{c} \) where \( B \) is \( m \times k \), \( C \) is \( l \times k \), \( \mathbf{b} \) is \( m \times 1 \) and \( \mathbf{c} \) is \( l \times 1 \). Then \( \mathbf{y} \) and \( \mathbf{z} \) are independent if and only if \( BAA'C = 0 \).

### 1.2.6 Matrix Variate Normal Distribution

If a random sample of \( n \) observations are independently selected from a \( N_k(\mathbf{\mu}, A) \) distribution, the joint density of \( \mathbf{X} = (\mathbf{x}_1, \mathbf{x}_2, \ldots, \mathbf{x}_n) \) can be written as

\[
dF(\mathbf{X}) = |A|^{-n} \prod_{j=1}^{n} \phi_k((AA')^{-1/2}(\mathbf{x}_j - \mathbf{\mu})) = \frac{1}{(2\pi)^{nk/2}|A|^{n}} e^{-\frac{1}{2} \sum_{j=1}^{n} (\mathbf{x}_j - \mathbf{\mu})'(AA')^{-1}(\mathbf{x}_j - \mathbf{\mu})}.
\]

A more general matrix variate normal distribution can be defined through linear transformation to an iid random sample of a \( k \) dimensional standard normal vectors both from right and left. Let \( \mathbf{z} \sim \phi_k(\mathbf{z}) \), and let \( \mathbf{Z} = (\mathbf{z}_1, \mathbf{z}_2, \ldots, \mathbf{z}_n) \) denote an iid random sample from this distribution. The joint density for \( \mathbf{Z} \) is given by

\[
dF(\mathbf{Z}) = \prod_{j=1}^{n} \phi_k(\mathbf{z}_j) = \frac{1}{(2\pi)^{np/2}} e^{-\frac{1}{2} \sum_{j=1}^{n} \mathbf{z}_j'\mathbf{z}_j} = \frac{1}{(2\pi)^{np/2}} e^{tr\left(-\frac{1}{2}\mathbf{Z}\mathbf{Z}'\right)}.
\]

We denote this by \( \mathbf{Z} \sim \phi_{k \times n}(\mathbf{Z}) \). Now let \( \mathbf{X} = A\mathbf{Z}B + \mathbf{M} \) for nonsingular matrices \( A \) and \( B \) of orders \( k \) and \( n \) and a \( k \times n \) constant matrix \( \mathbf{M} \). The Jacobian of this transformation is
The density of $X$ is

$$
\phi_{k\times n}(X; M, A, B) = \frac{1}{(2\pi)^{nk/2}|A|^{n/2}|B|^{k/2}}e^{\text{etr}(-\frac{1}{2}(AA')^{-1}(X-M)(B'B)^{-1}(X-M)'})
$$

### 1.2.7 Symmetric Distributions

A random variable $x$ is symmetric if $x \overset{d}{=} -x$. Multivariate symmetry is usually defined in terms of invariance of the distribution of a centered random vector $x$ in $\mathbb{R}^k$ under a suitable family of transformations. A random vector $x$ has centrally symmetric distribution about the origin if $x \overset{d}{=} -x$. An equivalent alternative description for central symmetry is through requiring $v'x \overset{d}{=} -v'x$ for any vector $v \in \mathbb{R}^k$, that is any projection of $x$ to a line through the origin has a symmetric distribution. The joint distribution of $k$ symmetrically distributed random variables is centrally symmetric. A random vector $x$ has a spherically symmetric distribution about the origin $0$ if an orthogonal rotation of $x$ about $0$ by any orthogonal matrix $A$ does not alter the distribution, i.e., $x \overset{d}{=} Ax$ for all orthogonal $k \times k$ matrices $A$.

A random vector that has uniform distribution on a $k$-cube of form $[-c, c]^k$ is centrally but not spherically symmetric, central symmetry is more general than spherical symmetry. Both of these definitions of symmetry reduce to the usual notion of symmetry in the univariate case.

### 1.2.8 Centrally Symmetric Distributions

The most obvious way of constructing centrally symmetric distributions over $\mathbb{R}^k$ is through considering the joint distribution of $k$ independent symmetric random variables. Because of independence, the joint density is the product of univariate marginals. Some examples of this kind of densities follows:

1. Multivariate uniform distribution on the $k$-cube over $[-1, 1]^k$. Density:

$$
g(x) = \frac{1}{2^k}, x \in [-1, 1]^k.
$$
2. Multivariate symmetric beta distribution on the \( k \)-cube over \([-1, 1]^k\). Density:

\[
g(x) = \left( \frac{\Gamma(2\theta)}{(\Gamma(\theta))^2} \right)^k \prod_{j=1}^k \left( \frac{e_j^' x - 1/2}{2} \left( 1 - \frac{e_j^' x - 1/2}{2} \right) \right)^{\theta-1}, x \in [-1, 1]^k.
\]

3. Multivariate centrally symmetric t-distribution with \( v \) degrees of freedom. Density:

\[
g(x) = \left( \frac{\Gamma\left(\frac{v+1}{2}\right)}{\sqrt{v\pi} \Gamma\left(\frac{v}{2}\right)} \right)^k \prod_{j=1}^k \left( 1 + \frac{(e_j^' x)^2}{v} \right)^{\left(\frac{v+1}{2}\right)}, x \in \mathbb{R}^k.
\]

4. Multivariate Laplace distribution. Density:

\[
g(x; v) = \frac{1}{2^k} \exp\left( -\sum_{j=1}^k |e_j^' x| \right), x \in \mathbb{R}^k.
\]

1.2.9 Spherically Symmetric Distributions

A spherically symmetric random vector \( \mathbf{x} \) has a jcf of the form \( h(t't) \) for some cf \( h(.) \); and a density, if it exists, of the form \( f(x'x) \) for some pdf \( f(.) \). For a spherically symmetric random vector \( \mathbf{x} \), the random variable defined as \( \mathbf{u}^{(k)} = \frac{x}{\sqrt{x'x}} \) is always distributed uniformly over the surface of the unit sphere in \( \mathbb{R}^k \), denoted by \( S^k \). A stochastic representation for the random vector \( \mathbf{x} \) which has spherically symmetric distribution is given by \( \mathbf{x} \overset{\text{i.i.d.}}{=} r \mathbf{u} \) where \( r \) is a nonnegative random variable with cdf \( K(r) \) that is independent of \( \mathbf{u} \) which is uniformly distributed over \( S^k \). In this case \( r \overset{\text{i.i.d.}}{=} \sqrt{(x'x)}, \mathbf{u} \overset{\text{i.i.d.}}{=} \frac{x}{\sqrt{(x'x)}} \) and they are independent. If both \( K'(r) = k(r) \in K \) the pdf of \( r \), and \( f(x'x) \) the jpdf of \( \mathbf{x} \) exists then they are related as follows:

\[
k(r) = \frac{2\pi^{\frac{k}{2}}}{\Gamma\left(\frac{k}{2}\right)} r^{k-1} f(r^2)
\]
or sometimes we like the expression for pdf of $v = r^2$

$$k(v) = \frac{\pi^{k/2}}{\Gamma(k/2)} v^{\frac{1}{2}(k-1)} f(v).$$  \hspace{1cm} (1.7)

Some examples follow:

(a) The standard multivariate normal $N_k(0_k, I_k)$ distribution with jpdf

$$\phi_k(x) = \frac{1}{(2\pi)^{k/2}} e^{-\frac{1}{2}x'x};$$

We denote the jcdf of $x$ by $\Phi_k(x)$.

(b) The multivariate spherical $t$ distribution with $v$ degrees of freedom with density function

$$dF(x) = \frac{\Gamma(\frac{1}{2}(v + k))}{\Gamma(\frac{1}{2}v)(v\pi)^{k/2}} \frac{1}{(1 + \frac{1}{v} x'x)^{(v+k)/2}}.$$

When $n = 1$ the multivariate $t$ distribution is called spherical Cauchy distribution.

Both centrally symmetric and spherically symmetric multivariate $t$ distributions approach to the multivariate normal density. In Figure 1.2.9 we display centrally symmetric $t$ distribution for various choices of degrees of freedom, $v$. As $v$ gets larger the contours approach to the contours of the multivariate normal density.

### 1.2.10 Location-Scale Families from Spherically Symmetric Distributions

Elliptical distributions that are obtained from an affine transform of a spherically symmetric random vector are usually utilized in the study of robustness of procedures developed under the normal distribution assumption. The family of the elliptical distributions is closed under linear transformations, marginalization and conditioning.
Figure 1.1: Centrally symmetric t distribution for various choices of degrees of freedom, v. As v gets larger the contours approach to the contours of the multivariate normal density.

Let $z$ be a $k -$dimensional random vector with spherical distribution $F$. A random vector $x \overset{d}{=} Az + \mu$ for $A$ a $m \times k$ constant matrix and $\mu$ a $m -$dimensional constant vector has elliptically contoured distribution with parameters $\mu \in \mathbb{R}^k$ and $A$ with kernel $F$, we denote this distribution by $E_m(\mu, A, F)$. $x$ has a jcf of the form $e^{i\mu' h_z(A't)}$ where $h_z(t) = h(t't)$ is the jcf of the spherically distributed variable $z$. When it exists the jmgf is given by $e^{\mu'Mx(A't)}$. When $AA'$ is nonsingular and jpdf $F' = f$ exists the jpdf of $x$ is written as follows:

$$E_k(x; \mu, A, f) = |A|^{-1} f((x - \mu)'(AA')^{-1}(x - \mu)).$$ \hspace{1cm} (1.8)

A characterization of $x$ is given by $x \overset{d}{=} Az + \mu \overset{d}{=} rA u^{(k)} + \mu$ where $r$ is a nonnegative random variable with cdf $K(r)$ that is independent of $u^{(k)}$ which is uniformly distributed over $S^k$. If $x \sim E_k(\mu, A, f)$ has finite first moment then it is given by $E(x) = \mu$. Any odd moments of $x - \mu$ are zero. Covariance matrix if it exists is
given by $E[(x - \mu)(x - \mu)^\prime] = \frac{E(x^2)}{k} AA'$. If fourth order moments exists, they are $E[(x_i - mu_i)(x_j - mu_j)(x_k - mu_k)(x_l - mu_l)] = \frac{E(x^4)}{k(k+2)}(\sigma_{ij}\sigma_{kl} + \sigma_{ik}\sigma_{jl} + \sigma_{il}\sigma_{jk})$ where $AA' = (\sigma_{ij})$.

If a random sample of $n$ observations are independently selected from a $E_k(\mu, A, f)$ distribution, the density of $X = (x_1, x_2, \ldots, x_n)$ can be written as

$$E_{k \times n}(X; \mu, A, f) = |A|^{-n} \prod_{j=1}^{n} f((x_j - \mu)'(AA')^{-1}(x_j - \mu)).$$

Gupta and Varga give a generalization of this model (37):

$$E_{k \times n}(X, M, A, B, f) = |A|^{-n}|B|^{-k} f(tr((X - M)'(AA')^{-1}(X - M)(B'B)^{-1})).$$
CHAPTER 2

A Class of Multivariate Skew Symmetric Distributions

The interest in flexible distribution models that can adequately accommodate asymmetric populations has increased during the last few years. Many distributions are skewed and so do not satisfy some standard assumptions such as normality, or even elliptical symmetry. A reasonable approach is to develop models that can represent both skewed and symmetric distributions. The book edited by Genton supplies a variety of applications for such models in areas such as economics, engineering, and biological sciences ([27]).

2.1 Background: Multivariate Skew Normal Distributions

The univariate skew normal density first appears in Roberts ([55]) as an example to weighted densities. In Aigner et al., skew normal model is obtained in the context of stochastic frontier production function models ([1]). Azzalini provides a formal study
of this distribution ([10]). The density of this model is given by

\[ \text{SN}_1(y, \sigma^2, \alpha) = f(y, \sigma^2, \alpha) = 2\phi(y, \sigma^2)\Phi(\alpha y), \]

(2.1)

where \( \alpha \) is a real scalar, \( \phi(., \sigma^2) \) is the univariate normal density function with variance \( \sigma^2 \) and \( \Phi(.) \) denotes the cumulative distribution function of the univariate standard normal variable.

Some basic properties of the univariate skew normal distribution are given in ([10]):

(a) \( \text{SN}_1(y, \sigma^2, \alpha = 0) = \phi(y, \sigma^2) \),

(b) If \( y \sim \text{SN}_1(y, \sigma^2, \alpha) \) then \( -y \sim \text{SN}_1(y, \sigma^2, -\alpha) \),

(c) As \( \alpha \to \pm \infty \) the density \( \text{SN}_1(y, \sigma^2, \alpha) \) approaches to the half normal density, i.e. to the distribution of \( \pm |z| \) when \( z \sim \phi(z, \sigma^2) \),

(d) If \( y \sim \text{SN}_1(y, \sigma^2 = 1, \alpha) \) then \( y^2 \sim \chi^2_1 \).

Properties 1, 2, and 3 follow directly from the definition while Property 4 follows immediately from the following lemma:

**Lemma 2.1.1.** ([53]) \( w^2 \sim \chi^2_1 \) if and only if the p.d.f. of \( w \) has the form \( f(w) = h(w)\exp(-w^2/2) \) where \( h(w) + h(-w) = \sqrt{2/\pi} \).

The univariate skew normal distribution family extends the widely employed family of normal distributions by introducing a skewness factor \( \alpha \). The advantage of this distribution family is that it maintains many statistical properties of the normal distribution family. The study of skew normal distributions explores an approach for statistical analysis without the assumption of symmetry for the underlying population distribution. The skew normal distribution family emerges to take into account the skewness property. Skew symmetric distribution is useful in many practical situations.
A more general form of the density in (2.1) arises as

\[ f(y, \alpha) = 2g(y)H(\alpha y) \]  

(2.2)

where \( \alpha \) is a real scalar, \( g(\cdot) \) is a univariate density function symmetric around 0 and \( H(\cdot) \) is a absolutely continuous cumulative distribution function with \( H'(\cdot) \) symmetric around 0. This family of densities is called skew symmetric family. Taking \( \alpha \) equal to zero in (2.2) gives a symmetric density, we obtain a skewed density for any other value of \( \alpha \). The parameter \( \alpha \) controls both skewness and kurtosis.

When applying the skew normal distribution family in statistical inference, frequently we need to discuss the joint distribution of a random sample from the population. This consequently necessitates the study of multivariate skew normal distribution.

Several generalizations of densities (2.1), (2.2) to the multivariate case have been studied. For example, Azzalini introduced the \( k \)-dimensional skew-normal density in two alternative forms ([10], [12]):

\[ f(y, \alpha, \Sigma) = c\phi(y, \Sigma) \prod_{i=1}^{k} \Phi(\alpha_i y_i) \]  

(2.3)

\[ f(y, \alpha, \Sigma) = 2\phi(y, \Sigma)\Phi(\alpha' y) \]  

(2.4)

where \( \alpha = (\alpha_1, \alpha_2, \ldots, \alpha_k)' \) is a \( k \)-dimensional real vector, \( \phi(\cdot, \Sigma) \) is the \( k \)-dimensional normal density function with covariance matrix \( \Sigma \), \( \Phi(\cdot) \) denotes the cumulative distribution function of the univariate standard normal variable, and \( c \) is the normalizing constant.

Chang and Gupta generalize the idea in (2.4) for cases other than normal ([31]). A function

\[ f(y, \alpha) = 2g(y)H(\alpha' y) \]  

(2.5)
defines a probability density function for \( y \in \mathbb{R}^k \) when \( \alpha \) is \( k \)-dimensional real vector, \( g(.) \) is a \( k \)-dimensional density function symmetric around \( 0_k \) and \( H(.) \) is a absolutely continuous cumulative distribution function with \( H'(.) \) is symmetric around 0.

The skew symmetric family generated by Equation 2.5 has been studied by many authors. For instance, one can define skew-t distributions (16, 41, 56), skew-Cauchy distributions (9), skew-elliptical distributions (13, 6), or other skew symmetric distributions (60). The models in (35) are obtained by taking both \( g(.) \) and \( H(.) \) in (2.2) to belong to one of normal, Cauchy, Laplace, logistic or uniform family. In (49), \( g(.) \) is taken to be a normal pdf while the cumulative distributive function \( H(.) \) is taken as the cumulative distribution function of normal, Students t, Cauchy, Laplace, logistic or uniform distribution. Alternatively, Gomez, Venegas, and Bolfarine consider the situation where \( H(.) \) is fixed to be the cumulative distribution function of the normal distribution while \( g(.) \) is taken as the density function of the normal, Student’s t, logistic, Laplace, and uniform distributions (30).

Univariate and multivariate skew symmetric family of distributions appear as a subset of the so called selection models. Suppose \( x \) is a \( k \) dimensional random vector with density \( f(x) \). The usual statistical analysis involves using a random sample \( x_1, x_2, \ldots, x_n \) to make inferences about \( f(x) \). Nevertheless, there are situations for which a weighted sample instead of a random sample is selected from \( f(x) \) because it is either difficult, costly, or even impossible to observe certain parts of the distribution. If we assume that a weight function used in selection, say \( g(x) \), then the sample of observation may be thought as coming from the following weighted density

\[
h(x) = \frac{f(x)g(x)}{\int_{\mathbb{R}^k} g(x)f(x)dx}.
\]

(2.6)

When the sample is only a subset of the population then the associated model would be called a selection model. This kind of densities originate from Fisher (26). Patil, Rao
and Zelen provide a survey on this family of distributions ([51]). In their recent article, Arellano-Valle, Branco, and Genton show that a very general class of skew symmetric densities emerge from selection models called fundamental skew distribution ([5]).

### 2.2 Skew-Centrally Symmetric Densities

In this section, we study a family of multivariate skew symmetric densities generated by multivariate centrally symmetric densities.

**Theorem 2.2.1.** Let \( g(.) \) be the \( k \)-dimensional jpdf for \( k \) independent variables centrally symmetric around \( 0 \), \( H(.) \) be a absolutely continuous cumulative distribution function with \( H'(.) \) is symmetric around \( 0 \), \( \alpha = (\alpha_1, \alpha_2, \ldots, \alpha_k)' \) be a \( k \)-dimensional real vector, and \( e_j \) for \( j = 1, 2, \ldots, k \) are the elementary vectors of \( \mathbb{R}^k \). Then

\[
    f(y, \alpha) = 2^k g(y) \prod_{j=1}^{k} H(\alpha_j e'_j y) \quad (2.7)
\]

defines a probability density function.

**Proof.** First note that \( f(y) \geq 0 \) for all \( y \in \mathbb{R}^k \). We need to show that

\[
k(\alpha_1, \alpha_2, \ldots, \alpha_k) = \int_{\mathbb{R}^k} 2^k g(y) \prod_{j=1}^{k} H(\alpha_j e'_j y) dy = 1.
\]
Observe that,

\[
\frac{\partial}{\partial \alpha_\ell} k(\alpha_1, \alpha_2, \ldots, \alpha_k) = \int_{\mathbb{R}^k} \frac{d}{d\alpha_\ell} 2^k g(y) \prod_{j=1}^k H(\alpha_j e'_j y) \, dy \\
= \int_{\mathbb{R}^k} 2^k y_\ell H'(\alpha_\ell e'_\ell y) \prod_{j \neq \ell}^k H(\alpha_j e'_j y) g(y) \, dy \\
= 0.
\]

The first equality is true because of Lebesgue dominated convergence theorem, the last equality is due to independence through

\[
E_g(y_\ell H'(\alpha_\ell e'_\ell y) \prod_{j \neq \ell}^k H(\alpha_j e'_j y)) = E_g(y_\ell H'(\alpha_\ell e'_\ell y)) E_g(\prod_{j \neq \ell}^k H(\alpha_j e'_j y)) = 0,
\]

and for \(g(.)\) is centrally symmetric around 0, \(y_\ell H'(\alpha_\ell e'_\ell y)\) is an odd function of \(y_\ell\).

Hence, \(k(\alpha_1, \alpha_2, \ldots, \alpha_k)\) is constant as a function of \(\alpha_j\) for all \(j = 1, 2, \ldots, k\); and when all \(\alpha_i = 0\), \(k(\alpha_1, \alpha_2, \ldots, \alpha_k) = 1\). This concludes the proof.

In the next theorem, we relate the distribution of the even powers of a skew symmetric random variable to those of its kernel’s.

