A Generalized Multivariate Skew-Normal Distribution with Application to Prediction

By

Tareq Taleb Suleiman Alodat

Supervisor

Prof. Ziad Al-Rawi

rogram: Statistics

February 21, 2010

A Generalized Multivariate Skew-Normal Distribution

with Application to Prediction

By

Tareq Taleb Suleiman Alodat

B.Sc. (Mathematics), Irbid National University, Irbid, 2006

thesis submitted in partial fulfillment of the requirements for the degree of Master

f Science in the Department of Statistics, Yarmouk University, Irbid, Jordan

pproved by:

rof. Ziad Al Rawi...Z. R. Al-Rawi.........................**Chairman.**

rofessor of Statistics, Yarmouk University.

r. Abdullah A. Smadi..**Member.**

ssociate Professor of Statistics, Yarmouk University.

r. Mohammed Hassan Baker Al-Haj Ebrahem ....**Member.**

ssociate Professor of Statistics, Yarmouk University.

rof. Muhammad S. Abu-Salih ....**Member.**

rofessor of Statistics, Amman Arab University for Graduate studies.

Acknowledgments

Thanks to Allah for everything. Then I acknowledge offering my gratitude to the followings

1. The assistance of my family, especially my mother and my father and brothers
2. My Supervisor Prof. Ziad Al-Rawi for advice, patience, steadfast encouragement, invaluable suggestion, his continuous help during the course of my work on the subject of this thesis, so all my thanks to him.
3. The defense committee, Dr. Abdullah A. Smadi Dr. Mohammed Hassan Baker Al-Haj Ebrahem and Prof. Muhammad S. Abu-Salih for their useful remarks and suggestions, which helped me a lot in improving the final draft
4. Special thanks to my brother Dr. Mohammad Alodat for his emotional and financial assistance and support during my bachelor's and master's degrees study. I also extend my sincere thanks to him, in particular in the completion of this master's thesis.
5. All members in the Department of Statistics
6. My friends in the department of statistics, Yarmouk University for their support and encouragement.

Contents

5. Chapter Five

Generalization of Skew Gaussian Process for Regression

6. Chapter Six

Index Abbreviations

1. $\mathbf{A}_{m\times n}$ $(\mathbf{A})$: $m \times n$ matrix

2. $\boldsymbol{a}_{n\times 1}(\boldsymbol{a})$: $n \times 1$ vector

3. $\mathbf{A}'$: *the transpose of a matrix* **A**

4. $\|\boldsymbol{x}\|$: the norm or length of $\boldsymbol{x}$

5. $\mathbf{0}_{m\times n}$ $(\mathbf{0})$: the zero matrix of dimension $m \times n$

6. $\mathbf{I}_n$: the identity matrix of dimension n

7. $\mathbf{D} = diag(\lambda_1, \dots, \lambda_n)$: the diagonal matrix of diagonal elements $\lambda_1, \dots, \lambda_n$

8. $\mathbf{A} = \mathrm{circ}(a_1, a_2, \dots, a_n)$: the circulant matrix generated by $a_1, a_2, \dots, a_n$

9. $|\mathbf{A}|$: the determinant of **A**

10. $r(\mathbf{A})$: the rank of **A**

11. $\mathbb{R}^n$: the n –dimensional Euclidian space

12. $X(\boldsymbol{t})$: a random process indexed by $\boldsymbol{t}$

13. $E(\boldsymbol{X})$: the expectation of $\boldsymbol{X}$

14. $Cov(\boldsymbol{X})$: the covariance matrix of $\boldsymbol{X}$

15. $\varphi_n(\boldsymbol{x}; \boldsymbol{\mu}, \boldsymbol{\Sigma})$: the multivariate normal density

16. $\Phi(x)$: cumulative distribution function of the standard normal distribution.

Abstract

Alodat, Tareq Taleb. A Generalized Multivariate Skew-Normal Distribution with Application to Prediction. Master of Science Thesis, Department of Statistics, Yarmouk University, 2010 (Supervisor: Prof. Ziad Al Rawi).

In this thesis, we generalize the multivariate skew-normal distribution of Arnold and Beaver (2002). Also we study its statistical properties such as moment-generating function and closure under marginal and conditional distribution. Moreover, we show that the new distribution is closed under convolution with multivariate normal distribution.

The new distribution is used to define random process called the generalized skew Gaussian process. We study two prediction problems under the generalized skew Gaussian process. For these two prediction problems, we derive the best linear predictors. Finally, we propose two algorithms to generate random observations from the new distribution.

Keywords: Gaussian process for regression; Ordinary kriging; Skew-normal distribution.

الملخص

العودات, طارق طالب. التوزيع الطبيعي المعمم الملتوي متعدد المتغيرات مع تطبيقات في التنبؤ.

رسالة ماجستير في العلوم، قسم الاحصاء، جامعة اليرموك، 2010.

(المشرف: الاستاذ الدكتور زياد الراوي)

في هذه الرسالة قمنا بتعميم التوزيع الطبيعي الملتوي المتعدد ل ارنولد وبيفر (2002). ايضا قمنا بدراسة الخصائص الاحصائية لهذا التوزيع مثل الدالة المولدة للعزوم، الانغلاق تحت الجمع، التوزيع الشرطي والتوزيع الهامشي. كذلك قمنا باثبات ان هذا التوزيع مغلق تحت عملية الجمع مع التوزيع الطبيعي المتعدد. كما قمنا باستخدام التوزيع الجديد لتعريف عملية احتمالية جديدة سميناها العملية الاحتمالية الجاوسية الملتوية المعممة، حيث درسنا مسألتين للتنبؤ لهذا العملية واوجدنا لها افضل متنبىء خطي. واخيرا اقترحنا خوارزميتين لتوليد عينه عشوائية من التوزيع الجديد.

Chapter One

Introduction

Until the last two decades, the multivariate normal distribution was one of the most popular distributions in modeling and analyzing various statistical data sets. This popularity comes from its exciting properties such as symmetry, unimodality, closure under convolution as well as marginal and conditional distributions. In various applications of statistics in environmental, financial and biological problems, the collected data after show skewness. In such cases, the practitioners try to use some transformations on their data to achieve normality and hence the normal distribution is used to analyze the transformed data sets. These transformations are not welcomed due to the following reasons: (i) It is not easy to find a suitable transformation to achieve normality for several multivariate data sets, (ii) since transformations are usually performed component-wise (where normality of marginal's does not guarantee the joint normality), then the statistical problem might be not invariant under these transformations, which leads to biased estimates (iii) any transformation on the data will reduce the amount of information in the original data unless the transformation is a sufficient statistic and (iv) despite of the difficulty in interpreting the transformed data, the data skewness has its natural interpretation and hence could not be ignored (Buccianti and Pawllowsky-Glahn, 2005).

Through the last two decades, several skewed multivariate statistical distributions have been introduced which possess properties that have a skewness parameter and coincide with or close to the properties of the normal distribution (Gupta and Gonzalez-Fraias., et al, 2003). Azzalini (1985, 1986) was the first to open such new area in multivariate analysis. He started that area by introducing the univariate skew-normal distribution:

$$f_X(x) = 2\varphi(x)\Phi(\alpha x), \quad x, \alpha \in \mathbb{R},$$

where $\varphi(.)$ and $\Phi(.)$ are respectively, the probability density function (pdf) and the cumulative distribution function (CDF) of the univariate standard normal distribution. The parameter α is called the skewness parameter. (Azzalini and Dalla Valle, 1996) and (Azzalini and Capitanio, 1999) extended the Azzalini distribution to the following multivariate version

$$f_X(\boldsymbol{x}) = 2\varphi_n(\boldsymbol{x}; \boldsymbol{0}, \boldsymbol{\Sigma})\Phi(\boldsymbol{\alpha}'\boldsymbol{x}), \quad \boldsymbol{x}, \boldsymbol{\alpha} \in \mathbb{R}^n \qquad (1.1)$$

where $\varphi_n(\boldsymbol{x}; \boldsymbol{0}, \boldsymbol{\Sigma})$ is the n-dimensional pdf of multivariate normal distribution with zero mean vector and a covariance matrix $\boldsymbol{\Sigma}$, while $\boldsymbol{\alpha}'$ stands for the transpose of $\boldsymbol{\alpha}$ and $\mathbb{R}^n$ is the n-dimensional Euclidean space.

Other kind of skew-normal distributions can be defined via the idea of hidden truncation (Arnold and Beaver, 2002). To illustrate the idea of hidden truncation, let $X_1, \ldots, X_n$ and U be independent standard normal random variables and $\lambda_0 \in \mathbb{R}$, $\boldsymbol{\lambda} \in \mathbb{R}^n$. Then the conditional distribution of $\boldsymbol{X} = (X_1, \ldots, X_n)'$ given that $\lambda_0 + \boldsymbol{\lambda}'X > U$ is called the hidden truncation distribution which has the pdf

$$f_X(\boldsymbol{x}) = \frac{\varphi_n(\boldsymbol{x}; \boldsymbol{0}, \mathbf{I}_n)}{\Phi(\delta_0)}\,\Phi(\lambda_0 + \boldsymbol{\lambda}'\boldsymbol{x}), \qquad \boldsymbol{x} \in \mathbb{R}^n \qquad (1.2)$$

where $\delta_0 = \lambda_0/\sqrt{1+\lambda'\lambda}$. Putting $\alpha = 0$ in (1.1) or $\lambda = 0$ in (1.2), we see that the family of normal distributions is a sub family of these skew-normal families. A comprehensive survey about skew-normal distributions and their applications can be found in Genton (2004a).

Despite of its simplicity and tractability, Gaussian models can lead to unrealistic estimates when they are used to analyze skewed data. Such data sets arise in various scientific fields where the observations are scattered in a subset of the space. For example, in oceanography, the sea surface elevation departs from Gaussianity and symmetry as the depth of the water decreases or the sea severity increases (Machado, 2002). Hence, non-Gaussian distributions are needed to capture the skewness as well as the non-Gaussianity in the data. Moreover, if the data are measured with time or at a set of points scattered in the space, then non-Gaussian random processes or fields are needed to capture their skewness and the non-Gaussianity.

In the statistical literature, two important statistical techniques are used to study the prediction problem for such data: the ordinary kriging and the Gaussian process for regression (see Cressie, 1993; Rasmussen, 1996). Based on these two techniques, data points are assumed to be correlated and follow a Gaussian random process. Once the skewness is present in the data, then it should be a limitation of these two prediction problems. Recently, some skewed Gaussian processes and fields have been introduced to analyze skewed data (Kim and Mallick, 2004; Alodat and Aludaat, 2006; Allard and Naveau, 2007; Zhang and El-Sharaawi, 2009; Alodat and Al-Rawwash, 2009). Allard and

Naveau (2007) have used the multivariate closed skew-normal distribution of Domingues et al., (2003) to define the so called closed skew-normal random field. The closed skew-normal random field is not suitable to generalize the Gaussian process for regression, since the dimension of the closed skew-normal distribution grows up as the sample size increases. This leads to a computations complexity. On the other hand, several skew Gaussian processes have been introduced in the literatures, but all of them have some limitations due to loss of closure under both convolution and conditioning (Zhang and El-Sharaawi, 2009; Alodat and Aludaat, 2006; Alodat and Al-Rawwash, 2009). Hence these limitations do not allow us to extend the two prediction problems that we have mentioned in this introduction.