**Theorem 2.2.2.** Let \(x\) be a random variable with probability density function \(g(x)\), let \(y\) be the random variable with probability density function

\[
f(y, \alpha) = 2^k g(y) \prod_{j=1}^k H(\alpha_j e'_j y)
\]

where \(g(.)\), \(H(.)\) and \(\alpha\) are defined as in Theorem 2.2.1. Then,

(a) the even moments of \(y\) and \(x\) are the same, i.e \(E(y y')^p = E(x x')^p\) for \(p\) even

and \(E(y' y)^m = E(x' x)^m\) for any natural number \(m\),

(b) \(y' y\) and \(x' x\) have the same distribution.
Proof. It suffices to show

\[ E(x_1^{n_1}x_2^{n_2} \ldots x_k^{n_k}) = E(y_1^{n_1}y_2^{n_2} \ldots y_k^{n_k}) \]

for \( n_1, n_2, \ldots, n_k \) even.

Let \( \Psi_y(t) \) be the characteristic function of \( y \). Then,

\[ \Psi_y(t) = \int_{\mathbb{R}^k} e^{it' y} 2^k g(y) \prod_{j=1}^k H(\alpha_j e'_j y) dy. \tag{2.8} \]

Let \( n_1 + n_2 + \ldots + n_k = n \). Taking the \( n_j \)th partial derivatives of (2.8) with respect to \( t_j \) for \( j = 1, 2, \ldots, k \) and putting \( t = 0 \),

\[
\frac{\partial^n \Psi_y(t)}{\partial t_1^{n_1} \partial t_2^{n_2} \ldots \partial t_k^{n_k}} |_{t=0_k} = \int_{\mathbb{R}^k} \frac{\partial^n}{\partial t_1^{n_1} \partial t_2^{n_2} \ldots \partial t_k^{n_k}} e^{it' y} 2^k g(y) \prod_{j=1}^k H(\alpha_j e'_j y) dy |_{t=0_k} \\
= \int_{\mathbb{R}^k} [e^{it' y} 2^k \prod_{j=1}^k H(\alpha_j e'_j y)] \prod_{\ell=1}^k y_{\ell}^{n_{\ell}} g(y) dy |_{t=0_k} \\
= \int_{\mathbb{R}^k} [2^k i^n \prod_{j=1}^k H(\alpha_j e'_j y)] \prod_{\ell=1}^k y_{\ell}^{n_{\ell}} g(y) dy. \tag{2.9}
\]

Taking derivative of (2.9) with respect to \( \alpha_m \),

\[
\frac{\partial}{\partial \alpha_m} \left( \int_{\mathbb{R}^k} [2^k i^n \prod_{j=1}^k H(\alpha_j e'_j y)] \prod_{\ell=1}^k y_{\ell}^{n_{\ell}} g(y) dy \right) \\
= 2^k i^n E_g(y) [y_m^{(n_m+1)} H'(\alpha_m e'_m y) \prod_{j \neq m} y_j^{n_j} H(\alpha_j e'_j y)] \\
= 2^k i^n E_g(y) [y_m^{(n_m+1)} H'(\alpha_m e'_m y)] E_g(y) \prod_{j \neq m} y_j^{n_j} H(\alpha_j e'_j y) \\
= 0
\]

The first equality is true because of Lebesgue dominated convergence theorem, the
second equality due to the independence of components. The last equality is due to the fact that
\[ y_m^{(m+1)} H'(\alpha_m y_m) \]
is an odd function of \( y_m \) and \( g(.) \) is centrally symmetric around \( 0 \).

Therefore, \( E(y_1^{n_1} y_2^{n_2} \ldots y_k^{n_k}) \), for \( n_1, n_2, \ldots, n_k \) even, is constant as a function of \( \alpha_m \).

If all \( \alpha_m = 0 \) then \( f(x) = g(x) \) and therefore
\[ E(x_1^{n_1} x_2^{n_2} \ldots x_k^{n_k}) = E(y_1^{n_1} y_2^{n_2} \ldots y_k^{n_k}) \]
Finally,
\[ E(x_1^{n_1} x_2^{n_2} \ldots x_k^{n_k}) = E(y_1^{n_1} y_2^{n_2} \ldots y_k^{n_k}) \]
is true for all \( \alpha_m \). The required results follow immediately.

The first theorem aids in constructing families of densities. As for an example take the family of jpdf’s generated by the uniform kernel.

**Example 2.2.1. Multivariate skew uniform distribution.** Let’s take the uniform density on the interval \((-1, 1)\). The density is given by
\[ u(z) = \frac{1}{2}, \quad z \in (-1, 1). \]

A \( k \)-dimensional extension of this density is obtained by considering \( k \) independent variables, \( x = (x_1, x_2, \ldots, x_k)' \), each with density \( u(.) \). The density of \( z \) is
\[ z \sim u^*(z) = \left(\frac{1}{2}\right)^k, \quad z \in (-1, 1)^k. \]

A skew symmetric density can be obtained by using the cdf of the standard normal distribution, \( \Phi(.) \). In Theorem 2.2.1 above, use \( g(.) = u^*(.), \) and \( H(.) = \Phi(.) \), the skew
2 variate skew uniform density for \( \alpha_1 = 2, \alpha_2 = -2 \)

Figure 2.1: Surface of skew uniform density in two dimensions; \( \alpha_1 = 2, \alpha_2 = -2, H(.) = \Phi(.) \).

Symmetric density becomes

\[
f(x, \alpha) = \begin{cases} 
\prod_{j=1}^{k} \Phi(\alpha_j e'_j x), & x \in (-1, 1)^k \\
0, & \text{elsewhere}. 
\end{cases}
\]

Figures 2.1 to 2.3 show the flexibility of the skew uniform distribution.

A skew normal density is obtained from normal kernel in the following example.

**Example 2.2.2.** Multivariate skew normal densities In Theorem 2.2.1 above, let \( g(.) = \phi_k(.) \) where \( \phi_k(.) \) is the \( k \)-dimensional standard normal density function. Also let \( H(.) \) and \( \alpha \) be defined as in Theorem 2.2.1. We can construct a density for \( k \)-dimensional
Figure 2.2: Surface of skew uniform density in two dimensions; \( \alpha_1 = 2, \alpha_2 = 0 \).
Figure 2.3: Surface of skew uniform density in two dimensions; $\alpha_1 = 2$, $\alpha_2 = 10$. 
2 variable skew normal density for various choices of the shape parameter

\[ \alpha_1 = -5, \alpha_2 = -5 \]

\[ \alpha_1 = -5, \alpha_2 = 0 \]

\[ \alpha_1 = 0, \alpha_2 = 0 \]

\[ \alpha_1 = 5, \alpha_2 = -5 \]

\[ \alpha_1 = 0, \alpha_2 = 5 \]

\[ \alpha_1 = 5, \alpha_2 = 0 \]

\[ \alpha_1 = 5, \alpha_2 = 5 \]

Figure 2.4: The contours for 2 dimensional skew normal pdf for different values for \( \alpha \) and for \( H(.) = \Phi(.) \).

Joint p.d.f's of the form

\[ f(y, \alpha) = 2^k \phi_k(y) \prod_{j=1}^{k} H(\alpha_j e_j' y) \] \hspace{1cm} (2.12)

Using Theorem 2.2.2 we can relate some properties of the density introduced by Example 2.2.2 with its kernel density, \( \phi_k(.) \). Let \( x \sim \phi_k(x) \). Let \( y \) be the random variable with probability density function

\[ f(y, \alpha) = 2^k \phi_k(y) \prod_{i=1}^{k} H(\alpha_i e_i' y). \]

Then,
Figure 2.5: The surface for 2 dimensional pdf for skew normal variable for $\alpha_1 = 3$, $\alpha_2 = 5$, $H(.) = \Phi(.)$. 
(a) the even moments of $y$ and $x$ are the same, i.e $E(yy')^p = E(xx')^p$ for $p$ even and $E(y'y)^m = E(x'x)^m$ for any natural number $m$.

(b) $y'y$ and $x'x$ both have $\chi_k^2$ distribution.

**Example 2.2.3.** Multivariate skew-Laplace densities. Take $g(.)$ as the joint density of $k$ iid Laplace variables:

$$g(x) = \frac{1}{2k} \exp(-\sum_{j=1}^{k} |x_j|).$$

Let $H(.)$ and $\alpha$ be defined as in Theorem 2.2.1. We get the following density:

$$\exp(-\sum_{j=1}^{k} |x_j|) \prod_{j=1}^{k} H(\alpha e'_j x).$$

### 2.2.1 Independent Observations, Location-Scale Family

In this section, we consider a location scale family of skew symmetric random variables. But first observe the following:

**Remark 2.2.1.** The family of densities introduced by Equation 2.5 are different models from the family of densities introduced in Theorem 2.2.1 except for the case $k = 1$.

**Remark 2.2.2.** (Joint Density for Random Sample). Let $g(.)$ be a density function symmetric about 0, $H(.)$ be an absolutely continuous cumulative distribution function with $H'(.)$ symmetric around 0. Let $y_1, y_2, \ldots, y_n$ constitute a random sample from a distribution with density $f(y, \alpha) = 2g(y)H(\alpha y)$. Then the joint p.d.f. of $y = (y_1, y_2, \ldots, y_n)'$ is

$$f(y, \alpha) = 2^n g^*(y) \prod_{i=1}^{n} H(\alpha y_i) = 2^n g^*(y) \prod_{i=1}^{n} H(\alpha e'_i y_i)$$

where $g^*(y) = \prod_{i=1}^{n} g(y_i)$, belongs to the family of densities introduced in Theorem 2.2.1. Observe that density of the random sample can not be represented by the family
Figure 2.6: The surface for 2–dimensional pdf for skew Laplace variable for $\alpha_1 = 3$, $\alpha_2 = 5$, $H(.) = \Phi(.)$. 
of multivariate skew symmetric densities introduced by Equation 2.5, i.e., in the form 
\( f(y, \alpha) = 2g(y)H(\alpha'y) \). This puts a doubt on the usefulness of the latter.

**Remark 2.2.3.** (Joint Density of Independent Skew-Symmetric Variables). Let \( g(.) \) be a density function symmetric about 0, \( H(.) \) be an absolutely continuous cumulative distribution function with \( H'(.) \) symmetric around 0. Let

\[ z_j \sim f(z_j, \alpha_j) = 2g(z)H(\alpha_j z) \]

with mean \( \mu(\alpha_j) \) and variance \( \sigma^2(\alpha_j) \) for \( j = 1, 2, \ldots, k \) be independent variables. Then the joint p.d.f. of \( z = (z_1, z_2, \ldots, z_k)' \) is

\[ f(z, \alpha) = 2^k g^*(z) \prod_{j=1}^{k} H(\alpha_j e'_j z) \]

where \( \alpha = (\alpha_1, \alpha_2, \ldots, \alpha_k)' \), \( g^*(z) = \prod_{j=1}^{k} g(z_j) \) and \( e'_j \) are the elementary vectors of the coordinate system \( \mathbb{R}^k \). This density belongs to the family of densities introduced in Theorem 2.2.1. Note that the density of \( z \) can not be represented by the family of multivariate skew symmetric densities introduced by Equation 2.5.

Let \( z \sim f(z, \alpha) = 2^k g^*(z) \prod_{j=1}^{k} H(\alpha_j e'_j z) \). By construction, the mean of \( z \) is

\[ E(z) = \mu_0 = (\mu(\alpha_1), \mu(\alpha_2), \ldots, \mu(\alpha_k))', \]

and covariance is

\[ C(z) = \Sigma_0 = \text{diag}(\sigma^2(\alpha_1), \sigma^2(\alpha_2), \ldots, \sigma^2(\alpha_k)). \]

Now let \( \Sigma \) be a positive definite covariance matrix of dimension \( k \), and let \( y = \Sigma^{1/2} z \).
Then
\[ y \sim f(y, \alpha, \Sigma) = \frac{2^k}{|\Sigma|^{1/2}} g^*(\Sigma^{-1/2} y) \prod_{j=1}^{k} H(\alpha_j e_j' \Sigma^{-1/2} y). \]

By letting \( \alpha_j \Sigma^{-1/2} e_j = \lambda_j \) for \( j = 1, 2, \ldots, k \), we get
\[ f(y, \lambda_1, \lambda_2 \ldots, \lambda_k, \Sigma) = \frac{2^k}{|\Sigma|^{1/2}} g^*(\Sigma^{-1/2} y) \prod_{j=1}^{k} H(\lambda_j' y). \]

This time, the mean of \( y \) is
\[ E(y) = \Sigma^{1/2} \mu_0, \]
and covariance matrix is
\[ C(y) = \Sigma^{1/2} \Sigma_0 \Sigma^{1/2}. \]

Next, let \( x = y + \mu \) where \( \mu \in \mathbb{R}^k \). The probability density function of \( x \) is
\[ f(x, \alpha, \Sigma, \mu) = \frac{2^k}{|\Sigma|^{1/2}} g^*(\Sigma^{-1/2}(x - \mu)) \prod_{j=1}^{k} H(\alpha_j e_j' \Sigma^{-1/2}(x - \mu)). \]

The mean of \( x \) is
\[ E(x) = \Sigma^{1/2} \mu_0 + \mu, \]
and covariance matrix of \( x \) is
\[ C(x) = \Sigma^{1/2} \Sigma_0 \Sigma^{1/2}. \]

The results in the previous section lead to the following definition:

**Definition 2.2.1. (Multivariate Skew-Symmetric Density).** Let \( g(.) \) be a density function symmetric about 0, \( H(.) \) be an absolutely continuous cumulative distribution function with \( H'(.) \) symmetric around 0. A variable \( x \) has skew-symmetric distribution if it
has probability density function

\[
f(x, \alpha, \Sigma, \mu) = \frac{2^k}{|\Sigma|^{1/2}} g^*(\Sigma^{-1/2}(x - \mu)) \prod_{j=1}^{k} H(\alpha_j e_j^\prime \Sigma^{-1/2}(x - \mu))
\]  

(2.13)

where \( \alpha_j \) are scalars, \( \alpha = (\alpha_1, \alpha_2, \ldots, \alpha_k)' \), \( \mu \in \mathbb{R}^k \), \( \Sigma \) is a positive definite covariance matrix, \( g^*(y) = \prod_{j=1}^{k} g(y_j) \) is a \( k \)-dimensional density function symmetric around \( 0 \).

We call this density skew-symmetric density with location parameter \( \mu \), scale parameter \( \Sigma^{1/2} \), and shape parameter \( \alpha \) and denote it as \( ss^{g,H}_{k}(\mu, \Sigma^{1/2}, \alpha) \).

Note that we could have taken the matrix \( \Sigma^{1/2} \) in the linear form \( x = \Sigma^{1/2}z + \mu \) to be any \((m \times k)\) matrix (for \( \mu \in \mathbb{R}^m \)). Then we do not always have a density for the random variable \( x \), nevertheless the distribution of \( x \) is defined by the moment generating function if it exists.

Let \( z \sim ss^{g,H}_{k}(0_k, I_k, \alpha) \). Then the moment generating function of \( z \) evaluated at \( t_k \in \mathbb{R}^k \), \( M_z(t_k) \), can be obtained as follows.

\[
M_z(t_k) = E(e^{t_k^\prime z}) = \int \cdots \int_{\mathbb{R}^k} e^{t_k^\prime z} 2^k g^*(z) \prod_{j=1}^{k} H(\alpha_j e_j^\prime z) dz = E_{g^*(z)}(e^{t_k^\prime z} \prod_{j=1}^{k} H(\alpha_j e_j^\prime z))
\]

Let \( x = Az + \mu \) for constant \((m \times k)\) matrix \( A \) and \( m \) dimensional constant vector \( \mu \).

The moment generating function of \( x \) evaluated at \( t_m \in \mathbb{R}^m \) is \( M_x(t_m) \) can be obtained as follows.

\[
M_x(t_m) = e^{t_m^\mu} M_z(A^t_m) = e^{t_m^\mu} E_{g^*(z)}(e^{A^t_m z} \prod_{j=1}^{k} H(\alpha_j e_j^\prime z))
\]

**Definition 2.2.2.** (Multivariate Skew Symmetric Random Variable) Let \( g(.) \) be a density function symmetric about \( 0 \), \( H(.) \) be an absolutely continuous cumulative distribution function with \( H'(.) \) symmetric around \( 0 \). Let \( z_j \sim f(z, \alpha_j) = 2g(z)H(\alpha_j z) \) for
\[ j = 1, 2, \ldots, k \text{ be independent variables. Then} \]

\[ z = (z_1, z_2, \ldots, z_k)' \sim f(z, \alpha, \Sigma) = 2^k g^*(z) \prod_{j=1}^{k} H(\alpha_j e_j' z) \]

where \( g^*(z) = \prod_{j=1}^{k} g(z_j) \) and \( e_j \) for \( j = 1, 2, \ldots, k \) are the elementary vectors of the coordinate system \( \mathbb{R}^k \). Let \( A \) be a \( m \times k \) constant matrix, and \( \mu \) be a \( k \)-dimensional constant real vector. A random variable \( x = Az + \mu \) is distributed with respect to multivariate skew symmetric distribution with location parameter \( \mu \), scale parameter \( A \), and shape parameter \( \alpha \). We denote this by \( x \sim SS_{m,k}^{g,H}(\mu, A, \alpha) \).

By this definition we can write the following properties:

**Property 2.2.1.** Let \( z_j \sim f(z, \alpha_j) = 2g(z)H(\alpha_j z) \) with mean \( \mu(\alpha_j) \) and variance \( \sigma^2(\alpha_j) \) for \( j = 1, 2, \ldots, k \) be independent variables. Let \( z = (z_1, z_2, \ldots, z_k)' \) and \( x = Az + \mu \) and \( \alpha = (\alpha_1, \alpha_2, \ldots, \alpha_k)' \). Then, \( x \sim SS_{m,k}^{g,H}(\mu, A, \alpha) \). By construction, the mean of \( x \) is \( E(x) = A\mu_0 + \mu \), and covariance matrix of \( x \) is \( C(x) = A\Sigma_0 A' \) where \( \mu_0 = (\mu(\alpha_1), \mu(\alpha_2), \ldots, \mu(\alpha_k))' \) and \( \Sigma_0 = diag(\sigma^2(\alpha_1), \sigma^2(\alpha_2), \ldots, \sigma^2(\alpha_k)) \).

### 2.3 Multivariate Skew-Normal Distribution

#### 2.3.1 Motivation: Skew-Spherically Symmetric Densities

**Theorem 2.3.1.** Let \( g(.) \) be a \( k \)-dimensional spherical jpdf, i.e. satisfies \( g(x) = k(x'|x) \) for all \( x \in \mathbb{R}^k \) for some density \( k(r^2) \) defined for \( r \geq 0 \), \( H(.) \) be a absolutely continuous cumulative distribution function with \( H'(.) \) is symmetric around 0, \( \alpha = (\alpha_1, \alpha_2, \ldots, \alpha_k) \) be a \( k \)-dimensional real vector, and \( e_j \) for \( j = 1, 2, \ldots, k \) are the elementary vectors of \( \mathbb{R}^k \). Then

\[
  f(y, \alpha) = c^{-1} g(y) \prod_{j=1}^{k} H(\alpha_j e_j' y)
\]  

(2.14)
defines a probability density function where \( c = E_x(\prod_{j=1}^{k} P(z_j \leq \alpha_j x_j) | x) \) for \((z_1, z_2, \ldots, z_k)\) are identically distributed random variables with \( cdf \ H(.) \) independent of \( x \sim g(x)\).

**Proof.** \( f(y) \) is nonnegative, we have to prove \( \int_{\mathbb{R}^k} f(y)dy = 1 \). Write

\[
\int_{\mathbb{R}^k} cf(y)dy = \int_{\mathbb{R}^k} g(y) \prod_{j=1}^{k} H(\alpha_j' e_j y) dy
\]

\[= E_x(\prod_{j=1}^{k} P(z_j \leq \alpha_j x_j) | x).\]

The last equality holds because \( z_1, z_2, \ldots, z_k \) are identically distributed random variables with \( cdf \ H(.) \) independent of \( x \sim g(x) \). This concludes the proof. \( \Box \)

When we choose \( g(.) = \phi_k(.) \) and \((z_1, z_2, \ldots, z_k)'\) independent, this reduces to the density introduced in Example 2.2.1. To see this observe that \( E_x(\prod_{j=1}^{k} P(z_j \leq \alpha_j x_j) | x) = \prod_{j=1}^{k} P(z_j \leq \alpha_j x_j) = \prod_{j=1}^{k} P(z_j - \alpha_j x_j \leq 0) = \frac{1}{2^k} \) for both \( H'(.) \) and \( \phi_1(.) \) are symmetric \( z_j - \alpha_j x_j \ j = 1, 2, \ldots, k \) have iid symmetric distribution.