Taking these points in consideration, we are going, in this thesis, to extend the multivariate skew-normal distribution of Arnold and Beaver (2002) to make it amenable for extending the above two spatial prediction problems. Moreover, we use it to define a skew random process called the generalized skew Gaussian process. Also, we propose two algorithms to simulate data from the proposed distribution.

The rest of this thesis is organized as follows. In Chapter 2, we introduce the reader to two spatial prediction problems, the ordinary kriging and the Gaussian process for regression. In Chapter 3, we extend the distribution of Arnold and Beaver (2002) to the multivariate generalized skew normal distribution. Also, we derive its moments, covariance matrix and moment-generation function, we the show that it is closed under convolution with the multivariate Gaussian distribution. In Chapter 4, the notion of generalized skew Gaussian process is introduced and the best linear unbiased predictor for an observation at non sampled location is derived. An alternative to Gaussian process for regression, called the generalized skew Gaussian process for regression, is also introduced in Chapter 5. Finally, we state our conclusions and future work in Chapter 6.

Chapter Two

Random Processes and Spatial Prediction

In this chapter, we introduce the reader to two spatial predictions namely, the ordinary kriging and the Gaussian process for regression. For more information about these two techniques see Cressie (1993) and Rasmussen (1996).

2.1 Random Processes

We may define the random process as a collection of random variables $\{Y(t): t \in D \subset \mathbb{R}^d\}$ together with a collection of measures or distribution functions of the form $F_{t_1,\dots,t_n}$ on $\mathcal{B}(\mathbb{R}^d)$, for $t_1, \dots, t_n \in \mathbb{R}^d$, $n = 1, 2, 3, \dots$

$$F_{t_1,t_2,\dots,t_n}(B) = \mathrm{P}((Y(t_1), \dots, Y(t_n))' \in B),$$

for every Borel σ −field, $B \in \mathbb{R}^d$. The collection of all such measures or, equivalently, the corresponding distribution function is known as the family of finite dimensional distributions for the field $Y(t)$. For a given $\omega \in \Omega, \{Y(t, \omega): t \in D\}$ is a deterministic real-valued function defined on $\mathbb{R}^d$, which is a realization of the field $Y(t)$. Moreover, the set $\{(t, Y(t)): t \in \mathbb{R}^d\}$ is called the sample function or sample path of $Y(t)$.

Definition 1. (Adler and Taylor, 2007) The random process $Y(t)$ is said to be strictly homogeneous or stationary if for any $k + 1$ *points* $\tau, t_1, \dots, t_k \in \mathbb{R}^d$, the following condition on its finite dimensional distribution holds

$$P(Y(t_1) \leq y_1, \dots, Y(t_k) \leq y_k) = P(Y(t_1 + \tau) \leq y_1, \dots, Y(t_k + \tau) \leq y_k).$$

Definition 2. (Adler and Taylor, 2007) The random process $Y(\boldsymbol{t})$ is said to be isotropic if for any k points $\boldsymbol{t}_1, \ldots, \boldsymbol{t}_k \in D$, the following condition on its finite dimensional distribution holds

$P(Y(\boldsymbol{t}_1) \leq y_1, \ldots, Y(\boldsymbol{t}_k) \leq y_2) = P(Y(\boldsymbol{t}_1\mathbf{Q}) \leq y_1, \ldots, Y(\boldsymbol{t}_k\mathbf{Q}) \leq y_2), \forall\, \mathbf{Q} \in O(\mathbb{R}^d)$, the set of all orthogonal matrices on $\mathbb{R}^d$.

For every random process $Y(\boldsymbol{t})$, we define two functions

1. Mean function

$$m(\boldsymbol{t}) = E\big(Y(\boldsymbol{t})\big), \forall\, \boldsymbol{t} \in D$$

2. Covariance function

$$R(\boldsymbol{s}, \boldsymbol{t}) = E\{\big(Y(\boldsymbol{s}) - m(\boldsymbol{s})\big)\big(Y(\boldsymbol{t}) - m(\boldsymbol{t})\big)\} \ \forall\, \boldsymbol{s}, \boldsymbol{t} \in D$$

Definition 3. (Adler and Taylor, 2007) A real-valued function $Y(\boldsymbol{t}), \boldsymbol{t} \in D$ is said to be a Gaussian random process, if for every k points $\{\boldsymbol{t}_1, \ldots, \boldsymbol{t}_k\} \subset D$, the vector $\big(Y(\boldsymbol{t}_1), \ldots, Y(\boldsymbol{t}_k)\big)'$ has a multivariate Gaussian distribution, i.e., all finite dimensional distributions are multivariate Gaussian.

A Gaussian random Process (GRP) with covariance function $R(\boldsymbol{s}, \boldsymbol{t})$ is stationary or homogenous, if its mean and covariance functions satisfy the following:

$$m(\boldsymbol{t}) = c \ \ \forall\, \boldsymbol{t} \in D \text{ and } c \in \mathbb{R}$$

and

$$R(\boldsymbol{s}, \boldsymbol{t}) = K(\boldsymbol{s} - \boldsymbol{t}) \ \ \forall\, \boldsymbol{s}, \boldsymbol{t} \in D,$$

and it is isotropic if its covariance function depends only on the distance between $\boldsymbol{s}$ and $\boldsymbol{t}$ as follows:

$$R(\boldsymbol{s},\boldsymbol{t}) = K(\|\boldsymbol{s}-\boldsymbol{t}\|)\ \forall\ \boldsymbol{s},\boldsymbol{t} \in D.$$

where $\|.\|$ denotes the Euclidean norm.