### 2.3.2 Multivariate Skew-Normal Distribution

**Definition**

In a recent article Chen and Gupta ([32]) pointed that neither of the multivariate skew-normal models (2.3, 2.4) cohere with the joint distribution of a random sample from a univariate skew-normal distribution and introduced an alternative multivariate skew-normal model which overcomes this problem ([32]):

\[
SN_k(y, \alpha, \Sigma) = f(y, \alpha, \Sigma) = 2^k \phi_k(y, \Sigma) \prod_{i=1}^{k} \Phi(\lambda_i'y)
\]

(2.15)
where $\alpha \in \mathbb{R}^k$ and $\Lambda = (\lambda_1, \lambda_2, \ldots, \lambda_k) = \Sigma^{-1/2} \text{diag}(\alpha_1, \alpha_2, \ldots, \alpha_k)$, $\phi_k(., \Sigma)$ is the k-dimensional normal density function with covariance $\Sigma$ and $\Phi(.)$ denotes the cumulative distribution function of the univariate standard normal variable. Chen and Gupta’s skew normal model ([32]) is in the form of skew symmetric density introduced in Definition 2.2.1.

First note that a random variable $z$ with probability density function

$$f(z, \alpha) = 2\phi(z)\Phi(\alpha z)$$

where $\alpha$ is a real scalar, $\phi(.)$ is the univariate standard normal density function and $\Phi(.)$ denotes the cumulative distribution function of the univariate standard normal variable has mean $\mu(\alpha) = \sqrt{\frac{2}{\pi(1+\alpha^2)}}$, variance $\sigma^2(\alpha) = 1 - \frac{2\alpha^2}{\pi(1+\alpha^2)}$. (Skewness: $\gamma_1 = \frac{4-\pi}{2} \frac{(\mu(\alpha))^3}{(\sigma^2(\alpha)^{3/2}}$, Kurtosis: $\gamma_2 = 2(\pi - 3) \frac{(\mu(\alpha))^4}{(\sigma^2(\alpha)^{3/2})}$)

Let $z_j \sim f(z, \alpha_j) = 2\phi(z)\Phi(\alpha_j z)$ for $j = 1, 2, \ldots, k$ be independent variables. Then the joint p.d.f. of $z = (z_1, z_2, \ldots, z_k)'$ is

$$f(z, \alpha_1, \alpha_2 \ldots, \alpha_k) = 2^k \phi_k(z) \prod_{j=1}^k \Phi(\alpha_j e_j' z)$$

where $\phi_k(z) = \prod_{j=1}^k \phi(z_j)$ is the k-dimensional standard normal variable and $e_j$ are the elementary vectors of the coordinate system $\mathbb{R}^k$. Mean of $z$ is

$$E(z) = \mu_0 = (\mu(\alpha_1), \mu(\alpha_2), \ldots, \mu(\alpha_k))',$n

and covariance is

$$C(z) = \Sigma_0 = \text{diag}(\sigma^2(\alpha_1), \sigma^2(\alpha_2), \ldots, \sigma^2(\alpha_k))$$
Now let $\Sigma$ be a positive definite covariance matrix, and let $y = \Sigma^{1/2}z$. Then

$$y \sim f(y, \alpha, \Sigma) = \frac{2^k}{|\Sigma|^{1/2}} \phi_k(\Sigma^{-1/2}y) \prod_{j=1}^k \Phi(\alpha_j e_j' \Sigma^{-1/2}y).$$

Let $x = y + \mu$ where $\mu \in \mathbb{R}^k$. The probability density function of $x$ is

$$f(x, \alpha, \Sigma, \mu) = \frac{2^k}{|\Sigma|^{1/2}} \phi_k(\Sigma^{-1/2}(x - \mu)) \prod_{j=1}^k \Phi(\alpha_j e_j' \Sigma^{-1/2}(x - \mu)).$$

Density

**Definition 2.3.1. (Multivariate Skew Normal Density).** We call the density

$$f(x, \alpha, \Sigma, \mu) = \frac{2^k}{|\Sigma|^{1/2}} \phi_k(\Sigma^{-1/2}(x - \mu)) \prod_{j=1}^k \Phi(\alpha_j e_j' \Sigma^{-1/2}(x - \mu))$$

(2.16)

as the density of a multivariate skew normal variable with location parameter $\mu$, scale parameter $\Sigma^{1/2}$, and shape parameter $\alpha = (\alpha_1, \alpha_2, \ldots, \alpha_k)' \in \mathbb{R}^k$, and denote it by $\text{sn}_k(\mu, \Sigma^{1/2}, \alpha)$.

**Remark 2.3.1.** Neither models (2.2, 2.3) can represent the joint density of independent univariate variables $x_j \sim SN_1(\sigma_j, \alpha_j)$ except for some special cases. The joint density of independent univariate skew normal variables belong the new family of densities introduced by Definition 2.3.1.

We can take the matrix $\Sigma^{1/2}$ in the linear form $x = \Sigma^{1/2}z + \mu$ to be any $(m \times k)$ matrix. Then we do not have a density for the random variable $x$ in all cases, nevertheless the distribution of $x$ is defined by the moment generating function if it exists.

**Moment Generating Function**

We need the following lemmas: See Zacks ([61]) and Chen and Gupta ([17]).
Lemma 2.3.1. Let $z \sim \phi_k(z)$. For scalar $b$, $a \in \mathbb{R}^k$, and for $\Sigma$ a positive definite matrix of order $k$ $E(\Phi(b + a'\Sigma^{1/2}z)) = \Phi\left(\frac{b}{(1+a'\Sigma a)^{1/2}}\right)$.

Lemma 2.3.2. Let $Z \sim \phi_{k \times n}(Z)$. For scalar $b$, $a \in \mathbb{R}^k$, and for $A$ and $B$ positive definite matrices of order $k$ and $n$ respectively, $E(\Phi_n(b + a'AZB)) = \Phi_n(a, (1 + a'AA'a)^{1/2}B)$.

Let $z \sim sn_k(0_k, I_k, \alpha)$. The moment generating function of $z$ evaluated at $t_k \in \mathbb{R}^k$ is $M_z(t_k)$ can be obtained as follows.

$$M_z(t_k) = E_{\phi_k(z)}(e^{t_k'z} \prod_{j=1}^k \Phi(\alpha_j e_j'z))$$

$$= 2^k \frac{k}{(2\pi)^{k/2}} \int_{\mathbb{R}^k} e^{-\frac{1}{2}z'z + t_k'z} \prod_{j=1}^k \Phi(\alpha_j e_j'z)dz$$

$$= 2^k \frac{k}{(2\pi)^{k/2}} \int_{\mathbb{R}^k} e^{-\frac{1}{2}(z'z - 2t_k'z)} \prod_{j=1}^k \Phi(\alpha_j e_j'z)dz$$

$$= 2^k e^{\frac{1}{2}t_k't_k} \frac{k}{(2\pi)^{k/2}} \int_{\mathbb{R}^k} e^{-\frac{1}{2}(z'z - 2t_k'z + t_k't_k)} \prod_{j=1}^k \Phi(\alpha_j e_j'z)dz$$

$$= 2^k e^{\frac{1}{2}t_k't_k} \frac{k}{(2\pi)^{k/2}} \int_{\mathbb{R}^k} e^{-\frac{1}{2}(z-t_k)'(z-t_k)} \prod_{j=1}^k \Phi(\alpha_j e_j'z)dz$$

$$= 2^k e^{\frac{1}{2}t_k't_k} \prod_{j=1}^k \int_{\mathbb{R}} \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}(y_j)^2} \Phi(\alpha_j y_j + \alpha_j (t_k)_j)dy_j$$

$$= 2^k e^{\frac{1}{2}t_k't_k} \prod_{j=1}^k \Phi\left(\frac{\alpha_j (t_k)_j}{\sqrt{1 + \alpha_j^2}}\right)$$

Let $x = Az + \mu$ for constant $(m \times k)$ matrix $A$ and $m$ dimensional constant vector $\mu$. Then the moment generating function of $x$ evaluated at $t_m \in \mathbb{R}^m$ is $M_x(t_m)$ can be
obtained as follows.

\[ M_x(t_m) = e^{t_m \mu} M_z(A't_m) = 2^k e^{t_m \mu + \frac{1}{2} t_m' A A' t_m} \prod_{j=1}^{k} \Phi\left( \frac{\alpha_j (A't_m)_j}{\sqrt{(1 + \alpha_j^2)}} \right) \]

Hence the following definition and theorem.

**Definition 2.3.2.** (Multivariate Skew Normal Random Variable) Let \( z_j \sim 2\phi(z) \Phi(\alpha_j z) \) for \( j = 1, 2, \ldots, k \) be independent univariate skew normal random variables. Then

\[ z = (z_1, z_2, \ldots, z_k)' \sim 2^k \phi_k(z) \prod_{j=1}^{k} \Phi(\alpha_j e'_j z) \]

where \( \phi_k(z) = \prod_{j=1}^{k} \phi(z_j) \) and \( e'_j \) are the elementary vectors of the coordinate system \( \mathbb{R}^k \). Let \( A \) be a \( m \times k \) constant matrix, and \( \mu \) be a \( k \)-dimensional constant real vector. A \( m \) dimensional random variable \( x = Az + \mu \) is distributed with respect to multivariate skew symmetric distribution with location parameter \( \mu \), scale parameter \( A \), and \( k \) dimensional shape parameter \( \alpha = (\alpha_1, \alpha_2, \ldots, \alpha_k)' \). We denote this by \( x \sim SN_{m,k}(\mu, A, \alpha) \).

**Theorem 2.3.2.** If \( x \) has multivariate skew-normal distribution \( SN_{m,k}(\mu, A, \alpha) \) then the moment generating function of \( x \) evaluated at \( t_m \in \mathbb{R}^m \) is given by

\[ M_x(t_m) = 2^k e^{t_m \mu + \frac{1}{2} t_m' A A' t_m} \prod_{j=1}^{k} \Phi\left( \frac{\alpha_j (A't_m)_j}{\sqrt{(1 + \alpha_j^2)}} \right). \]

Moments

The expectation and covariance of a multivariate skew normal variable \( x \) with distribution \( SN_{m,k}(\mu, A, \alpha) \) are easily calculated from the expectation and variance of the
univariate skew normal variable: Let $z \sim sn_k(0, I_k, \alpha)$. Then,

$$E(x) = AE(z) + \mu = \sqrt{\frac{2}{\pi}} A \left( \begin{array}{c} \frac{\alpha_1}{\sqrt{1+\alpha_1^2}} \\ \frac{\alpha_2}{\sqrt{1+\alpha_2^2}} \\ \vdots \\ \frac{\alpha_k}{\sqrt{1+\alpha_k^2}} \end{array} \right) + \mu,$$

$$Cov(x) = ACov(z)A' = A(I_k - \frac{2}{\pi} diag(\frac{\alpha_1^2}{1+\alpha_1^2}, \frac{\alpha_2^2}{1+\alpha_2^2}, \ldots, \frac{\alpha_k^2}{1+\alpha_k^2}))A'.$$

Given its jmgf, the cumulants of the multivariate skew normal distribution with $\mu = 0$ can be calculated. The cumulant generating function is given by

$$K_x(t) = log(M_x(t)) = k \log(2) + \frac{1}{2} t' A A' \sum_{j=1}^{k} \log(\Phi(\frac{\alpha_j(A't)}{\sqrt{1+\alpha_j^2}})).$$

$$\frac{\partial K_x(t)}{\partial t} \bigg|_{t=0} = AA' + \sum_{j=1}^{k} \frac{\alpha_j e_j' A'}{(1+\alpha_j^2)} \phi(\frac{\alpha_j e_j' A'}{\sqrt{1+\alpha_j^2}}) \bigg|_{t=0}$$

$$= \sqrt{\frac{2}{\pi}} A \left( \begin{array}{c} \frac{\alpha_1}{\sqrt{1+\alpha_1^2}} \\ \frac{\alpha_2}{\sqrt{1+\alpha_2^2}} \\ \vdots \\ \frac{\alpha_k}{\sqrt{1+\alpha_k^2}} \end{array} \right)$$

$$\frac{\partial^2 K_x(t)}{\partial t \partial t'} \bigg|_{t=0} = AA' + \sum_{j=1}^{k} \frac{\alpha_j^2}{1+\alpha_j^2} A e_j \otimes e_j' A' \left[ \phi'(\frac{\alpha_j e_j' A't}{\sqrt{1+\alpha_j^2}}) \bigg|_{t=0} \right] - \left( \frac{\phi(\frac{\alpha_j e_j' A't}{\sqrt{1+\alpha_j^2}})}{\phi'(\frac{\alpha_j e_j' A't}{\sqrt{1+\alpha_j^2}})} \right)^2 \bigg|_{t=0}$$

$$= A(I_k - \frac{2}{\pi} diag(\frac{\alpha_1^2}{1+\alpha_1^2}, \frac{\alpha_2^2}{1+\alpha_2^2}, \ldots, \frac{\alpha_k^2}{1+\alpha_k^2}))A'.$$
\[
\frac{\partial^3 K(t)}{\partial t \partial t' \partial t} \bigg|_{t=0} = \left( \frac{8}{\pi} - \frac{4\sqrt{2}}{\sqrt{\pi}} \right) \sum_{j=1}^{k} \left( \frac{\alpha_j}{\sqrt{1 + \alpha_j^2}} \right)^3 A e_j \otimes e'_j A' \otimes A e_j.
\]

Linear Forms

By Definition 2.3.2 we can write \( z \sim SN_{k,k}(0_k, I_k, \alpha) \), and prove the following theorems.

Theorem 2.3.3. Assume that \( y \sim SN_{m,k}(\mu, A, \alpha) \) and \( x = By + d \) with \( B \) a \((l \times m)\) matrix and \( d \) is a \( l \)-dimensional real vector. Then \( x \sim SN_{l,k}(B\mu + d, BA, \alpha) \).

Proof. From assumption we have \( y = Az + \mu \), and so \( x = (BA)z + (B\mu + d) \), i.e., \( x \sim SN_{l,k}(B\mu + d, BA, \alpha) \). \( \square \)

Corollary 2.3.1. Assume that \( y \sim SN_k(\mu, A, \alpha) \) where \( A \) is a full rank square matrix of order \( k \). Let \( \Sigma = AA' \), and let \( \Sigma^{1/2} \) be the square root of \( \Sigma \) as defined in Section 2.1.1. Assume \( x \sim SN_{k,k}(\mu, \Sigma^{1/2}, \alpha) \). Then \( x - \mu = O(y - \mu) \) where \( O = \Sigma^{1/2}A^{-1} \).

Corollary 2.3.2. Assume that \( y \sim SN_{m,k}(\mu, A, \alpha) \) and let

\[
y = \begin{pmatrix} y^{(1)} \\ y^{(2)} \end{pmatrix},
\]

\[
\mu = \begin{pmatrix} \mu^{(1)} \\ \mu^{(2)} \end{pmatrix},
\]

\[
A = \begin{bmatrix} A_{11} & A_{12} \\ A_{21} & A_{22} \end{bmatrix}
\]

where \( y^{(1)}, \mu^{(1)} \) are \( l \)-dimensional, \( A_{11} \) is \((l \times l)\). Then \( y^{(1)} \sim SN_{l,k} \left[ \mu^{(1)}; \left( \begin{array}{c} A_{11} \\ A_{12} \end{array} \right), \alpha \right] \).
Proof. In theorem above put \( B = \begin{bmatrix} I_l & 0_{l \times (m-l)} \end{bmatrix} \) and \( d = 0_l \).

Corollary 2.3.3. \( y^{(1)} \sim SN_{m,k}(\mu^{(1)}, A, \alpha^{(1)}) \), \( y^{(2)} \sim SN_{m,k}(\mu^{(2)}, B, \alpha^{(2)}) \) are independent, then \( x = y^{(1)} + y^{(2)} \) has \( SN_{m,2k}(\mu^{(1)} + \mu^{(2)}, [A, B], \begin{pmatrix} \alpha^{(1)} \\ \alpha^{(2)} \end{pmatrix}) \) distribution.

Corollary 2.3.4. Let \( y^{(i)} \sim SN_{m,k}(\mu, A, \alpha) \), be independent for \( i = 1, 2, \ldots, n \), then

\[
S = \sum_{i=1}^{n} y^{(i)} \text{ has } SN_{m,nk}(n\mu, [A, A, \ldots, A], \begin{pmatrix} \alpha \\ \alpha \\ \vdots \\ \alpha \end{pmatrix}) \text{ distribution. The jmgf can be written as}
\]

\[
M_S(t_m) = 2^{nk} e^{t_m'\mu + \frac{1}{2} t_m'AA' t_m} \prod_{j=1}^{k} \Phi\left( \frac{\alpha_j (A't_m)_j}{\sqrt{(1 + \alpha_j^2)}} \right)^n.
\]

The jmgf of \( S/n \) is

\[
M_{S/n}(t_m) = 2^{nk} e^{t_m'\mu + \frac{1}{2} t_m'AA' t_m} \prod_{j=1}^{k} \Phi\left( \frac{\alpha_j (A't_m)_j}{n\sqrt{(1 + \alpha_j^2)}} \right)^n.
\]

Some Illustrations

To illustrate the preceding theory, we consider the bivariate skew normal distribution.

Assume \((x_1, x_2)' \sim SN_{2,k}((\mu_1, \mu_2)', \begin{bmatrix} a_{11} & \ldots & a_{1k} \\ a_{21} & \ldots & a_{2k} \end{bmatrix}, (\alpha_1, \ldots, \alpha_k)')\). The mean vector is

\[
E((x_1, x_2)') = \begin{bmatrix} a_{11} & \ldots & a_{1k} \\ a_{21} & \ldots & a_{2k} \end{bmatrix} (\mu(\alpha_1)_0, \ldots, \mu(\alpha_k)_0)' + (\mu_1, \mu_2)',
\]

\[
= \begin{bmatrix} a_{11}\mu(\alpha_1)_0 + \ldots + a_{1k}\mu(\alpha_k)_0 + \mu_1 \\ a_{21}\mu(\alpha_1)_0 + \ldots + a_{2k}\mu(\alpha_k)_0 + \mu_2 \end{bmatrix}
\]

(2.18)
covariance can be written as

\[
C((x_1, x_2)') = \begin{bmatrix}
a_{11} & \ldots & a_{1k} \\
a_{21} & \ldots & a_{2k}
\end{bmatrix}
\begin{bmatrix}
\text{diag}(\sigma^2(\alpha_1), \ldots, \sigma^2(\alpha_k))
\end{bmatrix}
\begin{bmatrix}
a_{11} & \ldots & a_{1k} \\
a_{21} & \ldots & a_{2k}
\end{bmatrix}'
\]

\[
= \begin{bmatrix}
a_{11}\sigma^2(\alpha_1) & \ldots & a_{1k}\sigma^2(\alpha_k) \\
a_{21}\sigma^2(\alpha_1) & \ldots & a_{2k}\sigma^2(\alpha_k)
\end{bmatrix}
\begin{bmatrix}
a_{11} & \ldots & a_{1k} \\
a_{21} & \ldots & a_{2k}
\end{bmatrix}'
\]

\[
= \begin{bmatrix}
a_{11}^2\sigma^2(\alpha_1) + \ldots + a_{1k}^2\sigma^2(\alpha_k) & a_{11}a_{21}\sigma^2(\alpha_1) + \ldots + a_{1k}a_{2k}\sigma^2(\alpha_k) \\
a_{11}a_{21}\sigma^2(\alpha_1) + \ldots + a_{1k}a_{2k}\sigma^2(\alpha_k) & a_{21}^2\sigma^2(\alpha_1) + \ldots + a_{2k}^2\sigma^2(\alpha_k)
\end{bmatrix}
(2.19)
\]

where \(\mu(\alpha) = \sqrt{\frac{\pi}{2}} \frac{\alpha}{\sqrt{1+\alpha^2}}\), and \(\sigma^2(\alpha) = (1 - \frac{2\alpha^2}{\pi(1+\alpha^2)})\).