Definition 4: (Adler and Taylor, 2007) A random process $Y(\boldsymbol{t})$ satisfying the conditions

a. $E\big(Y(\boldsymbol{t})\big) = \mu,\ \forall\ \boldsymbol{t} \in D$ and

b. $Cov\big(Y(\boldsymbol{t}_1), Y(\boldsymbol{t}_2)\big) = K(\boldsymbol{t}_1 - \boldsymbol{t}_2),\ \forall\ \boldsymbol{t}_1, \boldsymbol{t}_2 \in D.$

is called second–order stationary and the function $K(.)$ is called covariogram or (stationary covariance function).

2.2 Ordinary Kriging

In various fields of science such as agriculture, geostatistics and ecology, prediction is an important result. Data sets arising in these fields could be treated as realizations of random fields or processes. The ordinary Kriging method, which is given in Cressie (1993), can be used to find the Best Linear Unbiased Predictor (BLUP) under a certain condition. This method is described as follows.

In spatial data, a random process $Y(\boldsymbol{t}), \boldsymbol{t} \in D$ is observed at known spatial locations $\{\boldsymbol{t}_1, \ldots, \boldsymbol{t}_n\} \subset D$, i.e., $Y(\boldsymbol{t}_1), \ldots, Y(\boldsymbol{t}_n)$ and we are interested in predicting the value of $Y(\boldsymbol{t})$ at $\boldsymbol{t}_0 \in D$. The BLUP depends only on the second-order properties of the process and no full distributional assumptions are made. We are given n observations and we wish to map the process $Y(\boldsymbol{t})$ within a region D. This predictor is available under:

i. Model assumption: The process $Y(\boldsymbol{t})$ is second order stationary with unknown mean μ,

$$Y(\boldsymbol{t}_i) = \mu + \delta(\boldsymbol{t}_i), \qquad \boldsymbol{t}_i \in D, \mu \in \mathbb{R},$$

where $\delta(.)$ is a zero-mean, second order stationary process with covariance function $C(\boldsymbol{h}), \boldsymbol{h} \in \mathbb{R}^n$.

ii. Predictor assumption: The predictor $P(\boldsymbol{Y}; \boldsymbol{t}_0)$ is linear and satisfies

$$P(\boldsymbol{Y}; \boldsymbol{t}_0) = \sum_{i=1}^{n} \lambda_i Y(\boldsymbol{t}_i),$$

where $\sum_{i=1}^{n} \lambda_i = 1$.The condition that the coefficients of the linear predictor sum to 1 guarantees uniform unbaisedness, i.e.

$$E\big(P(\boldsymbol{Y}; \boldsymbol{t}_0)\big) = E\left(\sum_{i=1}^{n} \lambda_i Y(\boldsymbol{t}_i)\right) = \sum_{i=1}^{n} \lambda_i E\big(Y(\boldsymbol{t}_i)\big) = \mu.$$

An optimal ordering predictor usually refers to the squared-error loss, i.e., we need to minimize the mean-squared prediction error

$$\sigma_e^2(\boldsymbol{t}_0) = E\big(Y(\boldsymbol{t}_0) - P(\boldsymbol{Y}; \boldsymbol{t}_0)\big)^2,$$

with respect to the predictor coefficients. The ordinary kriging predictor is given as follows. If we denote

$\mathbf{1}'_n = (1, 1, \dots, 1)$, $k' = (K(t_0 - t_1), K(t_0 - t_2), \dots, K(t_0 - t_n))$, K is the $n \times n$ covariance matrix of $K_{ij} = K(t_i - t_j)$, and $\zeta' = (\zeta_1, \zeta_2, \dots, \zeta_n)$. Then the best linear predictor (ordinary kriging predictor) is given by

$$\hat{P}(Y; t_0) = \hat{\zeta}'Y,$$

where

$$\hat{\zeta} = \mathrm{K}^{-1}(k + \boldsymbol{m}), \qquad \boldsymbol{m} = m\mathbf{1}_n$$

and

$$m = \frac{1 - \mathbf{1}'_n \mathrm{K}^{-1} k}{\mathbf{1}'_n \mathrm{K}^{-1} \mathbf{1}_n}.$$

The minimized mean-square prediction error is called the kriging variance $\sigma_e^2(t_0)$, where

$$\sigma_e^2(t_0) = k'\mathrm{K}^{-1}k - \frac{(\mathbf{1}'_n \mathrm{K}^{-1} k - 1)^2}{\mathbf{1}'_n \mathrm{K}^{-1} \mathbf{1}_n}.$$

For more details see Cressie (1993).

2.3. Gaussian Process for Regression (GPR)

In any regression problem, we are given a sample $S = \{(x_1, y_1), \dots, (x_n, y_n)\}$ in which x_i denotes the i^{th} observation of x and y_i is the corresponding target or response value. Then a relation between x_i and y_i is given via

$$y_i = f(x_i) + \epsilon(x_i),$$

where f is a function which maps the input vector $\boldsymbol{x}_i$ to the true target value $f(\boldsymbol{x}_i)$. The value $f(\boldsymbol{x}_i)$ is measured as y_i and is corrupted by a noise $\epsilon(\boldsymbol{x}_i)$. The Gaussian Process for Regression (GPR) is a non-parametric model that assumes a Gaussian process prior of the function $f(\boldsymbol{x})$, i.e., for every set of inputs $\boldsymbol{x}_1, \dots, \boldsymbol{x}_n$,

$$\boldsymbol{f} = \left(f(\boldsymbol{x}_1), \dots, f(\boldsymbol{x}_n)\right)' \sim N_n(\boldsymbol{0}, \mathbf{K}),$$

where $\mathbf{K}$ is a covariance matrix whose ij^{th} element is given by a covariance function $K(\boldsymbol{x}_i, \boldsymbol{x}_j)$, also $\epsilon(\boldsymbol{x}_j)$ is a white noise which is independent of $\boldsymbol{f}$ such that $\left(\epsilon(\boldsymbol{x}_1), \dots, \epsilon(\boldsymbol{x}_n)\right)' \sim N_n(\boldsymbol{0},\ \tau^2 \mathbf{I}_n)$. The main objective in Gaussian process for regression problem is to predict the value of $f(\boldsymbol{x})$ at a new input say $\boldsymbol{x}^*$, i.e., to predict $f(\boldsymbol{x}^*)$ given the data S and $\boldsymbol{x}^*$.