First, let’s take \(k = 1\).

Then the mean vector becomes

\[
E((x_1, x_2)') = (a_{11}, a_{21})'\mu(\alpha_1)_0 + (\mu_1, \mu_2)' = (\mu(\alpha_1)_0 a_{11} + \mu_1, \mu(\alpha_1)_0 a_{21} + \mu_2)'
(2.20)
\]

\[
= (\sqrt{\frac{\pi}{2}} \frac{\alpha_{11}}{\sqrt{1+\alpha_{11}^2}} + \mu_1, \sqrt{\frac{\pi}{2}} \frac{\alpha_{21}}{\sqrt{1+\alpha_{21}^2}} + \mu_2)'
(2.21)
\]

covariance can be written as

\[
C((x_1, x_2)') = (a_{11}, a_{12})\sigma^2(\alpha_1)(a_{11}, a_{12})' = (1 - \frac{2\alpha_1^2}{\pi(1+\alpha_1^2)}) \begin{bmatrix}
a_{11}^2 & a_{12}a_{21} \\
a_{12}a_{21} & a_{21}^2
\end{bmatrix}
(2.22)
\]

For \(k = 2\), the mean vector becomes

\[
E((x_1, x_2)') = \begin{bmatrix}
a_{11} & a_{12} \\
a_{21} & a_{22}
\end{bmatrix}
(\mu(\alpha_1)_0, \mu(\alpha_2)_0)' + (\mu_1, \mu_2)'
(2.23)
\]
covariance can be written as

\[
C((x_1, x_2)') = \begin{bmatrix}
    a_{11} & a_{12} \\
    a_{21} & a_{22}
\end{bmatrix}
\begin{bmatrix}
    \sigma^2(\alpha_1), \sigma^2(\alpha_2)
\end{bmatrix}
\begin{bmatrix}
    a_{11} & a_{12} \\
    a_{21} & a_{22}
\end{bmatrix}'
\] (2.25)

\[
= \begin{bmatrix}
    \sigma^2(\alpha_1)a_{11}^2 + \sigma^2(\alpha_2)a_{12}^2, & a_{11}a_{21}\sigma^2(\alpha_1) + a_{12}a_{22}\sigma^2(\alpha_2) \\
    a_{11}a_{21}\sigma^2(\alpha_1) + a_{12}a_{22}\sigma^2(\alpha_2), & \sigma^2(\alpha_1)a_{21}^2 + \sigma^2(\alpha_2)a_{22}^2
\end{bmatrix}.
\] (2.26)

For \(k = 3\), the mean is

\[
\begin{bmatrix}
    a_{11}\mu(\alpha_1)_0 + a_{12}\mu(\alpha_2)_0 + a_{13}\mu(\alpha_3)_0 + \mu_1 \\
    a_{21}\mu(\alpha_1)_0 + a_{22}\mu(\alpha_2)_0 + a_{23}\mu(\alpha_3)_0 + \mu_2
\end{bmatrix},
\] (2.27)

covariance is

\[
\begin{bmatrix}
    a_{11}\sigma^2(\alpha_1) + a_{13}\sigma^2(\alpha_3), & a_{11}a_{21}\sigma^2(\alpha_1) + a_{13}a_{23}\sigma^2(\alpha_3) \\
    a_{11}a_{21}\sigma^2(\alpha_1) + a_{13}a_{23}\sigma^2(\alpha_3), & a_{21}\sigma^2(\alpha_1) + a_{23}\sigma^2(\alpha_3)
\end{bmatrix}
\] (2.28)
Independence, Conditional Distributions

**Theorem 2.3.4.** Assume that \( y \sim sn_k(\mu, \Sigma^{1/2}, \alpha) \) and let

\[
y = \begin{pmatrix} y^{(1)} \\ y^{(2)} \end{pmatrix}
\]

\[
\mu = \begin{pmatrix} \mu^{(1)} \\ \mu^{(2)} \end{pmatrix}
\]

\[
\alpha = \begin{pmatrix} \alpha^{(1)} \\ \alpha^{(2)} \end{pmatrix}
\]

\[
\Sigma^{1/2} = \begin{pmatrix} \Sigma_{11}^{1/2} & \Sigma_{12}^{1/2} \\ \Sigma_{21}^{1/2} & \Sigma_{22}^{1/2} \end{pmatrix}
\]

where \( y^{(1)}, \mu^{(1)}, \alpha^{(1)} \) are \( l \)-dimensional, \( \Sigma_{11}^{1/2} \) is \((l \times l)\). If \( \Sigma_{12}^{1/2} = 0 \) and \( \Sigma_{21}^{1/2} = 0 \) then \( y^{(1)} \sim sn_l(\mu^{(1)}, \Sigma_{11}^{1/2}, \alpha^{(1)}) \) is independent of \( y^{(2)} \sim sn_{k-l}(\mu^{(2)}, \Sigma_{22}^{1/2}, \alpha^{(2)}) \).

**Proof.** The inverse of \( \Sigma^{1/2} \) is

\[
\Sigma^{-1/2} = \begin{pmatrix} \Sigma_{11}^{-1/2} & 0 \\ 0 & \Sigma_{22}^{-1/2} \end{pmatrix}
\]

Note that the quadratic form in the exponent of \( sn_k(\mu, \Sigma^{1/2}, \alpha) \) is

\[
Q = (y - \mu)' \Sigma^{-1} (y - \mu) = (y^{(1)} - \mu^{(1)})' \Sigma_{11}^{-1} (y^{(1)} - \mu^{(1)}) + (y^{(2)} - \mu^{(2)})' \Sigma_{22}^{-1} (y^{(2)} - \mu^{(2)}).
\]

Also

\[
\prod_{j=1}^{k} \Phi(\alpha_j e_j' \Sigma_{11}^{-1/2} (x - \mu)) = \prod_{j=1}^{l} \Phi(\alpha_j e_j' \Sigma_{11}^{-1/2} (x^{(1)} - \mu^{(1)})) \prod_{j=l+1}^{k} \Phi(\alpha_j e_j' \Sigma_{22}^{-1/2} (x^{(2)} - \mu^{(2)})).
\]
and $|\Sigma| = |\Sigma_{11}||\Sigma_{22}|$. The density of $y$ can be written as

$$sn_k(\mu, \Sigma^{1/2}, \alpha) = \frac{2^l}{|\Sigma_{11}|^{1/2}} \phi_l((\Sigma_{11}^{-1/2}(y^{(1)} - \mu^{(1)})) \prod_{j=1}^{l} \Phi(\alpha_j e_j' \Sigma_{11}^{-1/2}(y^{(1)} - \mu^{(1)}))$$

$$\times \frac{2^{k-l}}{|\Sigma_{22}|^{1/2}} \phi_{k-l}(\Sigma_{22}^{-1/2}(y^{(2)} - \mu^{(2)})) \prod_{j=1}^{k-l} \Phi(\alpha_j e_j' \Sigma_{22}^{-1/2}(y^{(2)} - \mu^{(2)}))$$

$$= sn_l(\mu^{(1)}, \Sigma_{11}^{1/2}, \alpha^{(1)}) sn_{k-l}(\mu^{(2)}, \Sigma_{22}^{1/2}, \alpha^{(2)}).$$

Since the ordering of variables are irrelevant, the above discussion also proves the following corollary:

**Corollary 2.3.5.** Let $y \sim sn_k(\mu, \Sigma^{1/2}, \alpha)$ and assume that $[i, j]$ partitions the indices $k = 1, 2, \ldots, k$ such that $(\Sigma_{ij}) = 0$ for all $i \in i$ and all $j \in j$. Then, the marginal joint distribution of variables with indices in $i$ is skew normal with location, scale, and shape parameters are obtained by taking the corresponding components of $\mu$, $\Sigma^{1/2}$ and $\alpha$, respectively.

We can also prove the following:

**Corollary 2.3.6.** Let $y \sim sn_k(\mu, A, \alpha)$ write $A = (a_1', a_2', \ldots, a_k')'$ and assume that $[i, j]$ are two disjoint subsets of the indices $k = 1, 2, \ldots, k$ such that $a_i' a_j = 0$ for all $i \in i$ and all $j \in j$. Let $A_i$ denote the matrix of rows of $A$ with indices in $i$ and $A_j$ be the matrix of rows of $A$ with indices in $j$. Then, there exists a $k$ dimensional skew normal variable, say $w$, that can be obtained by a nonsingular linear transformation to $y$, so that for $w$ the variables with indices in $i$ are independent of the variables with indices in $j$.

**Proof.** The hypothesis about the matrix of linear transformation $A$ requires that $AA' = \Sigma$ to be such that $(\Sigma)_{ij} = 0$ for all $i \in i$ and all $j \in j$. Let $x \sim sn_k(\mu, \Sigma^{1/2}, \alpha)$, and for
this distribution we have the variables with indices in \( i \) independent of the variables with indices in \( j \) by Corollary 2.3.5. By corollary 2.3.1 we can write \( y - \mu \overset{d}{=} O(x - \mu) \) for \( O = \Sigma^{1/2}A^{-1} \). This concludes the proof.

\[ \Box \]

**Corollary 2.3.7.** Assume that \( x \sim sn_k(\mu, A, \alpha) \) and let

\[
x = \begin{pmatrix} x^{(1)} \\ x^{(2)} \end{pmatrix},
\]

\[
\mu = \begin{pmatrix} \mu^{(1)} \\ \mu^{(2)} \end{pmatrix},
\]

\[
A = \begin{bmatrix} A_{11} & A_{12} \\ A_{21} & A_{22} \end{bmatrix}
\]

where \( x^{(1)}, \mu^{(1)} \) are \( l \)-dimensional, \( A_{11} \) is \((l \times l)\). Let

\[
C = \begin{bmatrix} I_l & -A_{12}A_{22}^{-1} \\ 0 & I_{k-l} \end{bmatrix}.
\]

Consider the variable

\[
y = (CAA'C')^{1/2}(CA)^{-1}(Cx - C\mu).
\]

Write

\[
y = \begin{pmatrix} y^{(1)} \\ y^{(2)} \end{pmatrix}
\]

where \( y^{(1)} \) is \( l \)-dimensional. Then \( y \sim sn_k(0_k, (CAA'C')^{1/2}, \alpha) \), and \( y^{(1)} \) is independent of \( y^{(2)} \).
Corollary 2.3.8. Assume that \( x \sim sn_k(\mu, \Sigma^{1/2}_c, \alpha) \) and let

\[
x = \begin{pmatrix} x^{(1)} \\ x^{(2)} \end{pmatrix},
\]
\[
\mu = \begin{pmatrix} \mu^{(1)} \\ \mu^{(2)} \end{pmatrix},
\]
\[
\alpha = \begin{pmatrix} \alpha^{(1)} \\ \alpha^{(2)} \end{pmatrix},
\]
\[
\Sigma^{1/2}_c = \begin{bmatrix} A_{11} & 0 \\ A_{21} & A_{22} \end{bmatrix}
\]

where \( x^{(1)}, \mu^{(1)} \) are \( l \)-dimensional, \( A_{11} \) is \((l \times l)\) lower triangular matrix with positive diagonal elements, \( A_{22} \) is \(((k-l) \times (k-l))\) lower triangular matrix with positive diagonal elements. Let

\[
C = \begin{bmatrix} I_l & 0 \\ -A_{21}A_{11}^{-1} & I_{k-l} \end{bmatrix}.
\]

Consider the variable

\[
y = Cx = \begin{pmatrix} x^{(1)} \\ y^{(2)} \end{pmatrix}.
\]

Then \( x^{(1)} \) is independent of \( y^{(2)} = x^{(2)} - A_{21}A_{11}^{-1}x^{(1)} \) which has \( sn_{k-l}(\mu^{(2)} - A_{21}A_{11}^{-1}\mu^{(1)}, A_{22}, \alpha^{(2)}) \) distribution. The joint distribution of \( y \) is given by

\[
\text{sn}_l(\mu^{(1)}, A_{11}, \alpha^{(1)}) \text{sn}_{k-l}(\mu^{(2)} - A_{21}A_{11}^{-1}\mu^{(1)}, A_{22}, \alpha^{(2)}).
\]

The density of \( x \) then can be obtained by substituting \( y^{(2)} = x^{(2)} - A_{21}A_{11}^{-1}x^{(1)} \), the
The conditional distribution of \( x^{(2)} \) given \( x^{(1)} \) is
\[
\text{sn}_{k-l}(\mu^{(2)} + A_{21}A_{11}^{-1}(x^{(1)} - \mu^{(1)})), A_{22}, \alpha^{(2)}).
\]

### 2.3.3 Generalized Multivariate Skew-Normal Distribution

We can define a multivariate skew normal density with any choice of skewing cdf \( H \) that has properties in Theorem 2.2.1. In the next section, we define and study some properties of this generalized multivariate skew normal density.

**Definition 2.3.3. (Generalized Multivariate Skew Normal Density)** Let \( x \) have a multivariate skew symmetric distribution with kernel \( \phi_k(.) \) and skewing distribution \( H(.) \) defined as in Definition 2.2.1 with location parameter \( \mu = 0_k \), scale parameter \( \Sigma^{1/2} \), and shape parameter \( \alpha = (\alpha_1, \alpha_2, \ldots, \alpha_k)' \in \mathbb{R}^k \), i.e. the density of \( x \) is
\[
f(x, \alpha, \Sigma, \mu = 0_k) = \frac{2^k}{|\Sigma|^{1/2}} \phi_k(\Sigma^{-1/2}x) \prod_{j=1}^k H(\alpha_j e'_j \Sigma^{-1/2}x).
\]

We call this density the generalized skew normal density, and denote it by \( \text{gsn}_k^H(\mu, \Sigma, \alpha) \).

**Definition 2.3.4. (Generalized Multivariate Skew Normal Random Variable)** Let \( H(.) \) be defined as in Theorem 2.2.1. Let \( z_j \sim 2\phi(z)H(\alpha_j z) \) for \( j = 1, 2, \ldots, k \) be independent univariate skew symmetric random variables. Then
\[
z = (z_1, z_2, \ldots, z_k)' \sim 2^k \phi_k(z) \prod_{j=1}^k H(\alpha_j e'_j z)
\]
where \( \phi_k(z) = \prod_{j=1}^{k} \phi(z_j) \) and \( \mathbf{e}_j' \) are the elementary vectors of the coordinate system \( \mathbb{R}^k \). Let \( A \) be a \( m \times k \) constant matrix, and \( \mu \) be a \( k \)-dimensional constant real vector.

A \( m \) dimensional random variable \( x = Az + \mu \) is distributed with respect to multivariate skew symmetric distribution with location parameter \( \mu \), scale parameter \( A \), and \( k \) dimensional shape parameter \( \alpha = (\alpha_1, \alpha_2, \ldots, \alpha_k)' \). We denote this by \( x \sim GS_{m,k}^H(\mu, A, \alpha) \).

**Theorem 2.3.5.** If \( x \) has generalized multivariate skew symmetric distribution \( GS_{m,k}^H(\mu, A, \alpha) \) then the moment generating function of \( x \) evaluated at \( t_m \in \mathbb{R}^m \) is given by

\[
M_x(t_m) = 2^k e^{t_m \mu + \frac{1}{2} t_m' A A' t_m} \prod_{j=1}^{k} E_{\phi(y)}\left(H(\alpha_j y_j + \alpha_j (A' t_m)_j)\right) \tag{2.29}
\]

**Proof.** If \( z \) has generalized multivariate skew symmetric density \( gsn_k^H(0_k, I_k, \alpha) \) then the moment generating function of \( z \) evaluated at \( t_k \in \mathbb{R}^k \) is given by
Let $x = Az + \mu$ for constant $(m \times k)$ matrix $A$ and $m$ dimensional constant vector $\mu$. Then the moment generating function of $x$ evaluated at $t_m \in \mathbb{R}^m$ is $M_x(t_m)$ can be obtained as follows.

$$M_x(t_m) = e^{t_m'\mu} M_z(A't_m) = 2^k e^{t_m'\mu + \frac{1}{2} t_m'AA't_m} \prod_{j=1}^k E_{\phi(y)}(H(\alpha_j y_j + \alpha_j(A't_m)_j))$$

By Definition 2.3.4 we can write $z \sim gsn^H_{k,k}(0_k, I_k, \alpha)$, and prove the following theorem.

**Theorem 2.3.6.** Assume that $y \sim GSN^H_{m,k}(\mu, A, \alpha)$ and $x = By + d$ with $B$ a $(l \times m)$ matrix and $d$ is a $l$-dimensional real vector. Then $x \sim GSN^H_{l,k}(B\mu + d, BA, \alpha)$. 

Proof. From assumption we have \( y = Az + \mu \), and so \( x = (BA)z + (B\mu + d) \), i.e., \( x \sim GSN^H_{l,k}(B\mu + d, BA, \alpha) \).

**Theorem 2.3.7.** Assume that \( y \sim GSN^H_{m,k}(\mu, A, \alpha) \) and let

\[
\begin{align*}
y &= \begin{pmatrix} y^{(1)} \\ y^{(2)} \end{pmatrix}, \\
\mu &= \begin{pmatrix} \mu^{(1)} \\ \mu^{(2)} \end{pmatrix}, \\
A &= \begin{bmatrix} A_{11} & A_{12} \\ A_{21} & A_{22} \end{bmatrix}
\end{align*}
\]

where \( y^{(1)}, \mu^{(1)} \) are \( l \)-dimensional, \( A_{11} \) is \((l \times l)\). Then \( y^{(1)} \sim GSN^H_{l,k}(\mu^{(1)}, A_{11}, \alpha) \).