O'Hagan (1978) was the first to propose a Gaussian process prior for regression. After the publication of Neal (1995), the GPR became a popular non-parametric model. In his Ph.D. thesis, Rasmussen (1996) showed that GPR can achieve a good prediction performance. O'Hagan has derived the predictive distribution of $f^* = f(\boldsymbol{x}^*)$ given $\boldsymbol{x}^*$ and S. If $P(f^*|\boldsymbol{x}^*, S)$ denotes the predictive distribution of f^* given $\boldsymbol{x}^*$ and S, then

$$(f^*|\boldsymbol{x}^*, S) \sim N\left(\boldsymbol{k}(\boldsymbol{x}^*)'(\mathbf{K} + \tau^2\mathbf{I}_n)^{-1}\boldsymbol{y}, K(\boldsymbol{x}^*, \boldsymbol{x}^*) - \boldsymbol{k}(\boldsymbol{x}^*)'(\mathbf{K} + \tau^2\mathbf{I}_n)^{-1}\boldsymbol{k}(\boldsymbol{x}^*)\right),$$

where $\boldsymbol{k}(\boldsymbol{x}^*) = \left(K(\boldsymbol{x}_1, \boldsymbol{x}^*), \dots, K(\boldsymbol{x}_n, \boldsymbol{x}^*)\right)'$, $\boldsymbol{y} = (y_1, \dots, y_n)'$ and $\mathbf{K} = \left(K(\boldsymbol{x}_i, \boldsymbol{x}_j)\right)_{i,j=1}^{n}$.

The mean $E(f^*|\boldsymbol{x}^*, S)$ serves as an estimate of the output function $f(\boldsymbol{x}^*)$ with an uncertainty $\sigma(\boldsymbol{x}^*)$, where

$$E(f^*|\boldsymbol{x}^*, S) = \boldsymbol{k}(\boldsymbol{x}^*)'(\mathrm{K} + \tau^2\mathbf{I}_n)^{-1}\boldsymbol{y},$$

and

$$\sigma^2(\boldsymbol{x}^*) = K(\boldsymbol{x}^*, \boldsymbol{x}^*) - \boldsymbol{k}(\boldsymbol{x}^*)'(\mathrm{K} + \tau^2\mathbf{I}_n)^{-1}\boldsymbol{k}(\boldsymbol{x}^*).$$

The covariance function $K(.,.)$ has to be chosen in such a way that it reflects the prior information about the problem. For high-dimensional problems, the Gaussian covariance function is the most widely used one. The Gaussian covariance function is given by

$$K(\boldsymbol{x}_i, \boldsymbol{x}_j) = \rho^2 exp\left(-\frac{1}{2}\sum_{k=1}^{n}\frac{\|x_{ik} - x_{jk}\|^2}{h_k^2}\right),$$

where ρ^2 and $h_k{}'s$ are called the hyper parameters. In some prediction problem, Gaussian process cannot capture the asymmetry in the data (Zhang, H. and El-Sharaawi, 2009).

Chapter Three

Generalized Multivariate Skew- Normal Distribution

In this chapter, we use the idea of hidden truncation to generalize the multivariate skew normal distribution of Arnold and Beaver (2002). Moreover, we study some properties of that distribution such as, moment generating function, mean, covariance matrix, closure under marginal and conditional distributions.

3.1 Generalization of Multivariate Skew- Normal Distribution

To illustrate the hidden truncation, let $(\boldsymbol{X}', X_{n+1})' = (X_1, X_2, \dots, X_n, X_{n+1})'$ have an $(n+1)$-dimensional multivariate normal distribution with mean vector $\mathbf{0}$ $\big((\mathrm{n}+1)\times 1\big)$ and a positive definite covariance matrix $\boldsymbol{\Delta}$ $\big((n+1)\times(n+1)\big)$. Let A denote the event $\{\lambda_0 + \boldsymbol{\lambda}'\boldsymbol{X} > X_{n+1}\}$, where $\lambda_0 \in \mathbb{R}$ and $\boldsymbol{\lambda} = (\lambda_1, \lambda_2 \dots, \lambda_n)' \in \mathbb{R}^n$. If $\boldsymbol{\Delta} = \begin{pmatrix} \boldsymbol{\Sigma}_{n\times n} & \boldsymbol{\gamma}_{n\times 1} \\ \boldsymbol{\gamma}'_{n\times 1} & \sigma^2 \end{pmatrix}$, where $\boldsymbol{\gamma}$ be an $n \times 1$ real vector and $\sigma^2 > 0$, then the conditional distribution of $\boldsymbol{X}$ given the event A is said to have an n −dimensional generalized skew-normal distribution (GSN), denoted by $\boldsymbol{X}|A \sim GSN_n^1(\lambda_0, \sigma^2, \boldsymbol{\lambda}, \boldsymbol{\gamma}, \boldsymbol{\Sigma})$. Hence