Proof. In theorem above put \( B = \begin{bmatrix} I_l & 0_{l \times (m-l)} \end{bmatrix} \) and \( d = 0_l \). Then we obtain \( y^{(1)} \sim GSN^H_{l,k} \left( \mu^{(1)}, ( A_{11} & A_{12} ), \alpha \right) \).
If we put $H(.) = H_1(.)$ in the definition of generalized skew normal density in 2.3.3 we get the skew normal-Laplace density. To evaluate the jmgf of skew normal-Laplace random vector we use the following lemma ([40]):

**Lemma 2.3.3.** Let $z \sim \phi(z)$. Then

$$E_z(H_2(a + bz)) = \frac{1}{2} \exp(a + \frac{1}{2}b^2) \Phi\left(\frac{-a + b^2}{|b|}\right) + \Phi\left(\frac{a}{|b|}\right) - \frac{1}{2} \exp(-a + \frac{1}{2}b^2) \Phi\left(\frac{a - b^2}{|b|}\right).$$

**Corollary 2.3.9.** The jmgf of a $k-$dimensional skew normal-Laplace random vector with location parameter $\mu$, scale parameter $A$, and shape parameter $\alpha$; call $x$, is given by

$$M_x(t_m) = 2^k e^{t'\mu + \frac{1}{2}t'At_m} \prod_{j=1}^{k} \left(\frac{1}{2} e^{\alpha_j + \frac{1}{2}(At_j)^2} \Phi\left(\frac{-(\alpha_j + (At_j)^2)}{|At_j|^2}\right) + \Phi\left(\frac{\alpha_j}{|At_j|^2}\right) - \frac{1}{2} e^{-\alpha_j + \frac{1}{2}(At_j)^2} \Phi\left(\frac{\alpha_j - (At_j)^2}{|At_j|^2}\right)\right).$$

**Proof.** By Theorem 2.3.5 and Lemma 2.3.3

The odd moments of the standard univariate skew normal-Laplace random variable are given in ([50]), even moments are the same as the even moments of univariate standard normal random variable. Suppose $x$ is a random variable with standard skew normal-Laplace distribution with shape parameter $\alpha$. The first moment of $x$ is given by

$$\mu(\alpha) = E(x) = 2\alpha e^{\frac{\alpha^2}{2}} \Phi(-\alpha),$$

and we can calculate the variance from first and second moments:

$$\sigma^2(\alpha) = var(x) = E(x^2) - E^2(x) = 1 - (2\alpha e^{\frac{\alpha^2}{2}} \Phi(-\alpha))^2.$$

Let $x$ have $k-$dimensional skew normal-Laplace distribution with location parameter
\( \mu \), scale parameter \( A \), and shape parameter \( \alpha \), then

\[
E(x) = A \left( \begin{array}{c}
\mu(\alpha_1) \\
\mu(\alpha_2) \\
\vdots \\
\mu(\alpha_k)
\end{array} \right) + \mu,
\]

\[
\text{Cov}(x) = A(I_k - \text{diag}(\mu(\alpha_1)^2, \mu(\alpha_2)^2, \ldots, \mu(\alpha_k)^2))A'.
\]

**Multivariate Skew Normal-Logistic Distribution**

The cdf of logistic distribution is given by

\[
H_2(x) = \frac{1}{1 + e^{-x}}
\]

\[
= \begin{cases}
\sum_{\ell=0}^{\infty} \left( \begin{array}{c}
\ell + 1 \\
\ell
\end{array} \right) e^{-\ell x}, & x \geq 0, \\
\left( \begin{array}{c}
0 \\
\ell
\end{array} \right) e^{\ell x}, & x < 0
\end{cases}
\]

If we put \( H(.) = H_2(.) \) in the definition of generalized skew normal density in 2.3.3 we get the skew normal-logistic density. To evaluate the jmgf of skew normal-logistic random vector we use the following lemma (H0):

**Lemma 2.3.4.** Let \( z \sim \phi(z) \). Then

\[
E_z(H_2(a+bz)) = \sum_{\ell=0}^{\infty} \left( \begin{array}{c}
\ell + 1 \\
\ell
\end{array} \right) e^{-\ell a + \frac{(\alpha_2^2 + (\ell+1)^2 b^2)}{2}} \Phi\left( \frac{a - (\ell+1)b^2}{|b|} \right) + e^{(\ell+1)a + (\ell+2)(\alpha^2 b^2)} \Phi\left( \frac{a - (\ell+1)b^2}{|b|} \right).
\]

**Corollary 2.3.10.** The jmgf of a \( k \)-dimensional skew normal-logistic random vector with location parameter \( \mu \), scale parameter \( A \), and shape parameter \( \alpha \); call \( x \), is given
by

\[ M_x(t_m) = 2^k e^{t'\mu + \frac{1}{2}t'AA't_m} \]
\[
\times \prod_{j=1}^{k} \sum_{\ell=0}^{\infty} \left( -1 \right)^{\ell} e^{-\ell\alpha_j \frac{\ell \alpha_j^2}{2}} \Phi\left( \frac{\alpha_j - \ell (A't_j)^2}{|A't_j|} \right)
\]
\[
+ \frac{e^{(\ell+1)\alpha_j \frac{(\ell+1)\alpha_j^2}{2}}}{2} \Phi\left( \frac{-\alpha_j - (\ell + 1) (A't_j)^2}{|A't_j|} \right).
\]

**Proof.** By Theorem 2.3.5 and Lemma 2.3.4.

The odd moments of the standard univariate skew normal-logistic random variable are given in (50), even moments are the same as the even moments of univariate standard normal random variable. Suppose \( x \) is a random variable with standard skew normal-logistic distribution with shape parameter \( \alpha \). The first moment of \( x \) is given by

\[ \mu(\alpha) = E(x) \]
\[
= -2\alpha \left[ \sum_{\ell=0}^{\infty} \ell \left( -1 \right)^{\ell} e^{\frac{\ell \alpha^2}{2}} \Phi(-\ell \alpha) - \sum_{\ell=0}^{\infty} (\ell + 1) \left( -1 \right)^{\ell} e^{\frac{(\ell + 1)\alpha^2}{2}} \Phi(-(\ell + 1)\alpha) \right],
\]

and we can calculate the variance from first and second moments:

\[ \sigma^2(\alpha) = var(x) = E(x^2) - E^2(x) = 1 - (\mu(\alpha))^2. \]

Let \( x \) have \( k \)--dimensional skew normal-logistic distribution with location parameter
\( \mu \), scale parameter \( A \), and shape parameter \( \alpha \), then

\[
E(x) = A \left( \begin{array}{c}
\mu(\alpha_1) \\
\mu(\alpha_2) \\
\vdots \\
\mu(\alpha_k)
\end{array} \right) + \mu.
\]

\[
Cov(x) = A(I_k - \text{diag}(\mu(\alpha_1)^2, \mu(\alpha_2)^2, \ldots, \mu(\alpha_k)^2))A'.
\]

**Multivariate Skew Normal-uniform Distribution**

The cdf of uniform distribution is given by

\[
H_3(x) = \begin{cases} 
0, & x < -h, \\
\frac{x+h}{2h}, & -h \leq x \leq h, \\
1, & h < x
\end{cases}
\]

If we put \( H(.) = H_3(.) \) in the definition of generalized skew normal density in 2.3.3 we get the skew normal-uniform density. To evaluate the mgf of skew normal-uniform random vector we use the following lemma (10):

**Lemma 2.3.5.** Let \( z \sim \phi(z) \). Then

\[
E_z(H_3(a + bz)) = \frac{|b|}{2h\sqrt{2\pi}} \exp\left(-\frac{(h + a)^2}{2b^2}\right) \\
- \frac{|b|}{2h\sqrt{2\pi}} \exp\left(-\frac{(h - a)^2}{2b^2}\right) \\
+ \left(\frac{a}{2h} - \frac{1}{2}\right)\Phi\left(\frac{h - a}{|b|}\right) - 1 \\
+ \left(\frac{a}{2h} + \frac{1}{2}\right)\Phi\left(\frac{h + a}{|b|}\right).
\]

**Corollary 2.3.11.** The mgf of a \( k \)-dimensional skew normal-uniform variable with
location parameter $\mu$, scale parameter $A$, and shape parameter $\alpha$; call $x$, is given by

$$M_x(t_m) = 2^k e^{t'\mu + \frac{1}{2}t'A't_m} \prod_{j=1}^{k} \left| (A't)_j \right| \exp\left(-\frac{(h + \alpha_j)^2}{2(A't)_j^2}\right)$$

$$- \left| (A't)_j \right| \exp\left(-\frac{(h - \alpha_j)^2}{2((A't)_j)^2}\right)$$

$$+ \left( \frac{\alpha_j}{2h} - \frac{1}{2} \right) \Phi\left( \frac{h - \alpha_j}{|A't_j|} \right) - 1)$$

$$+ \left( \frac{\alpha_j}{2h} + \frac{1}{2} \right) \Phi\left( \frac{h + \alpha_j}{|A't_j|} \right).$$

Proof. By Theorem 2.3.5 and Lemma 2.3.5

The odd moments of the standard univariate skew normal-uniform random variable are given in (50), even moments are the same as the even moments of univariate standard normal random variable. Suppose $x$ is a random variable with standard skew normal-uniform distribution with shape parameter $\alpha$. The first moment of $x$ is given by

$$\mu(\alpha) = E(x) = \frac{\alpha}{h} (2\Phi\left( \frac{h}{\alpha} \right) - 1),$$

and we can calculate the variance from first and second moments:

$$\sigma^2(\alpha) = var(x) = E(x^2) - E^2(x) = 1 - (\mu(\alpha))^2.$$
\[ \text{Cov}(\mathbf{x}) = A(I_k - \text{diag}(\mu(\alpha_1)^2, \mu(\alpha_2)^2, \ldots, \mu(\alpha_k)^2))A'.} \]

### 2.3.4 Stochastic Representation, Random Number Generation

We have defined \( k \)-dimensional multivariate skew symmetric random variable through a location scale family based on \( k \) independent univariate skew symmetric random variables. This fact can be utilized to give a stochastic representation for the multivariate skew symmetric random variable which in turn can be used for random number generation.

By putting \( k = 1 \) in Theorem 2.2.1 we get a density of univariate skew symmetric variable

\[ f(y, \alpha) = 2g(y)H(\alpha y) \quad (2.30) \]

where \( \alpha \) is a real scalar, \( g(.) \) is a univariate density function symmetric around 0 and \( H(.) \) is a absolutely continuous cumulative distribution function with \( H'(.) = h(.) \) is symmetric around 0.

A probabilistic representation of a univariate skew symmetric random variable is given in the following ([10]).

**Theorem 2.3.8.** Let \( h(.) \) and \( g(.) \) be defined as above. If \( v \sim h(v) \) and \( u \sim g(u) \) are independently distributed random variables then the random variable

\[ y = \begin{cases} u, & v \leq \alpha u, \\ -u, & v > \alpha u \end{cases} \]

has density \( f(y, \alpha) \) as given in (2.30) above.

**Corollary 2.3.12.** Let

\[ z_j \sim f(z, \alpha_j) = 2g(z)H(\alpha_j z) \]
for \( j = 1, 2, \ldots, k \) be independent variables. The joint p.d.f. of \( z = (z_1, z_2, \ldots, z_k) \) is

\[
f(z, \alpha) = 2^k g^*(z) \prod_{j=1}^{k} H(\alpha_j e'_j z)
\]

where \( g^*(z) = \prod_{j=1}^{k} g(z_j) \) and \( e'_j \) are the elementary vectors of the coordinate system \( \mathbb{R}^k \). For \( \mu \in \mathbb{R}^k \) Let \( \Sigma^{1/2} \) be a positive definite matrix, \( x = \Sigma^{1/2} z + \mu \) has p.d.f.

\[
f(x, \alpha, \Sigma^{1/2}, \mu) = 2^k |\Sigma|^{1/2} g^*(\Sigma^{-1/2}(x - \mu)) \prod_{j=1}^{k} H(\alpha_j e'_j \Sigma^{-1/2}(x - \mu)). \tag{2.31}
\]

A random variable \( y \) with density \( f(y, \alpha, \Sigma^{1/2}, \mu) \) has the same distribution with \( x \).

Using Theorem 2.3.8 and Corrolary 2.3.12, we can generate skew symmetric random vectors from \( sn_k(\mu, \Sigma^{1/2}, \alpha) \) density.

### 2.4 Extensions: Some Other Skew Symmetric Forms

In Theorem 2.2.1 and 2.2.2 we have seen that a function of the form

\[
f(y, \alpha) = c^{-1} g(y) \prod_{j=1}^{k} H(\alpha_j y_j)
\]

is a jpdf for a \( k \)-dimensional random variable if certain assumptions are met. Genton and Ma ([15]) obtain a flexible family of densities by replacing \( \alpha y \) in this density by an odd function, say \( w(y) \). This motivates the following theorem.

**Theorem 2.4.1.** Let \( g(.) \) be the \( k \)-dimensional density centrally symmetric around \( 0_k \) (i.e. \( g(x) = g(-x) \)), \( H(.) \) be a absolutely continuous cumulative distribution function.
with $H'(.)$ is symmetric around 0, let $w(.)$ be an odd function. Then

$$f(y) = c^{-1} g(y) \prod_{j=1}^{k} H(w(e'_j y))$$

(2.32)

is a jpdf for a $k$-dimensional random variable where $c = E_x(\prod_{j=1}^{k} P(z_j \leq w(x_j)|x)$ for $(z_1, z_2, \ldots, z_k)'$ iid random variables with cdf $H(.)$ independent of $x \sim g(x)$.

Proof. $g(y) \prod_{j=1}^{k} H(w(y_j))$ is positive. We need to calculate $c$.

$$c = \int_{R^k} g(y) \prod_{j=1}^{k} H(w(y_j)) dy$$

$$= E_x(\prod_{j=1}^{k} P(z_j \leq w(x_j)|x)$$

since $(z_1, z_2, \ldots, z_k)'$ iid random variables with cdf $H(.)$ independent of $x \sim g(x)$.

Observe that when $g(x)$ is a density for $k$ independent variables $x_1, x_2, \ldots, x_k$, for example, when we choose $g(x) = \frac{1}{2\pi}, g(x) = \phi_k(x)$, or $g(x) = \frac{1}{2\pi} \exp(-\sum_{j=1}^{k} |x_j|)$ like in Examples 2.2.1, 2.2.2, and 2.2.3 correspondingly, $\prod_{j=1}^{k} P(z_j \leq w(x_j)) = \prod_{j=1}^{k} P(z_j - w(x_j) \leq 0)$ becomes $\frac{1}{2\pi}$ since assuming both $H'(x)$ and $\phi_1(x)$ are symmetric implies that $w(x_j)$ and therefore $z_j - w(x_j)$ for $j = 1, 2, \ldots, k$ have independent and identical symmetric distributions.

Some examples of models motivated by 2.4.1 are presented next:

**Example 2.4.1.** Take $w(x) = \frac{\lambda x}{\sqrt{1+\lambda^2 x^2}},$ we obtain a multivariate form of the skew symmetric family introduced by Arellano-Valle at al. ([7]). This density is illustrated in two dimensions in Figure 2.7.

**Example 2.4.2.** Take $w(x) = \alpha x + \beta x^3,$ we obtain a multivariate form of the skew symmetric family introduced by Ma and Genton ([43]). This density is illustrated in two dimensions in Figure 2.8.
Figure 2.7: The surface for 2 dimensional pdf for skew-normal variable in Example 2.4.1 for
\( \alpha_1 = 2, \alpha_2 = 2, \beta_1 = 5, \beta_2 = 10: \)
Example 2.4.2 with $\alpha_1=2$, $\alpha_2=1$, $\beta_1=5$, $\beta_2=10$

Figure 2.8: The surface for 2 dimensional pdf for skew-normal variable in Example 2.4.2 for $\alpha_1 = 2$, $\alpha_2 = 2$, $\beta_1 = 5$, $\beta_2 = 10$: 
Figure 2.9: The surface for 2 dimensional pdf for skew-normal variable in Example 2.4.3 for $\alpha_1 = 2$, $\alpha_2 = 4$, $\lambda_1 = 1$, $\lambda_2 = 2$:

**Example 2.4.3.** Take $w(x) = \text{sign}(x)|x|^{\alpha/2}\lambda(2/\alpha)^{1/2}$, we obtain a multivariate form of the skew symmetric family introduced by DiCiccio and Monti ([23]). This density is illustrated in two dimensions in Figures 2.9, 2.10, 2.11 and 2.4.3.

If we take the linear transformation of this variable we get the following graphs:

By the definition of multivariate weighted distribution in Equation 2.6, the skewing function of the multivariate variate skew normal density in Definition 2.3.1 i.e.

$$\prod_{j=1}^{k} \Phi(\alpha_j e_j^T (A^{-1}(x - \mu)))$$

can be replaced by

$$\Phi_m(\Gamma A^{-1}(x - \mu))$$
Figure 2.10: The contours for 2 dimensional pdf for skew-normal variable in Example 2.4.3 for \( \alpha_1 = 2, \alpha_2 = 4, \lambda_1 = 1, \lambda_2 = 2 \):
Figure 2.11: The surface for 2 dimensional pdf for skew-normal variable in Example 2.4.3 for $\alpha_1 = 2$, $\alpha_2 = 4$, $\lambda_1 = 1$, $\lambda_2 = 2$ with location $\mu = (0, 0)'$, $\Sigma^{1/2} = \begin{pmatrix} 1 & 0 \\ 1 & 2 \end{pmatrix}$
Figure 2.12: The contours for 2 dimensional pdf for skew-normal variable in Example 2.4.3 for $\alpha_1 = 2$, $\alpha_2 = 4$, $\lambda_1 = 1$, $\lambda_2 = 2$ with location $\mu = (0, 0)'$, $\Sigma^{1/2} = \begin{pmatrix} 1 & 0 \\ 1 & 2 \end{pmatrix}$.
where $\Gamma$ is a matrix of dimension $m \times k$. In this case, the normalizing constant $2^k$ will have to be changed to

$$E_z(P(y < \Gamma A^{-1}(z - \mu))|z)$$

and for $y \sim \phi_k(0, I_m)$ and $z \sim \phi_k(\mu, AA')$ this is equal to

$$\Phi_m(0; I_m + \Gamma \Gamma').$$

This extension to multivariate skew normal density is given by Gupta et al. [34] and is similar to the fundamental skew distribution of Arellano-Valle, Branco, and Genton ([5]). Note that if $k = m$ and $\Gamma$ is diagonal then we are back to the density in Definition 2.3.1.
CHAPTER 3

A Class of Matrix Variate Skew Symmetric Distributions

3.1 Distribution of Random Sample

In order to make inferences about the population parameters of a $k-$dimensional distribution, we work with a random sample of $n$ individuals from this population which can be represented by a $k \times n$ matrix $X$. The typical multivariate analysis assumes independence among the individuals. However, in many situations this assumption may be too restrictive. Matrix variable distributions can account for dependence among individual observations.

In the normal case, the matrix variate density of $X$ is written as

$$
\phi_{k \times n}(X; M, AA', B'B) = \frac{e^{tr\left(-\frac{1}{2}(AA')^{-1}(X - M)(B'B)^{-1}(X - M)'ight)}}{(2\pi)^{nk/2}|AA'|^{n/2}|B'B|^{k/2}}. \tag{3.1}
$$

Like in the multivariate case the matrix $A$ determines how the variables are related, also a matrix $B$ is introduced to account for dependence among individuals. Note that, for the density in Equation 3.1 be defined, $AA'$ and $B'B$ should be positive definite. In
practice, however, there will be natural restrictions on the parameter space. First, the knowledge about the sample design through which the variable $X$ is being observed can be reflected to the parameter space. Second, simplifying assumptions about the relationship of $k$ variables, like independence or zero correlation assumptions, can be utilized to achieve model parsimony. Finally, if there are no restrictions on $AA'$ or $B'B$ then these parameters are confounded (37).

A detailed investigation of the sources, magnitude and impact of errors is necessary to identify how survey design and procedures may be improved. For example, most data collection and sample designs involve some overlapping between interviewer workload and the sampling units (clusters). For those cases, a proportion of the measurement variance which is due to interviewers is reflected to some degree in the sampling variance calculations. The variable effects that interviewers have on respondent answers are sometimes labeled the correlated response variance (14). In the next example, we illustrate the use of matrix variate normal distribution to account for the correlated response variance.

**Example 3.1.1.** Suppose that the $k$–dimensional random variable $x$ has spherical multivariate normal density over a population, i.e. $x \sim \frac{1}{\sigma^k} \phi_k ((\sigma I_k)^{-1} x)$. In order to make inferences about the scale parameter $\sigma^2$ a random sample of individuals from this population are to be observed. In a realistic setting, there will be more than one interviewer to accomplish the work, each interviewer would observe a partition of the random sample of individuals. It is reasonable to assume that each interviewer introduces a degree of correlation to their observations and that observations from different interviewers are uncorrelated. Now, for simplicity, assume that there are only two interviewers available to observe the random sample of $n$ individuals, say I-1 and I-2. I-1 is going to observe $n_1$ of the individuals and introduces a correlation $\rho_1$ and I-2 is responsible of observing the remaining of the sample and introduces a correlation of $\rho_2$. As the following results show, the measurement variance, which is due to the
interviewers, is reflected to some degree in the sampling variance calculations.

Likelihood function given the observations can be represented with the matrix variate normal density: \( X = (x_1, x_2, \ldots, x_n) \sim \phi_{k \times n}(X; M, AA', B'B) \) with \( M = 0_{k \times n}, AA' = \sigma^2 I_k \) and

\[
B'B = \begin{bmatrix}
\Psi_1 & 0_{n-n_1 \times n-n_1} \\
0_{n-n_1 \times n-n_1} & \Psi_2
\end{bmatrix} = \Psi.
\]

\( \Psi_1 \) is a \( n_1 \times n_1 \) symmetric matrix with all diagonal entries equal to one and all its off-diagonal entries equal to \( \rho_1 \). Similarly \( \Psi_2 \) is a \( n - n_1 \times n - n_1 \) symmetric matrix with all diagonal entries equal to one and all its off-diagonal entries equal to \( \rho_2 \).

The log-likelihood is

\[
l(\sigma, \rho_1, \rho_2; X) = \log\left(\frac{1}{(2\pi)^{nk/2}}\right) - \log(\sigma^2)^{nk/2} - \log|\Psi_1|^{k/2} - \log|\Psi_2|^{k/2} - \frac{\sigma^{-2}}{2} tr(X'X\Psi^{-1}).
\]

In order to find the ML estimators of \( \rho_1, \rho_2 \) and \( \sigma^2 \), we equate the partial derivatives of \( l(\sigma, \rho_1, \rho_2; X) \). According to this, if both \( \rho_1 \) and \( \rho_2 \) are known, asymptotically unbiased maximum likelihood estimator of \( \sigma^2 \) is given by

\[
\hat{\sigma}^2(\Psi) = \frac{tr(X'X\Psi^{-1})}{nk}.
\]

The usual maximum likelihood estimator from the uncorrelated model, \( \hat{\sigma}^2 = \frac{tr(X'X)}{nk} \), is asymptotically biased for \( \sigma^2 \) where the size of bias depends on the magnitudes of \( \rho_1 \) and \( \rho_2 \). A detailed investigation of the sources, magnitude and impact of errors is necessary to identify how survey design and procedures may be improved.