$$f\,(\boldsymbol{x}|A) = \frac{\varphi_n(\boldsymbol{x}; \mathbf{0}, \boldsymbol{\Sigma}\,)}{\Phi(\delta_o)} \Phi\left(\frac{\lambda_o + (\boldsymbol{\lambda}' - \boldsymbol{\gamma}'\boldsymbol{\Sigma}^{-1})\boldsymbol{x}}{\sqrt{\sigma^2 - \boldsymbol{\gamma}'\boldsymbol{\Sigma}^{-1}\boldsymbol{\gamma}}}\right), \qquad (3.1)$$

where $\delta_0 = \lambda_0/\sqrt{\sigma^2 - 2\boldsymbol{\lambda}'\boldsymbol{\gamma} + \boldsymbol{\lambda}'\boldsymbol{\Sigma}\boldsymbol{\lambda}}$, $\varphi_n(\boldsymbol{x}; \mathbf{0}, \boldsymbol{\Sigma}\,)$ is the n −dimensional multivariate normal density function with mean vector $\mathbf{0}$ and covariance matrix $\boldsymbol{\Sigma}$. This density can be derived as follows. From multivariate normal theory, we have the following two facts:

1. $(X_{n+1}|\boldsymbol{X} = \boldsymbol{x}) \sim N(\boldsymbol{\gamma}'\boldsymbol{\Sigma}^{-1}\boldsymbol{x}, \sigma^2 - \boldsymbol{\gamma}'\boldsymbol{\Sigma}^{-1}\boldsymbol{\gamma})$,

2. $P(A|X=x) = \Phi\left(\frac{\lambda_0 + \boldsymbol{\lambda}'\boldsymbol{x} - \boldsymbol{\gamma}'\boldsymbol{\Sigma}^{-1}\boldsymbol{x}}{\sqrt{\sigma^2 - \boldsymbol{\gamma}'\boldsymbol{\Sigma}^{-1}\boldsymbol{\gamma}}}\right) = \Phi\left(\frac{\lambda_0 + (\boldsymbol{\lambda}' - \boldsymbol{\gamma}'\boldsymbol{\Sigma}^{-1})\boldsymbol{x}}{\sqrt{\sigma^2 - \boldsymbol{\gamma}'\boldsymbol{\Sigma}^{-1}\boldsymbol{\gamma}}}\right).$

The proof is straightforward from the following results:

$$(\boldsymbol{X}', \quad X_{n+1})' = \begin{pmatrix} \boldsymbol{X} \\ X_{n+1} \end{pmatrix} \sim N_{n+1}(\boldsymbol{0}, \ \boldsymbol{\Delta}).$$

Since

$$X_{n+1} - \boldsymbol{\lambda}'\boldsymbol{X} = (-\boldsymbol{\lambda}' \quad 1)\begin{pmatrix} \boldsymbol{X} \\ X_{n+1} \end{pmatrix},$$

then

$$E(X_{n+1} - \boldsymbol{\lambda}'\boldsymbol{X}) = E(X_{n+1}) - \boldsymbol{\lambda}'E(\boldsymbol{X}) = 0$$

and

$$\begin{aligned} Var(X_{n+1} - \boldsymbol{\lambda}'\boldsymbol{X}) &= Var\left((-\boldsymbol{\lambda}' \quad 1)\begin{pmatrix} \boldsymbol{X} \\ X_{n+1} \end{pmatrix}\right), \\ &= (-\boldsymbol{\lambda}' \quad 1)\begin{pmatrix} \boldsymbol{\Sigma} & \boldsymbol{\gamma} \\ \boldsymbol{\gamma}' & \sigma^2 \end{pmatrix}\begin{pmatrix} -\boldsymbol{\lambda} \\ 1 \end{pmatrix}, \\ &= (-\boldsymbol{\lambda}'\boldsymbol{\Sigma} + \boldsymbol{\gamma}' \quad -\boldsymbol{\lambda}'\boldsymbol{\gamma} + \sigma^2)\begin{pmatrix} -\boldsymbol{\lambda} \\ 1 \end{pmatrix}, \\ &= \boldsymbol{\lambda}'\boldsymbol{\Sigma}\boldsymbol{\lambda} - 2\boldsymbol{\lambda}'\boldsymbol{\gamma} + \sigma^2. \end{aligned}$$

So we have

$$X_{n+1} - \boldsymbol{\lambda}'\boldsymbol{X} \sim N(0, \boldsymbol{\lambda}'\boldsymbol{\Sigma}\boldsymbol{\lambda} - 2\boldsymbol{\lambda}'\boldsymbol{\gamma} + \sigma^2).$$

Using the facts 1, 2 and the Bayes' rule, we get

$$\begin{aligned} f(\boldsymbol{x}|A) &= \frac{P(A|\boldsymbol{x})\varphi_n(\boldsymbol{x}; \boldsymbol{0}, \boldsymbol{\Sigma})}{P(\lambda_0 + \boldsymbol{\lambda}'\boldsymbol{X} > X_{n+1})}, \\ &= \frac{\Phi\left(\frac{\lambda_0 + \boldsymbol{\lambda}'\boldsymbol{x} - \boldsymbol{\gamma}'\boldsymbol{\Sigma}^{-1}\boldsymbol{x}}{\sqrt{\sigma^2 - \boldsymbol{\gamma}'\boldsymbol{\Sigma}^{-1}\boldsymbol{\gamma}}}\right)\varphi_n(\boldsymbol{x}; \boldsymbol{0}, \boldsymbol{\Sigma})}{P(X_{n+1} - \boldsymbol{\lambda}'\boldsymbol{X} < \lambda_0)}, \end{aligned}$$