Chen and Gupta extend the matrix normal distribution to accommodate skewness in the following form (177):

\[
f_1(X, \Sigma \otimes \Psi, b) = c^{*}_1 \phi_{k \times n}(X; \Sigma \otimes \Psi) \Phi_n(X'b, \Psi)
\]
where \( c_1^* = (\Phi_n(0, (1 + b'\Sigma b)\Psi))^{-1} \). A drawback of this definition is that it allows independence only over its rows or columns, but not both. Harrar ([39]) give two more definitions for the matrix variate skew normal density:

\[
f_2(X, \Sigma, \Psi, b, \Omega) = c_2^* \phi_{k \times n}(X; \Sigma \otimes \Psi) \Phi_n(X'b, \Omega)
\] (3.4)

and

\[
f_3(X, \Sigma, \Psi, b, B) = c_3^* \phi_{k \times n}(X; \Sigma \otimes \Psi) \Phi_n(tr(B'X))
\] (3.5)

where \( c_2^* = (\Phi_n(0, (\Omega + b'\Sigma b)\Psi))^{-1} \), \( c_3^* = 2 \); \( \Sigma \), \( \Psi \), and \( \Omega \) are positive definite covariance matrices of dimensions \( k \), \( n \) and \( n \) respectively, \( B \) is a matrix of dimension \( k \times n \). Note that if \( \Omega = \Psi \) then \( f_2 \) is the same as \( f_1 \). Although, more general than \( f_1 \), the densities \( f_2 \) and \( f_3 \) still do not permit independence of rows and columns simultaneously.

A very general definition of skew symmetric variable for the matrix case can be obtained from matrix variate selection models. Suppose \( X \) is a \( k \times n \) random matrix with density \( f(X) \), let \( g(X) \) be a weight function. A weighted form of density \( f(X) \) is given by

\[
h(X) = \frac{f(X)g(X)}{\int_{\mathbb{R}^{k \times n}} g(X)f(X)dX}.
\] (3.6)

When the sample is only a subset of the population then the associated model would be called a selection model. In their recent article article Domínguez-Molina, González-Farías, Ramos-Quiroga and Gupta introduced the matrix variable closed skew normal distribution in this form ([39]).

In this chapter, a construction for a family of matrix variable skew-symmetric densities that allows for independence among both variables and individuals is studied. We also consider certain extensions and relation to matrix variable closed skew normal distribution.
### 3.2 Matrix Variable Skew Symmetric Distribution

To define a matrix variate distribution from the multivariate skew symmetric distribution first assume that $z_i \sim ss_k^{g,H}(0_k, I_k, \alpha_i)$ for $i = 1, 2, \ldots, n$ are independently distributed random variables. Write $Z = (z_1, z_2, \ldots, z_n)$. We can write the density of $X$ as a product as follows,

$$\prod_{i=1}^{n} 2^k g^*(Z) \prod_{j=1}^{k} H(\alpha_{ji} e_j^t z_i).$$

This is equal to

$$2^{nk} g^{**}(Z) \prod_{i=1}^{n} \prod_{j=1}^{k} H(\alpha_{ji} e_j^t Z c_i).$$

Let $A$, and $B$ be nonsingular symmetric matrices of order $k$, and $n$ respectively, also assume $M$ is a $k \times n$ matrix. Define the matrix variate skew symmetric variable as $X = AZB + M$.

**Definition 3.2.1.** (Matrix Variate Skew-Symmetric Density) Let $g(.)$ be a density function symmetric about 0, $H(.)$ be an absolutely continuous cumulative distribution function with $H'(.)$ symmetric around 0. A variable $X$ has matrix variate skew symmetric distribution if it has probability density function

$$\frac{2^{nk} g^{**}(A^{-1}(X - M)B^{-1}) \prod_{i=1}^{n} \prod_{j=1}^{k} H(\alpha_{ji} e_j^t (A^{-1}(X - M)B^{-1}) c_i)}{|A|^n |B|^k}$$

(3.7)

where $\alpha_{ji}$ are real scalars, $M \in \mathbb{R}^{k \times n}$, $A$ and $B$ be nonsingular symmetric matrices of order $k$ and $n$ respectively. Finally, $g^{**}(X) = \prod_{i=1}^{n} \prod_{j=1}^{k} g(y_{ij})$. The density is called matrix variate skew-symmetric density with location parameter $M$, scale parameters $(A, B)$, and shape parameter $\Delta = (\alpha_{ji})$, and it is denoted by $mss_k^{g,H}(M, A, B, \Delta)$.

Let $Z \sim mss_k^{g,H}(0_{k \times n}, I_k, I_n, \Delta)$. The the moment generating function of $Z$ evaluated
at $T_{k \times n} \in \mathbb{R}^{k \times n}$ is $M_Z(T_{k \times n})$ and can be obtained as follows:

$$M_Z(T_{k \times n}) = E(etr(T_{k \times n}'Z))$$

$$= \int_{\mathbb{R}^{k \times n}} etr(T_{k \times n}'Z)2^k g^{**}(Z) \prod_{i=1}^{n} \prod_{j=1}^{k} H(\alpha_{ji} e'_j Z c_i) dZ$$

$$= E g^{**}(Z)(etr(T_{k \times n}'Z) \prod_{i=1}^{n} \prod_{j=1}^{k} H(\alpha_{ji} e'_j Z c_i)).$$

Let $X = AZB + M$ for constant $(k \times k)$ matrix $A$, $(n \times n)$ matrix $B$ and $k \times n$ dimensional constant matrix $M$. The moment generating function of $X$ evaluated at $T_{k \times n}$ is $M_X(T_{k \times n})$:

$$M_X(T_{k \times n}) = etr(T_{k \times n}'M)M_Z(A'T_{k \times n}B')$$

$$= etr(T_{k \times n}'M)E g^{**}(Z)(etr((BT_{k \times n}'A)Z) \prod_{i=1}^{n} \prod_{j=1}^{k} H(\alpha_{ji} e'_j Z c_i)).$$

**Definition 3.2.2. (Matrix Variate Skew Symmetric Random Variable)** Let $g(.)$ be a density function symmetric about 0, $H(.)$ be an absolutely continuous cumulative distribution function with $H'(.)$ symmetric around 0. Let $z_{ij} \sim f(z_{ij}, \alpha_{ji}) = 2g(z_{ij})H(\alpha_{ji}z)$ for $i = 1, 2, \ldots, n$, and $j = 1, 2, \ldots, k$ be independent variables. Then the matrix variate random variable $Z = (z_{ij})$ has density $2^{nk} \prod_{|A|^n |B|} g^{**}(Z) \prod_{i=1}^{n} \prod_{j=1}^{k} H(\alpha_{ji} e'_j Z c_i)$ where $g^{**}(z) = \prod_{i=1}^{n} \prod_{j=1}^{k} g(z_{ij})$, and $e'_j$ and $c'_i$ are the elementary vectors of the coordinate system $\mathbb{R}^k$ and $\mathbb{R}^n$ respectively. Let $A$ be a $k \times k$ constant matrix, $B$ be a $n \times n$ constant matrix and $M$ be a $k \times n$-dimensional constant matrix. A random variable $X = AZB + M$ is distributed with respect to matrix variate skew symmetric distribution with location parameter $M$, scale parameters $(A, B)$, and shape parameter $\Delta = (\alpha_{ji})$. We denote this by $X \sim MSS_{k \times n}^{g,H}(M, A, B, \Delta)$. 
3.3 Matrix Variate Skew-Normal Distribution

Definition 3.3.1. (Matrix Variate Skew Normal Density). We call the density

\[
f(X, M, A, B, \Delta) = \frac{2^{kn} \phi_{k\times n}(A^{-1}(X - M)B^{-1})}{|A|^{n/2}|B|^{k/2}} \prod_{j=1}^{k} \prod_{i=1}^{n} \phi(\alpha_{ji}e_j'(A^{-1}(X - M)B^{-1})c_i)
\]

as the density of a matrix variate skew normal variable with location parameter \(M\), scale parameters \((A, B)\), and shape parameter \(\Delta\). We denote it by \(msn_{k\times n}(M, A, B, \Delta)\).

Let \(Z \sim msn_{k\times n}(0_{k\times n}, I_k, I_n, \Delta)\). Then the moment generating function of \(Z\) evaluated at \(T_{k\times n}\) is \(M_Z(T_{k\times n})\). It can be obtained as follows:

\[
M_Z(T_{k\times n}) = E_{\phi_{k\times n}(Z)}(etr(T_{k\times n}'Z) \prod_{i=1}^{n} \prod_{j=1}^{k} \phi(\alpha_{ji}e_j'Zc_i))
\]

\[
= \frac{2^k}{(2\pi)^{k/2}} \prod_{i=1}^{n} \int_{R^k} e^{-\frac{1}{2}z_i^2+\alpha_{ji}z_i} \prod_{j=1}^{k} \phi(\alpha_{ji}e_j'z_i)dz
\]

\[
= \frac{2^k}{(2\pi)^{k/2}} \prod_{i=1}^{n} \int_{R^k} e^{-\frac{1}{2}(z_i^2-2\alpha_{ji}z_i)} \prod_{j=1}^{k} \phi(\alpha_{ji}e_j'z_i)dz
\]

\[
= \prod_{i=1}^{n} 2^k e^{\frac{1}{2}t_i^2} \prod_{j=1}^{k} \int_{R^k} e^{-\frac{1}{2}(z_i^2-2\alpha_{ji}z_i)} \phi(\alpha_{ji}e_j'z_i)dz
\]

\[
= \prod_{i=1}^{n} 2^k e^{\frac{1}{2}t_i^2} \prod_{j=1}^{k} \int_{R} e^{-\frac{1}{2}(z_i^2-2\alpha_{ji}z_i)} \phi(\alpha_{ji}z_i)dz
\]

\[
= \prod_{i=1}^{n} 2^k e^{\frac{1}{2}t_i^2} \prod_{j=1}^{k} \int_{R} e^{-\frac{1}{2}(z_i^2-2\alpha_{ji}z_i)} \phi(y_{ij})dy_{ij}
\]

\[
= \prod_{i=1}^{n} 2^k e^{\frac{1}{2}t_i^2} \prod_{j=1}^{k} \Phi(\frac{\alpha_{ji}(T)_{ij}}{\sqrt{1+\alpha_{ji}^2}})
\]

\[
= 2^{nk} etr(\frac{1}{2}T_{k\times n}'T_{k\times n}) \prod_{i=1}^{n} \prod_{j=1}^{k} \Phi(\frac{\alpha_{ji}(T)_{ij}}{\sqrt{1+\alpha_{ji}^2}}).
\]
Let $X = AZB + M$ for constant $(k \times k)$ matrix $A$, $(n \times n)$ matrix $B$ and $k \times n$ dimensional constant matrix $M$. Then the moment generating function of $X$ evaluated at $T_{k \times n} \in \mathbb{R}^{k \times n}$ is $M_X(T_{k \times n})$, this can be obtained as follows:

$$M_X(T_{k \times n}) = 2^{nk}etr(T'_{k \times n}M + \frac{1}{2}(A'T_{k \times n}B')'A'T_{k \times n}B') \prod_{i=1}^{k} \prod_{j=1}^{n} \Phi(\alpha_{ji}(A'T_{k \times n}B')_{ij}) \sqrt{1 + \alpha_{ji}^2}.$$  

Hence the following definition and theorems.

**Definition 3.3.2. (Matrix Variate Skew Normal Random Variable) Let**

$$z_{ij} \sim 2\phi(z_{ij})\Phi(\alpha_{ji}z_{ij})$$

for $i = 1, 2, \ldots, n$, and $j = 1, 2, \ldots, k$ be independent univariate skew normal random variables. Then the matrix variate random variable $Z = (z_{ij})$ has density

$$2^{nk}\phi_{k \times n}(Z) \prod_{i=1}^{n} \prod_{j=1}^{k} \Phi(\alpha_{ji}e'_jZc_i)$$

where $\phi_{k \times n}(Z) = \prod_{i=1}^{n} \prod_{j=1}^{k} \phi(z_{ij})$, and $e_j$ and $c_i$ are the elementary vectors of the coordinate system $\mathbb{R}^k$ and $\mathbb{R}^n$ respectively. Let $A$ be a $k \times k$ constant matrix, $B$ be a $n \times n$ constant matrix and $M$ be a $k \times n$-dimensional constant matrix. A random variable $X = AZB + M$ is distributed with respect to matrix variate skew symmetric distribution with location parameter $M$, scale parameters $(A, B)$, and shape parameter $\Delta = (\alpha_{ji})$. We denote this by $X \sim MSN_{k \times n}(M, A, B, \Delta)$. If the density exists it is given in Equation 3.8. We denote this case by writing $X \sim \text{msn}_{k \times n}(M, A, B, \Delta)$.

**Theorem 3.3.1. If $X$ has multivariate skew-normal distribution $MSN_{k \times n}(M, A, B, \Delta)$**
then the moment generating function of $X$ evaluated at $T_{k \times n}$ is given by

$$M_X(T_{k \times n}) = 2^{nk} \text{etr}(T_{k \times n}^T M + \frac{1}{2} (A'T_{k \times n}B')'A'T_{k \times n}B')$$

$$\times \prod_{i=1}^{n} \prod_{j=1}^{k} \Phi(\frac{\alpha_{ij}(A'T_{k \times n}B')_{ij}}{\sqrt{1 + \alpha_{ij}^2}}). \quad (3.9)$$

By Definition 3.4.2 we can write $Z \sim MSN_{k \times n}(0, I_k, I_n, \Delta)$, and prove the following theorems.

**Theorem 3.3.2.** Assume that $Y \sim MSN_{k \times n}(M, A, B, \Delta)$ and $X = CYD + N$. Then $X \sim MSN_{k \times n}(CMD + N, CA, BD, \Delta)$.

**Proof.** From assumption, we have $Y = AZB + M$, and so $X = CAZBD + (CMD + N)$, i.e., $X \sim MSN_{k \times n}(CMD + N, CA, BD, \Delta)$. \qed

**Theorem 3.3.3.** Let $x_1, x_2, \ldots, x_n$ be independent, where $x_i$ is distributed according to $sn_k(0_k, \Sigma^{1/2}, \alpha)$. Then,

$$\sum_{j=1}^{n} x_j' \Sigma^{-1} x_j \sim \chi^2_{kn}.$$

**Proof.** Let $y \sim N_k(\mu = 0, \Sigma)$. Then $y' \Sigma^{-1} y \sim \chi^2_k$, and $x_j' \Sigma^{-1} x_j$ and $y' \Sigma^{-1} y$ have the same distribution from Theorem 2.2.2. Moreover, $x_j' \Sigma^{-1} x_j$ are independent. Then the desired property is proven by the addition property of $\chi^2$ distribution. \qed

It is well known that if $X \sim \phi_{k \times n}(M, AA', \Psi = I_n)$ then the matrix variable $XX'$ has the Wishart distribution with the moment generating function given as $|(I - 2(AA'T)|^{-n/2}, (AA')^{-1} - 2T$ being a positive definite matrix. The following theorem implies that the decomposition for a Wishart matrix is not unique.

**Theorem 3.3.4.** If a $k \times n$ matrix variate random variable $X$ has $sn_{k \times n}(0_{k \times n}, A, I_n, \Delta)$ distribution for constant positive definite matrix $A$ of order $k$ and $\Delta = \alpha \mathbf{1}'$, then $XX' \sim W_k(n, AA')$. 

Proof. The moment generating function of the quadratic form $XX'$ can be obtained as follows, for any $T \in \mathbb{R}^{k \times k}$, with $(AA')^{-1} - 2T$ being a positive definite matrix,

$$E(etr(XX'T)) = \int_{\mathbb{R}^{k \times n}} etr(XX'T)dFX$$

$$= \int_{\mathbb{R}^{k \times n}} 2^{nk}etr(-\frac{1}{2}(AA')^{-1}XX' + XX'T) \prod_{i=1}^{n} \prod_{j=1}^{k} \Phi(\alpha_{ji}e_j' A^{-1} X c_i) dX$$

$$= \int_{\mathbb{R}^{k \times n}} 2^{nk}etr(-\frac{1}{2}X'((AA')^{-1} - 2T)X) \prod_{i=1}^{n} \prod_{j=1}^{k} \Phi(\alpha_{ji}e_j' A^{-1} X c_i) dX$$

$$= \int_{\mathbb{R}^{k \times n}} 2^{nk}etr(-\frac{1}{2}Z'Z) \prod_{i=1}^{n} \prod_{j=1}^{k} \Phi(\alpha_{ji}e_j' A^{-1} ((AA')^{-1} - 2T)^{1/2} Z c_i) dZ$$

$$= \int_{\mathbb{R}^{k \times n}} 2^{nk} \prod_{j=1}^{k} \prod_{i=1}^{n} e^{-\frac{1}{2} z_i'z_i} \Phi(\alpha_{ji}e_j' A^{-1} ((AA')^{-1} - 2T)^{1/2} z_i) (2\pi)^{nk/2}(I - 2(AA')T)^{n/2} dZ$$

$$= 2^{nk} (E_z \Phi_k(\alpha_{ji}e_j' A^{-1} ((AA')^{-1} - 2T)^{1/2} z_i)) \prod_{j=1}^{k} \prod_{i=1}^{n} e^{-\frac{1}{2} z_i'z_i} \Phi(\alpha_{ji}e_j' A^{-1} ((AA')^{-1} - 2T)^{1/2} z_i) (2\pi)^{nk/2}(I - 2(AA')T)^{n/2}$$

$$= |(I - 2(AA')T)|^{-n/2}.$$

\[ \square \]

### 3.4 Generalized Matrix Variate Skew Normal Distribution

**Definition 3.4.1.** (Generalized Matrix Variate Skew Normal Density). We call the density

$$f(X, M, A, B, \Delta) = \frac{2^{kn} \phi_{k \times n}(A^{-1}(X - M)B^{-1}) \prod_{j=1}^{k} H(\alpha_{ji}e_j' A^{-1}(X - M)B^{-1})c_i}{|A|^n|B|^k}$$

(3.10)

as the density of a matrix variate skew normal variable with location parameter $M$, scale
parameters \((A, B)\), and shape parameter \(\Delta\). We denote it by \(gmsn_{k \times n}^H(M, A, B, \Delta)\).