$$= \frac{\varphi_n(x; 0, \Sigma)}{\Phi\left(\frac{\lambda_0}{\sqrt{\boldsymbol{\lambda}'\boldsymbol{\Sigma}\boldsymbol{\lambda} - 2\boldsymbol{\lambda}'\gamma + \sigma^2}}\right)} \Phi\left(\frac{\lambda_0 + (\boldsymbol{\lambda}' - \gamma\boldsymbol{\Sigma}^{-1})x}{\sqrt{\sigma^2 - \gamma\boldsymbol{\Sigma}^{-1}\gamma}}\right),$$

i.e,

$$f\,(x|A) = \frac{\varphi_n(x; 0, \Sigma)}{\Phi(\delta_0)} \Phi\left(\frac{\lambda_0 + (\boldsymbol{\lambda}' - \gamma'\boldsymbol{\Sigma}^{-1})x}{\sqrt{\sigma^2 - \gamma'\boldsymbol{\Sigma}^{-1}\gamma}}\right).$$

A particular case of this density was obtained by Arnold and Beaver (2002) for choices $\sigma^2 = 1, \boldsymbol{\gamma} = \mathbf{0}$ and $\boldsymbol{\Sigma} = \mathbf{I}_n$. Moreover, this density is reduced to the density that was derived by Azzalini and Dalla Valle (1996) and Azzalini and Capitanio (1999) for $\sigma^2 = 1$, $\lambda_0 = 0$ and $\boldsymbol{\gamma} = \mathbf{0}$. For $\boldsymbol{\lambda} = \mathbf{0}$ and $\boldsymbol{\gamma} = \mathbf{0}$, this density is reduced to the well known multivariate normal distribution.

3.2 Location –Scale Generalized Skew- Normal Distribution

A more general form of the generalized n −skew normal distribution is obtained by introducing a location parameter $\boldsymbol{\mu}$ and a scale parameter $\boldsymbol{\Omega}$ in model (3.1). Here $\boldsymbol{\mu}$ is $n \times 1$ real vector and $\boldsymbol{\Omega}$ is $n \times n$ symmetric and positive definite matrix. Since $\boldsymbol{\Omega}$ is a symmetric and positive definite matrix, it has a unique positive definite square root matrix that we will denote by $\boldsymbol{\Omega}^{\frac{1}{2}}$. To illustrate the location-scale distribution, we prove the following theorem:

Theorem 3.2.1. Let $\boldsymbol{X} \sim GSN_n^1(\lambda_0, \sigma^2, \boldsymbol{\lambda}, \boldsymbol{\gamma}, \boldsymbol{\Sigma})$, and define $\boldsymbol{Y} = \boldsymbol{\mu} + \boldsymbol{\Omega}^{\frac{1}{2}}\boldsymbol{X}$, then

$$f_Y(\boldsymbol{y}) = \frac{\varphi_n\left(\boldsymbol{y}; \boldsymbol{\mu}, \boldsymbol{\Omega}^{\frac{1}{2}}\boldsymbol{\Sigma}\boldsymbol{\Omega}^{\frac{1}{2}}\right)}{\Phi(\delta_0)} \Phi\left(\frac{\lambda_0 + (\boldsymbol{\lambda}' - \boldsymbol{\gamma}'\boldsymbol{\Sigma}^{-1})\boldsymbol{\Omega}^{-\frac{1}{2}}(\boldsymbol{y} - \boldsymbol{\mu})}{\sqrt{\sigma^2 - \boldsymbol{\gamma}'\boldsymbol{\Sigma}^{-1}\boldsymbol{\gamma}}}\right).$$

Proof. Since $Y = \mu + \Omega^{\frac{1}{2}}X$, then $X = \Omega^{-\frac{1}{2}}(Y - \mu)$ and the jacobian of this transformation is $|J| = |\Omega|^{-\frac{1}{2}}$. So Y has the pdf $f_Y(y)$, as

$$f_Y(y) = \frac{exp\left(-\frac{1}{2}(y-\mu)'\Omega^{-\frac{1}{2}}\Sigma^{-1}\Omega^{-\frac{1}{2}}(y-\mu)\right)}{(2\pi)^{\frac{n}{2}}\Phi(\delta_0)|\Sigma|^{\frac{1}{2}}|\Omega|^{\frac{1}{2}}}\Phi\left(\frac{\lambda_0 + (\lambda' - \gamma'\Sigma^{-1})\Omega^{-\frac{1}{2}}(y-\mu)}{\sqrt{\sigma^2 - \gamma'\Sigma^{-1}\gamma}}\right),$$

$$= \frac{exp\left(-\frac{1}{2}(y-\mu)'\left(\Omega^{\frac{1}{2}}\Sigma\Omega^{\frac{1}{2}}\right)^{-1}(y-\mu)\right)}{(2\pi)^{\frac{n}{2}}\Phi(\delta_0)|\Sigma\Omega|^{\frac{1}{2}}}\Phi\left(\frac{\lambda_0 + (\lambda' - \gamma'\Sigma^{-1})\Omega^{-\frac{1}{2}}(y-\mu)}{\sqrt{\sigma^2 - \gamma'\Sigma^{-1}\gamma}}\right),$$

$$= \frac{\varphi_n\left(y; \mu, \Omega^{\frac{1}{2}}\Sigma\Omega^{\frac{1}{2}}\right)}{\Phi(\delta_0)}\Phi\left(\frac{\lambda_0 + (\lambda' - \gamma'\Sigma^{-1})\Omega^{-\frac{1}{2}}(y-\mu)}{\sqrt{\sigma^2 - \gamma'\Sigma^{-1}\gamma}}\right).$$

We use the notation $Y \sim GSN_n^2\left(\mu, \Omega^{\frac{1}{2}}\Sigma\Omega^{\frac{1}{2}}, \lambda_0, \sigma^2, \lambda, \gamma, \Sigma\right)$ to mean that Y follows the location-scale generalized n −skew normal distribution.