Let \(Z \sim gmsn_{k \times n}^H(0_{k \times n}, I_k, I_n\Delta)\). The moment generating function of \(Z\) evaluated at \(T_{k \times n}\) is \(M_Z(T_{k \times n})\), it can be obtained as follows:

\[
M_Z(T_{k \times n}) = E_{H_{k \times n}(Z)}(et\{(T'_{k \times n}Z) \prod_{i=1}^n \prod_{j=1}^k H(\alpha_{ji}e'_{ij}Zc_i)\})
\]

\[
= \frac{2^k}{(2\pi)^{k/2}} \prod_{i=1}^n \int_{R^k} e^{-\frac{1}{2}(z_i'z_i + t_i'z_i)} \prod_{j=1}^k H(\alpha_{ji}e'_{ij}z_i)dz
\]

\[
= \frac{2^k}{(2\pi)^{k/2}} \prod_{i=1}^n \int_{R^k} e^{-\frac{1}{2}(z_i'z_i - 2t_i'z_i)} \prod_{j=1}^k H(\alpha_{ji}e'_{ij}z_i)dz
\]

\[
= \prod_{i=1}^n \frac{2^k}{(2\pi)^{k/2}} \int_{R^k} e^{-\frac{1}{2}(z_i'z_i - 2t_i'z_i + t_i't_i)} \prod_{j=1}^k H(\alpha_{ji}e'_{ij}z_i)dz
\]

\[
= \prod_{i=1}^n \left(\frac{2^k}{(2\pi)^{k/2}} \int_{R} e^{-\frac{1}{2}(z_i')^2} \prod_{j=1}^k H(\alpha_{ji}z_{ij})dz_{ij}\right)
\]

\[
= \prod_{i=1}^n \frac{2^k}{(2\pi)^{k/2}} \int_{R} \frac{1}{\sqrt{(2\pi)}} e^{-\frac{1}{2}(y_{ij})^2} H(\alpha_{ji}y_{ij} + \alpha_{ji}(t_{ij})dy_{ij}
\]

\[
= \prod_{i=1}^n \frac{2^k}{(2\pi)^{k/2}} \int_{R} H(\frac{\alpha_{ji}(t_{ij})}{\sqrt{1 + \alpha_{ji}^2}})
\]

\[
= 2^k et(\frac{1}{2}T_{k \times n}^T T_{k \times n} - n) \prod_{i=1}^n \prod_{j=1}^k H(\frac{\alpha_{ji}(T_{ij})}{\sqrt{1 + \alpha_{ji}^2}}).
\]

Let \(X = AZB + M\) for constant \((k \times k)\) matrix \(A\), \((n \times n)\) matrix \(B\) and \(k \times n\) dimensional constant matrix \(M\). Then the moment generating function of \(X\) evaluated at \(T_{k \times n} \in \mathbb{R}^{k \times n}\) is \(M_X(T_{k \times n})\), this can be obtained as follows:

\[
M_X(T_{k \times n}) = 2^k et(\frac{1}{2}T_{k \times n}^T M + \frac{1}{2}(A'T_{k \times n}B')A'T_{k \times n}B') \prod_{i=1}^n \prod_{j=1}^k H(\frac{\alpha_{ji}(A'T_{k \times n}B')_{ij}}{\sqrt{1 + \alpha_{ji}^2}}).
\]

Hence the following definition and theorems.
Definition 3.4.2. (Matrix Variate Generalized Skew Normal Random Variable) Let
\[ z_{ij} \sim 2\phi(z_{ij})H(\alpha_{ji}z_{ij}) \]
for \( i = 1, 2, \ldots, n \), and \( j = 1, 2, \ldots, k \) be independent univariate generalized skew normal random variables. The matrix variate random variable \( Z = (z_{ij}) \) has density
\[
2^{nk}\phi_{k\times n}(Z)\prod_{i=1}^{n}\prod_{j=1}^{k}H(\alpha_{ji}e'_{j}Zc_{i})
\]
where \( \phi_{k\times n}(Z) = \prod_{i=1}^{n}\prod_{j=1}^{k}\phi(z_{ij}) \), and \( e'_{j} \) and \( c'_{i} \) are the elementary vectors of the coordinate system \( \mathbb{R}^{k} \) and \( \mathbb{R}^{n} \) respectively. Let \( A \) be a \( k \times k \) constant matrix, \( B \) be a \( n \times n \) constant matrix and \( M \) be a \( k \times n \)-dimensional constant matrix. A random variable \( X = AZB + M \) is distributed with respect to generalized matrix variate skew symmetric distribution with location parameter \( M \), scale parameters \((A, B)\), and shape parameter \( \Delta = (\alpha_{ji}) \). We denote this by \( X \sim gmsn_{k\times n}^{H}(M, A, B, \Delta) \).

Theorem 3.4.1. If \( X \) has generalized matrix variate skew-normal distribution, \( Gmsn_{k\times n}^{H}(M, A, B, \Delta) \), then the moment generating function of \( X \) evaluated at \( T_{k\times n} \) is given by
\[
M_X(T_{k\times n}) = 2^{nk}\text{etr}(T_{k\times n}^{\prime}M + \frac{1}{2}(A'T_{k\times n}B')^{\prime}A'T_{k\times n}B')
\times \prod_{i=1}^{n}\prod_{j=1}^{k}H\left(\frac{\alpha_{ji}(A'T_{k\times n}B')_{ij}}{\sqrt{1 + \alpha_{ji}^{2}}}\right). \tag{3.11}
\]

By Definition 3.4.2 we can write \( Z \sim Gmsn_{k\times n}^{H}(0, I_k, I_n, \Delta) \), and prove the following theorems.

Theorem 3.4.2. Assume that \( Y \sim Gmsn_{k\times n}^{H}(M, A, B, \Delta) \) and \( X = CYD + N \). Then \( X \sim Gmsn_{k\times n}^{H}(CMD + N, CA, BD, \Delta) \).
Proof. From assumption we have $Y = AZB + M$, and so $X = CAZBD + (CMD + N)$, i.e., $X \sim MSN_{k \times n}(CMD + N, CA, BD, \Delta)$.

Theorem 3.4.3. Let $x_1, x_2, \ldots, x_n$ be independent, where $x_i$ is distributed according to $gsn_{k}^{n}(0, \Sigma^{1/2}, \alpha)$. Then, \[ \sum_{j=1}^{n} x_j' \Sigma^{-1} x_j \sim \chi^{2}_{k \times n} \]

Proof. Let $y \sim N_k(\mu = 0, \Sigma)$. Then $y' \Sigma^{-1} y \sim \chi^{2}_{k}$, and $x_j' \Sigma^{-1} x_j$ and $y' \Sigma^{-1} y$ have the same distribution from Theorem 2.2.2. Moreover, $x_j' \Sigma^{-1} x_j$ are independent. Then the desired property is proved by the addition property of $\chi^{2}$ distribution.

3.5 Extensions

As in Chapter 2, an odd function, say $w(x)$, can be used to replace the term in the form $\alpha_{ji}x$ in the skewing function to give more flexible families of densities. As before we can take $w(x_{ji}) = \frac{\lambda_{ji}x}{\sqrt{1+\lambda_{ji}x^2}}$, to obtain a matrix variable form of the skew symmetric family introduced by Arellano-Valle et al. (17); take $w(x) = \alpha x + \beta x^3$, we obtain a matrix variate form of the skew symmetric family introduced by Ma and Genton (45); or take $w(x) = \text{sign}(x)|x|^{\alpha/2}\lambda(2/\alpha)^{1/2}$ to obtain a matrix variate form of the skew symmetric family introduced by DiCiccio and Monti (23).

Also note that, by the definition of matrix variate weighted distribution in Equation 3.6, the skewing function of the matrix variate skew normal density in Equation 3.8, i.e.,

\[ \prod_{j=1}^{k} \prod_{i=1}^{n} \Phi(\alpha_{ji}e_j' (A^{-1}(X - M)B^{-1})c_i), \]

can be replaced by

\[ \Phi_{k \times n}(\Gamma A^{-1}(Z - M)B^{-1}\Lambda) \]
where $\Gamma$ and $\Lambda$ are matrices of dimensions $k^*k$ and $n \times n^*$ correspondingly. In this case, the normalizing constant $2^{kn}$ will have to be changed to $E_Z(P(Y < \Gamma A^{-1}(Z - M)B^{-1}\Lambda|Z))$ for $Y \sim \phi_{k \times n}(0_{k \times n}, I_k, I_n)$ and $Z \sim \phi_{k \times n}(M, AA', BB')$. This extension to matrix variate skew normal density is similar to the matrix variable closed skew normal distribution \cite{8}. Also, if $k^* = k$, $n^* = n$ and both $\Gamma$ and $\Lambda$ are diagonal such that $\alpha_{ij} = \Gamma_{ii}\Lambda_{jj}$ we return to the family introduced in Definition 3.4.1.
CHAPTER 4

Estimation and Some Applications

Pewsey discusses some of the problems related to inference about the parameters of the skew normal distribution ([52]). The method of moments approach fails for the samples for which the sample skewness index is outside the admissibility range of the skew normal model. Also, the maximum likelihood estimator of the shape parameter $\alpha$ may take infinite values with positive probability. To deal with this problem certain methods have been proposed: Azzalini and Capitanio recommends that maximization of the log-likelihood function be stopped when the likelihood reaches a value not significantly lower than the maximum ([12]). Sartori uses bias prevention of maximum likelihood estimates for scalar skew normal and skew-$t$ distributions and obtains finite valued estimators ([57]). Bayesian estimation of the shape parameter is studied by Liseo and Loperfido ([44]). Minimum $\chi^2$ estimation method and an asymptotically equivalent maximum likelihood method are proposed by Monti ([46]). Chen and Gupta discuss goodness of fit procedures for the skew normal distribution ([33]).

Maximum products of spacings (MPS) estimation is independently introduced by Cheng and Amin ([12]), and Ranneby ([53]). It is a general method for estimating parameters in univariate continuous distributions and is known to give consistent and asymptotically efficient estimates under general conditions. In this chapter, we consider
estimation of the parameters of $sn_k(\alpha, \mu, \Sigma^{1/2})$ distribution introduced in Chapter 2, using the MPS procedure. We also show that MPS estimation procedure can be extended to give a class of bounded statistical model selection criteria which is suitable even for cases where Akaike’s and other likelihood based model selection criteria do not exist.

4.1 Maximum products of spacings (MPS) estimation

Consider the common statistical inference problem of estimating a parameter $\theta$ from observed data $x_1, \ldots, x_n$, where the $x_i$’s are mutually independent observations from a distribution function $F_{\theta_0}(x)$. The distribution function $F_{\theta}(x)$ is assumed to belong to a family $F = \{F_{\theta}(x) : \theta \in \Theta\}$ of mutually different distribution functions, where the parameter $\theta$ may be vector-valued.

Let $x_{(1)}, x_{(2)}, \ldots, x_{(n)}$ be the ordered random sample from $F_{\theta_0}(x)$. MLE method arises from maximization of the log product of density function,

$$l(\theta, X) = \frac{1}{n} \sum_{i=1}^{n} \log f_{\theta}(x_{(i)}) = \frac{1}{n} \sum_{i=1}^{n} l_i(\theta).\quad (4.1)$$

Let the maximum of this be denoted by $l(\hat{\theta})$. MPS method arises from maximization of the log product of spacings,

$$S(\theta; X) = \frac{1}{n+1} \sum_{i=1}^{n+1} s_i(\theta) = \frac{1}{n+1} \sum_{i=1}^{n+1} \log(F_{\theta}(x_{(i)}) - F_{\theta}(x_{(i-1)})),\quad (4.2)$$

with respect to $\theta$. It is assumed that $F_{\theta}(x_0) = 0$ and $F_{\theta}(x_{(n+1)}) = 1$. Let the maximum of this function be denoted by $S(\hat{\theta})$. Under very general conditions, Cheng and Amin (18) prove that the estimators produced by it have the desirable properties of being
consistent, asymptotically normal, and asymptotically efficient since ML method and MPS method are asymptotically equivalent. For the details of this proof and the necessary conditions see ([18]).

An intuitive comparison of the log likelihood with the log products of spacings is helpful in understanding the similarities and differences. By the mean value theorem,

\[ \log(s(i)) = \log\left( \int_{x(i-1)}^{x(i)} f_{\theta}(x) \, dx \right) = l_i(\theta) + \log(x(i) - x(i-1)) + R(x(i-1), x(i), \theta). \quad (4.3) \]

for each \( i \). Here, \( R(x(i-1), x(i), \theta) \) is essentially of \( O(x(i) - x(i-1)) \). Now, let the notation \( y = O_p(g(n)) \) mean that, given \( \epsilon > 0 \), \( K \) and \( N \) can be found, depending possibly on \( \theta_0 \), such that \( P(|y| < Kg(n)) > 1 - \epsilon \), for all \( n > N \). Similarly \( y = o_p(g(n)) \) mean that, given \( \epsilon > 0, \delta > 0 \), \( N \) can be found, depending possibly on \( \theta_0 \), such that \( P(|y| < \delta g(n)) > 1 - \epsilon \), for all \( n > N \). Under standard regulatory assumptions, the following result holds ([22]). Given any \( \epsilon > 0 \), there is a \( K = K(\epsilon, \theta_0) \) and \( N \) such that \( \sup(x(i) - x(i-1)) < K \frac{\log(n)}{n} \) with probability \( 1 - \epsilon \) for all \( n \geq N \). Therefore, the difference between \( l(\theta, X) \) and \( S(\theta, X) \) is approximately \( \sum_{i=1}^{n+1} \log(x(i) - x(i-1))/(n+1) \) which is \( O_p(\frac{\log(n)}{n}) \). Consequently, MPS and ML estimation are asymptotically equal and have the same asymptotic sufficiency, efficiency, and consistency properties. However \( S(\theta; X) \) is always bounded above by \( \log(\frac{1}{n+1}) \), MPS approach gives consistent estimators even when MLE fails.

The following theorem is proved under mild conditions in Cheng and Stephens ([19]), it justifies the use of MPS approach in its relation to the likelihood and Moran Statistic.

**Theorem 4.1.1.** Let \( \tilde{\theta} \) be the MPS estimator of \( \theta \). Then the MPS estimator, \( \tilde{\theta} \), exists in probability, and the following holds:

1. \( n^{1/2}(\tilde{\theta} - \theta_0) \sim n(0, -E\left( (\frac{\partial^2 l(\theta)}{\partial \theta^2})^{-1} |_{\theta_0} \right)) \) asymptotically.
2. If $\hat{\theta}$ is the likelihood estimator, then

$$\hat{\theta} - \theta_0 = o_p(n^{-1/2}).$$

3. Also

$$S(\hat{\theta}) = S(\theta_0) + \frac{1}{2n} Q + o_p(n^{-1}),$$

where $Q$ is distributed as $\chi^2_k$.

ML method and MPS method can be seen as objective functions that are approximations to Kullback-Leibler divergence ([24]). Ranneby ([53]) gives approximations of information measures other than the Kullback-Leibler information. In Ranneby & Ekström ([54]) a class of estimation methods, called generalized maximum spacing (GMPS) methods, is derived from approximations based on simple spacings of so-called $\varphi$-divergences. Ghosh & Jammalamadaka ([28]) show in a simulation study that GMPS methods other than the MPS method can perform better in terms of mean square error. Ekstrom ([24]) gives strong consistency theorems for GMPS estimators.

There are two disadvantages to MPS estimation: MPS procedure fails if there are ties in the sample. Although this will happen with probability zero for continuous densities, when data is grouped or rounded this may become an issue. Titterington ([59]) and Ekström & Ranneby([54]) give methods to deal with ties in data using higher order spacings. The second disadvantage is that optimization related to MPS approach usually requires computational methods for solution.
4.2 A Bounded Information Criterion for Statistical Model Selection

A fundamental problem in statistical analysis is about the choice of an appropriate model. The area of statistical model identification, model evaluation, or model selection deals with choosing the best approximating model among a set of competing models to describe a given data set. In a series of papers, Akaike ([2] and [3]) develops the field of statistical data modeling from the principle of Boltzmann’s ([15]) generalized entropy or Kullback-Leibler ([42]) divergence. As a measure of the goodness of fit of an estimated statistical model among several competing models, Akaike proposes the so-called Akaike information criterion (AIC). AIC provides a trade off between precision and complexity of a model, given a data set competing models may be ranked according to their AIC, the model with smallest AIC is considered the best approximating model. The Kullback-Leibler divergence between the true density $f_{\theta_0}(x)$ and its estimate $f_\theta(x)$ is given by

$$KLD(\theta_0, \theta) = \int f_{\theta_0}(x) \log(\frac{f_{\theta_0}(x)}{f_\theta(x)}) dx.$$  \hspace{1cm} (4.4)

The most obvious estimator of (4.4) is given by

$$\hat{KLD}(\theta_0, \theta) = \frac{1}{n} \sum_{i=1}^{n} \log(\frac{f_{\theta_0}(x(i))}{f_\theta(x(i)))}.$$ \hspace{1cm} (4.5)

Minimization of the expression in (4.5) with respect to $\theta$ is equivalent to maximization of the log likelihood, in Equation 4.6. The maximum log likelihood method can thus be used to estimate the values of parameters. However, it cannot be used to compare different models without some corrections because it is biased. An unbiased estimate of $-2$ times the mean expected maximum value of log likelihood yields the famous
AIC:

\[ AIC = -2l(\hat{\theta}) + 2k/n. \] (4.6)

For small sample sizes, the penalty term of AIC, i.e. \( 2k/n \), is usually not sufficient and may be replaced in search for improvement. For example the Bayes information criterion (BIC) of Schwarz ([58]) is obtained for \( 2k \log(n)/n \), Hannan-Quinn ([38]) information criterion is obtained for \( 2k \log(\log(n))/n \). Many other model selection criteria are obtained by similar adjustments to the penalty term. However, all these different criteria share the same problem: There are many cases for which the likelihood function is unbounded as a function of the parameters (for example see ([21])), therefore in such situations likelihood based inference and model selection is not appropriate. Next, we develop a model selection criterion which overcomes the unbounded likelihood problem based on the MPS estimation.

In Akaike ([2]), AIC is recommended as an approximation to the Kullback-Leibler divergence between the true density and its estimate ([2]). The connection of \( S(\theta, X) \) to the Kullback-Leibler divergence and minimum information principle is studied by Ranneby [53] and Lind ([43]). Ranneby introduces MPS from an approximation to the Kullback-Leibler divergence [53]. Under mild assumptions the following holds

\[ P(S(\hat{\theta}) < \sup_{\theta}(-\gamma - KLD(\theta, \theta) + \epsilon)) \to 1 \]

as \( n \to \infty \) for every \( \epsilon > 0 \) where \( \gamma \approx 0.57722 \) is the Euler’s constant.

In Cheng and Stephens ([19]), the relation of MPS method to Moran’s goodness of fit statistic ([17]) is studied. Moran’s statistics is the negative of \( S(\theta, X) \), and its distribution evaluate at the true value of the parameter does not depend on the particular distribution function it is calculated for. The percentage points of Moran’s statistic are given by Cheng and Thornton ([20]).
Cheng and Stephens extend the Moran’s test ([48]) to the situation where parameters have to be estimated from sample data. Under mild assumptions, when \( \hat{\theta} \) is an efficient estimate of \( \theta \) (this may be the maximum likelihood estimate, or the value, \( \theta \), which minimizes \( S(\theta; X) \)) then asymptotically the following holds:

\[
S(\hat{\theta}) = S(\theta_0; X) + \frac{1}{2n} Q + o_p(n^{-1})
\]

where \( Q \) is a \( \chi^2_k \) random variable with expected value \( k \). As \( n \to \infty \) expected value of the statistic \( -S(\hat{\theta}) + \frac{k}{2n} \) does not depend on the particular distribution function it is calculated for.

Because of the above reasons, we propose the statistic

\[
MPSIC = -S(\hat{\theta}) + \frac{k}{2n}
\]

as a model selection criterion and we refer to it as the maximum products of spacings information criterion (MPSIC) here on. Note that the penalty term in \( MPSIC \) can be replaced with \( \frac{k \log(n)}{2n} \) or \( \frac{k \log(\log(n))}{2n} \) to obtain other criteria of this kind, let’s call these variants \( MPSIC_2 \) and \( MPSIC_3 \). A

In Section 4.1, we have argued that \( S(\theta_0; X) \) should be a close to the log likelihood in 4.1 when the latter exists; however, the main difference between \( S(\theta_0; X) \) and the log likelihood in 4.1 is the term \( \sum_{i=1}^{n+1} \log(x_{(i)} - x_{(i-1)})/(n+1) \) which depends on the unknown true parameter value \( \theta_0 \). Because of this term the distribution of the log likelihood depends on the unknown true parameter value and therefore is not suitable for model identification purposes.

As mentioned earlier in skew normal model likelihood function is often unbounded. Therefore, estimation in skew normal model leaves a lot of room for testing MPSIC as a model selection criterion. In Table 4.1 we display the results of the following
experiment: We generate \( n \) observations from \( sn_1(\alpha,0,1) \), and then we compare the fit of this model to standard normal and half normal models with \( MPSC \), \( MPSC_2 \), and \( MPSC_3 \). We record the percentage of times the skew normal model is selected by these model selection criterion.

<table>
<thead>
<tr>
<th>( n )</th>
<th>( \alpha )</th>
<th>MPSIC</th>
<th>MPSIC2</th>
<th>MPSIC3</th>
</tr>
</thead>
<tbody>
<tr>
<td>20</td>
<td>0</td>
<td>0.272</td>
<td>0.056</td>
<td>0.251</td>
</tr>
<tr>
<td>50</td>
<td>0</td>
<td>0.265</td>
<td>0.030</td>
<td>0.190</td>
</tr>
<tr>
<td>100</td>
<td>0</td>
<td>0.294</td>
<td>0.030</td>
<td>0.182</td>
</tr>
<tr>
<td>20</td>
<td>2</td>
<td>0.961</td>
<td>0.953</td>
<td>0.961</td>
</tr>
<tr>
<td>50</td>
<td>2</td>
<td>1.000</td>
<td>1.000</td>
<td>1.000</td>
</tr>
<tr>
<td>100</td>
<td>2</td>
<td>1.000</td>
<td>1.000</td>
<td>1.000</td>
</tr>
<tr>
<td>20</td>
<td>5</td>
<td>0.806</td>
<td>0.777</td>
<td>0.803</td>
</tr>
<tr>
<td>50</td>
<td>5</td>
<td>0.960</td>
<td>0.959</td>
<td>0.960</td>
</tr>
<tr>
<td>100</td>
<td>5</td>
<td>0.999</td>
<td>0.999</td>
<td>0.999</td>
</tr>
<tr>
<td>20</td>
<td>10</td>
<td>0.536</td>
<td>0.483</td>
<td>0.529</td>
</tr>
<tr>
<td>50</td>
<td>10</td>
<td>0.842</td>
<td>0.822</td>
<td>0.835</td>
</tr>
<tr>
<td>100</td>
<td>10</td>
<td>0.972</td>
<td>0.966</td>
<td>0.972</td>
</tr>
</tbody>
</table>

Table 4.1: \( MPSIC \), \( MPSIC_2 \), and \( MPSIC_3 \) are compared for samples of size 20, 50 and 100 for different levels of \( \alpha \).

### 4.3 Estimation of parameters of \( sn_k(\alpha, \mu, \Sigma) \)

We need the following result in this section: Let \( x_{(1)}, x_{(2)}, \ldots, x_{(n)} \) be the ordered sample from \( sn_1(\alpha,0,1) \), and let the cdf of this density be denoted as \( F_\alpha(\cdot) \). Then

\[
\frac{\partial}{\partial \alpha} S(\alpha; X) = -\frac{\sqrt{2 \pi}}{2} \frac{\phi\left(\frac{x_{(1)}}{(\alpha^2+1)^{1/2}}\right)}{F_\alpha(x_{(1)})(\alpha^2 + 1)} \\
+ \frac{\sqrt{2 \pi}}{2} \frac{\phi\left(\frac{x_{(n)}}{(\alpha^2+1)^{1/2}}\right)}{(1 - F_\alpha(x_{(n)}))(\alpha^2 + 1)} \\
+ \sum_{i=2}^{n} \frac{\sqrt{2 \pi}}{2} \left(\phi\left(\frac{x_{(i-1)}}{(\alpha^2+1)^{1/2}}\right) - \phi\left(\frac{x_{(i)}}{(\alpha^2+1)^{1/2}}\right)\right) \\
\cdot \left(F_\alpha(x_{(i)}) - F_\alpha(x_{(i-1)})\right)(\alpha^2 + 1). \tag{4.9}
\]
4.3.1 Case 1: MPS estimation of \( sn_1(\alpha, 0, 1) \)

Assume \( X \sim sn_{k \times n}(\mu 1_k, \Sigma^{1/2}_c I_k, \alpha 1_k) \). When \( \mu, \Sigma^{1/2}_c \) are known, we can write \( z_i = \Sigma^{-1/2}_c (x_i - \mu) \) and estimate components \( \alpha_j \) of \( \alpha \) by using the corresponding i.i.d. \( sn_1(\alpha_j, 0, 1) \) random variables \( z_{j1}, z_{j2}, \ldots, z_{jn} \).

**Example 4.3.1.** The "frontier data" 50 observations generated by a \( sn_1(0, 1, \alpha = 5) \) distribution reported in Azzalini and Capitino ([12]) is an example for which the usual estimation methods like method of moments or maximum likelihood fail. Figure 4.1 gives the histogram, kernel density estimate, density curves for \( sn_1(0, 1, \alpha = 5) \) and \( sn_1(0, 1, \hat{\alpha} = 4.9632) \) distributions. For the frontier data, the Bayesian estimates recommended by Liseo and Loperfido gives \( \hat{\alpha} = 2.12 \) ([47]), the estimate through bias prevention method by Sartori gives \( \hat{\alpha} = 6.243 \) ([47]). Moreover the MPSIC for skew normal model is 4.3999. The MPSIC for the half normal model is 5.0495. Therefore, MPSIC identifies the correct model.

For any value shape parameter \( \alpha \), the absolute value of the skew normal variable has half normal distribution. If we take the absolute values of the frontier data, the appropriate model should be the half normal model instead of a skew normal model with large shape parameter. This fact is supported by MPSIC: Estimate of \( \alpha \) by MPS method is 38.3906, and the MPSIC for this model is 4.136. The MPSIC for the half normal model is 4.4057.

The following Table 4.2 gives summary statistics for the distribution of MPS estimator for different values of \( n \) and \( \alpha \) obtained from 5000 simulations. These results can be compared with the simulation results for the ML estimator in Table 4.2.

Although MPS estimation provides an improvement on ML estimation, highly skew distribution of the estimator, as seen in Figure 4.2, points to the tendency in some samples to give extremely large estimates for the shape parameter. Since \( \alpha \to \pm \infty \) the density \( SN_1(y, \alpha) \) approaches to the half normal density, the question whether skew
Figure 4.1: The histogram, kernel density estimate (blue), $sn_1(0,1,\alpha = 5)$ curve (red) and $sn_1(0,1,\hat{\alpha} = 4.9681)$ curve (black) for the "frontier data".

<table>
<thead>
<tr>
<th>n</th>
<th>$\alpha$</th>
<th>Mean ($\hat{\alpha}$)</th>
<th>std($\hat{\alpha}$)</th>
<th>Median($\hat{\alpha}$)</th>
<th>MAD($\hat{\alpha}$)</th>
</tr>
</thead>
<tbody>
<tr>
<td>20</td>
<td>0</td>
<td>0.0028</td>
<td>0.2702</td>
<td>0.0045</td>
<td>0.1578</td>
</tr>
<tr>
<td>50</td>
<td>0</td>
<td>0.0040</td>
<td>0.1572</td>
<td>0.0094</td>
<td>0.1048</td>
</tr>
<tr>
<td>100</td>
<td>0</td>
<td>0.0057</td>
<td>0.1193</td>
<td>0.0065</td>
<td>0.0849</td>
</tr>
<tr>
<td>20</td>
<td>2</td>
<td>2.9409</td>
<td>7.9874</td>
<td>1.6648</td>
<td>0.4516</td>
</tr>
<tr>
<td>50</td>
<td>2</td>
<td>1.9662</td>
<td>1.1007</td>
<td>1.8120</td>
<td>0.2981</td>
</tr>
<tr>
<td>100</td>
<td>2</td>
<td>1.9363</td>
<td>0.3932</td>
<td>1.8913</td>
<td>0.2446</td>
</tr>
<tr>
<td>20</td>
<td>5</td>
<td>10.2334</td>
<td>19.1683</td>
<td>3.6313</td>
<td>1.1751</td>
</tr>
<tr>
<td>50</td>
<td>5</td>
<td>5.5383</td>
<td>6.7330</td>
<td>4.2047</td>
<td>0.9611</td>
</tr>
<tr>
<td>100</td>
<td>5</td>
<td>4.7936</td>
<td>4.1561</td>
<td>4.4462</td>
<td>0.7202</td>
</tr>
<tr>
<td>20</td>
<td>10</td>
<td>21.5252</td>
<td>29.5315</td>
<td>6.2980</td>
<td>2.6961</td>
</tr>
<tr>
<td>50</td>
<td>10</td>
<td>11.4908</td>
<td>15.1492</td>
<td>7.3934</td>
<td>1.9900</td>
</tr>
<tr>
<td>100</td>
<td>10</td>
<td>10.0525</td>
<td>7.0268</td>
<td>8.3873</td>
<td>1.8476</td>
</tr>
</tbody>
</table>

Table 4.2: MPS estimation of $\alpha$: Mean, standard deviation, median and mean absolute deviation of $\hat{\alpha}$ from 5000 simulations.
normal model is appropriate for the data arises. When \( \hat{\alpha} \) is large, perhaps it is better to use the half normal model with one less parameter for sake of model parsimony. This can always be checked by comparing half normal model and skew normal model by one of the model selection criteria, \( MPSIC \), \( MPSIC_2 \), or \( MPSIC_3 \).

**4.3.2 Case 2: Estimation of \( \mu, \Sigma_{c}^{1/2} \) given \( \alpha \)**

For \( X \sim sn_{k \times n}(\mu 1_k, \Sigma_{c}^{1/2}, I_k, \alpha 1_k^1) \), when \( \alpha \) is known, we can use method of moments to estimate the parameters \( \mu, \Sigma_{c}^{1/2} \). Note that

\[
E(x) = \sqrt{2 \pi} \Sigma_{c}^{1/2} \left( \begin{array}{c}
\frac{\alpha_1}{\sqrt{1+\alpha_1^2}} \\
\frac{\alpha_2}{\sqrt{1+\alpha_2^2}} \\
\vdots \\
\frac{\alpha_k}{\sqrt{1+\alpha_k^2}}
\end{array} \right) + \mu = \Sigma_{c}^{1/2} \mu_0(\alpha) + \mu,
\]

\[
Cov(x) = \Sigma_{c}^{1/2}(I_k - \frac{2}{\pi} diag(\frac{\alpha_1^2}{1+\alpha_1^2}, \frac{\alpha_2^2}{1+\alpha_2^2}, \ldots, \frac{\alpha_k^2}{1+\alpha_k^2})) \Sigma_{c}^{1/2} = \Sigma_{c}^{1/2} \Sigma_0(\alpha) \Sigma_{c}^{1/2}.
\]
Figure 4.2: Histogram of 5000 simulated values for $\hat{\alpha}$ by MPS. The MPS estimator is highly skewed suggesting that half normal model might be more suitable for cases where extreme values of $\hat{\alpha}$ are observed.
Therefore, given $\alpha$, we can estimate $\Sigma$ and $\mu$ by solving the moment equations

$$S = (1/n) \sum_{j=1}^{n} (x_j - \bar{x})(x_j - \bar{x})' = \Sigma_c^{1/2} \Sigma_0(\alpha) \Sigma_c^{1/2}$$

(4.10)

$$\bar{x} = (1/n) \sum_{j=1}^{n} x_j = \Sigma_c^{1/2} \mu_0(\alpha) + \mu.$$  

(4.11)

Solution to the above equations are

$$\hat{\Sigma}_c^{1/2} = S_c^{1/2}(I_k - \frac{2}{\pi} \text{diag}(\frac{\alpha_1^2}{1 + \alpha_1^2}, \frac{\alpha_2^2}{1 + \alpha_2^2}, \ldots, \frac{\alpha_k^2}{1 + \alpha_k^2}))^{-1/2},$$

(4.12)

$$\hat{\mu} = \bar{x} - S_c^{1/2}(I_k - \text{diag}(\mu(\alpha_1)^2, \mu(\alpha_2)^2, \ldots, \mu(\alpha_k)^2))^{-1/2} \begin{pmatrix} \frac{\alpha_1}{\sqrt{1+\alpha_1^2}} \\ \frac{\alpha_2}{\sqrt{1+\alpha_2^2}} \\ \vdots \\ \frac{\alpha_k}{\sqrt{1+\alpha_k^2}} \end{pmatrix}.$$  

(4.13)

Note that $(\hat{\Sigma}_c^{1/2}, \hat{\mu})$ is an equivariant estimator of $(\Sigma_c^{1/2}, \mu)$ under the group affine transformations obtained by scaling with lower triangular matrix with positive diagonal elements.

4.3.3 Case 3: Simultaneous estimation of $\mu$, $\Sigma_c^{1/2}$ and $\alpha$.

Given an initial guess of $\alpha$, the following estimation procedure gives reasonably well results.

(a) Start with an initial estimate for $\alpha$, say $\alpha_0$.

(b) Given an estimate of $\alpha$, estimate $\Sigma_c^{1/2}$ and $\mu$ by [4.12] and [4.13]

(c) Given an estimate of $\mu$ and $\Sigma$, estimate $\alpha$ by MPS.
(d) Repeat steps 2-3 until convergence.

First, we apply this technique to univariate data.

**Example 4.3.2.** Bolfarine et. al. ([27]), consider the 1150 height measurements at 1 millisecond intervals along the drum of a roller (i.e. parallel to the axis of the roller). Data is a part of a study on surface roughness of the rollers. Since there are ties in the sample before analyzing the data we have adjusted the data for ties. We apply the procedure described above, Figure 4.3 summarizes the results.

![Histogram, kernel density estimate, and estimated density curve](image)

Figure 4.3: The histogram, kernel density estimate (blue), and estimated density curve $sn_1(\mu = 4.2865, \sigma^{1/2} = 0.9938, \alpha = -2.9928)$ (black) for 1150 height measurements.
Table 4.4 gives summary statistics for the distribution of estimators of $\alpha$, $\mu = 0$ and $\sigma = 1$ for different of $n$ and $\alpha$.

<table>
<thead>
<tr>
<th>n</th>
<th>$\alpha$</th>
<th>mean $\hat{\alpha}$</th>
<th>std $\hat{\alpha}$</th>
<th>median $\hat{\alpha}$</th>
<th>MAD $\hat{\alpha}$</th>
<th>mean $\hat{\mu}$</th>
<th>std $\hat{\mu}$</th>
<th>mean $\hat{\sigma}$</th>
<th>std $\hat{\sigma}$</th>
</tr>
</thead>
<tbody>
<tr>
<td>30</td>
<td>1.5</td>
<td>2.6013</td>
<td>2.1136</td>
<td>1.9277</td>
<td>1.4755</td>
<td>-0.0722</td>
<td>0.2303</td>
<td>1.0560</td>
<td>0.2003</td>
</tr>
<tr>
<td>50</td>
<td>1.5</td>
<td>2.0914</td>
<td>1.2355</td>
<td>1.7920</td>
<td>0.8614</td>
<td>-0.0536</td>
<td>0.1971</td>
<td>1.0432</td>
<td>0.1634</td>
</tr>
<tr>
<td>100</td>
<td>1.5</td>
<td>1.8096</td>
<td>0.6817</td>
<td>1.6584</td>
<td>0.5292</td>
<td>-0.0372</td>
<td>0.1443</td>
<td>1.0228</td>
<td>0.1157</td>
</tr>
<tr>
<td>30</td>
<td>3.0</td>
<td>4.0288</td>
<td>2.7718</td>
<td>3.2239</td>
<td>2.0730</td>
<td>-0.0229</td>
<td>0.1848</td>
<td>0.9836</td>
<td>0.1840</td>
</tr>
<tr>
<td>50</td>
<td>3.0</td>
<td>3.7638</td>
<td>2.0545</td>
<td>3.2340</td>
<td>1.5469</td>
<td>0.0072</td>
<td>0.1426</td>
<td>0.9934</td>
<td>0.1509</td>
</tr>
<tr>
<td>100</td>
<td>3.0</td>
<td>3.5131</td>
<td>1.5431</td>
<td>3.1676</td>
<td>1.1380</td>
<td>0.0012</td>
<td>0.1090</td>
<td>0.9981</td>
<td>0.1090</td>
</tr>
</tbody>
</table>

Table 4.4: Table above gives summary statistics for the distribution of estimators of $\alpha$, $\mu = 0$ and $\sigma = 1$ for different of $n$ and $\alpha$.

Example 4.3.3. We generate 100 observations from $\text{sn}_2(\mu = (0,0)', \Sigma_{1/2} = \begin{pmatrix} 1 & 0 \\ .5 & 2 \end{pmatrix}$, $\alpha = (5,-3)'$) distribution. We estimate the parameters with the procedure described above with initial estimate for $\alpha$, i.e. $\alpha_0 = (10,-10)'$. Figure 4.4 gives the the scatter plot, and the contour plots of the true, and the estimated densities.

Example 4.3.4. Australian Athletes Data: This data is collected on 202 athletes and mentioned in Azzalini (\[11\]). We take the variables LBM and BMI and fit a bivariate skew normal model with the procedure described above with initial estimate for $\alpha$, i.e. $\alpha_0 = (10,10)'$. Figure 4.5 gives the the scatter plot, and the contour plot the estimated density.

Example 4.3.5. Australian Athletes Data 2: This data is collected on 202 athletes and mentioned in Azzalini (\[11\]). We take the variables LBM and SSF and fit a bivariate skew normal model with the procedure described above with initial estimate for $\alpha$, i.e. $\alpha_0 = (10,10)'$. Figure 4.6 gives the the scatter plot, and the contour plot the estimated density.
Figure 4.4: The scatter plot for 100 generated observations (black dots) and, contours of $sn_2(\mu = (0,0)', \Sigma_{cc}^{1/2} = \begin{pmatrix} 1 & 0 \\ 0.5 & 2 \end{pmatrix}, \alpha = (5,-3)'$) (blue), and estimated density curve $sn_2(\mu = (0.0795, 0.0579)', \Sigma_{cc}^{1/2} = \begin{pmatrix} 0.9404 & 0 \\ 0.5845 & 1.1092 \end{pmatrix}, \alpha = (5.6825, -3.6566)'$) (red).
Figure 4.5: The scatter plot for LBM and BMI in Australian Athletes Data, contours of the estimated density $sn_2(\mu = (19.7807, 48.8092)', \Sigma_1^{1/2} = \begin{pmatrix} 4.2760 & 0 \\ 13.9306 & 10.7933 \end{pmatrix}, \alpha = (2.5429, 0.8070)'$ (red).

Figure 4.6: The scatter plot for LBM and SSF in Australian Athletes Data, contours of the estimated density $sn_2(\mu = (19.7807, 17.3347)', \Sigma_1^{1/2} = \begin{pmatrix} 4.2760 & 0 \\ 15.6127 & 50.5834 \end{pmatrix}, \alpha = (2.5429, 8.6482)'$ (red).
Table 4.5: The simulation results about \( \hat{\mu} \) and \( \hat{\alpha} \) in estimation of \( s\bar{n}_2(\mu = (1, 1)\), \( \Sigma_1^{1/2} = \begin{pmatrix} 1 & 0 \\ 0.5 & 1 \end{pmatrix} \), \( \alpha = (1, 1)' \).
Table 4.6: The simulation results about $\hat{\Sigma}_{1/2}$ in estimation of $sn_2(\mu = (1, 1)', \Sigma_{c}^{1/2} = \begin{pmatrix} 1 & 0 \\ .5 & 1 \end{pmatrix}, \alpha = (1, 1)')$. 

<table>
<thead>
<tr>
<th>$n$</th>
<th>mean</th>
<th>cov</th>
<th>mean</th>
<th>cov</th>
<th>mean</th>
<th>cov</th>
<th>mean</th>
<th>cov</th>
</tr>
</thead>
<tbody>
<tr>
<td>10</td>
<td>1.1266</td>
<td>0.1179</td>
<td>0.046</td>
<td>-0.0015</td>
<td>0.5627</td>
<td>0.046</td>
<td>0.1867</td>
<td>0.0108</td>
</tr>
<tr>
<td>20</td>
<td>1.0797</td>
<td>0.0656</td>
<td>0.0349</td>
<td>-0.0042</td>
<td>0.5229</td>
<td>0.0349</td>
<td>0.0841</td>
<td>0.0024</td>
</tr>
<tr>
<td>30</td>
<td>1.0582</td>
<td>0.044</td>
<td>0.0226</td>
<td>-0.0002</td>
<td>0.5038</td>
<td>0.0226</td>
<td>0.0514</td>
<td>0.0018</td>
</tr>
<tr>
<td>50</td>
<td>1.0319</td>
<td>0.0293</td>
<td>0.0147</td>
<td>0.0029</td>
<td>0.5201</td>
<td>0.0147</td>
<td>0.03</td>
<td>0.0014</td>
</tr>
<tr>
<td>100</td>
<td>1.0268</td>
<td>0.0159</td>
<td>0.0078</td>
<td>0.0003</td>
<td>0.5168</td>
<td>0.0078</td>
<td>0.0147</td>
<td>0.0011</td>
</tr>
<tr>
<td>200</td>
<td>1.0179</td>
<td>0.0085</td>
<td>0.0039</td>
<td>-0.0003</td>
<td>0.5112</td>
<td>0.0039</td>
<td>0.0068</td>
<td>0.0002</td>
</tr>
</tbody>
</table>
BIBLIOGRAPHY