3.3 Marginal and Conditional Distributions of GSN

In this section, we derive the marginal and conditional distribution of the GSN. The following theorem shows that the generalized skew normal family is closed under marginal and conditional distributions.

Theorem 3.3.1. If $X \sim GSN_n^1(\lambda_0, \sigma^2, \boldsymbol{\lambda}, \boldsymbol{\gamma}, \boldsymbol{\Sigma})$ and if we partition $\mathbf{X}, \boldsymbol{\Sigma}^{-1}, \boldsymbol{\gamma}$ and $\boldsymbol{\lambda}$ as

$$X = \begin{array}{c} \\ \begin{pmatrix} X_1 \\ X_2 \end{pmatrix} \end{array} \begin{array}{l} m \\ n-m \end{array}, \boldsymbol{\Sigma}^{-1} = \mathbf{B} = \begin{array}{cc} m & n-m \\ \multicolumn{2}{c}{\begin{pmatrix} \mathbf{B}_{11} & \mathbf{B}_{12} \\ \mathbf{B}_{21} & \mathbf{B}_{22} \end{pmatrix}} \end{array} \begin{array}{l} m \\ n-m \end{array}, \boldsymbol{\gamma} = \begin{pmatrix} \gamma_1 \\ \gamma_2 \end{pmatrix} \begin{array}{l} m \\ n-m \end{array} \text{ and } \boldsymbol{\lambda} = \begin{pmatrix} \boldsymbol{\lambda}_1 \\ \boldsymbol{\lambda}_2 \end{pmatrix} \begin{array}{l} m \\ n-m \end{array},$$

where

$$\mathbf{B}_{11} = (\boldsymbol{\Sigma}_{11} - \boldsymbol{\Sigma}_{12}\boldsymbol{\Sigma}_{22}^{-1}\boldsymbol{\Sigma}_{21})^{-1}, \mathbf{B}_{12} = -\boldsymbol{\Sigma}_{11}^{-1}\boldsymbol{\Sigma}_{12}(\boldsymbol{\Sigma}_{22} - \boldsymbol{\Sigma}_{21}\boldsymbol{\Sigma}_{11}^{-1}\boldsymbol{\Sigma}_{12})^{-1},$$

$$\mathbf{B}_{21} = -\boldsymbol{\Sigma}_{22}^{-1}\boldsymbol{\Sigma}_{21}(\boldsymbol{\Sigma}_{11} - \boldsymbol{\Sigma}_{12}\boldsymbol{\Sigma}_{22}^{-1}\boldsymbol{\Sigma}_{21})^{-1} \text{ and } \mathbf{B}_{22} = (\boldsymbol{\Sigma}_{22} - \boldsymbol{\Sigma}_{21}\boldsymbol{\Sigma}_{11}^{-1}\boldsymbol{\Sigma}_{12})^{-1},$$

then

1. $X_1 \sim GSN_m^1(\lambda_0, \dot{\sigma}_*^2, \boldsymbol{\lambda}_*, \boldsymbol{\gamma}_*, \boldsymbol{\Sigma}_{11})$,

where

$\dot{\sigma}_*^2 = \sigma^{*2} - \boldsymbol{\gamma}_2'\mathbf{B}_{22}\boldsymbol{\gamma}_2 + \boldsymbol{\gamma}_*'\boldsymbol{\Sigma}_{11}^{-1}\boldsymbol{\gamma}_*, \sigma^{*2} = \sigma^2 - \boldsymbol{\gamma}_1'\mathbf{B}_{11}\boldsymbol{\gamma}_1 - 2\boldsymbol{\gamma}_1'\mathbf{B}_{12}\boldsymbol{\gamma}_2, \boldsymbol{\gamma}_* = \boldsymbol{\Sigma}_{11}\mathbf{B}_{11}\boldsymbol{\gamma}_1,$

$\boldsymbol{\lambda}_* = \boldsymbol{\lambda}_1 + \boldsymbol{\Sigma}_{11}^{-1}\boldsymbol{\Sigma}_{12}\boldsymbol{\lambda}_2^*$ and $\boldsymbol{\lambda}_2^{*\prime} = \boldsymbol{\lambda}_2' - \boldsymbol{\gamma}_1'\mathbf{B}_{12}$,

2. $(X_2|\mathbf{X}_1 = x_1) \sim GSN_{n-m}^2\left(\boldsymbol{\mu}_{x_1}, \boldsymbol{\Sigma}_{22}^{\frac{1}{2}}\mathbf{M}\boldsymbol{\Sigma}_{22}^{\frac{1}{2}}, \dot{\lambda}_0^*, \sigma^{*2}, \boldsymbol{\lambda}_2^*, \dot{\boldsymbol{\gamma}}_2, \mathbf{M}\right)$,

where

$$\boldsymbol{\mu}_{x_1} = \boldsymbol{\Sigma}_{21}\boldsymbol{\Sigma}_{11}^{-1}x_1, \quad \mathbf{M} = \boldsymbol{\Sigma}_{22}^{\frac{1}{2}}\mathbf{A}\boldsymbol{\Sigma}_{22}^{\frac{1}{2}}, \mathbf{A}^{-1} = \boldsymbol{\Sigma}_{22} + (\boldsymbol{\Sigma}_{22}^{-1} - (\boldsymbol{\Sigma}_{21}\boldsymbol{\Sigma}_{11}^{-1}\boldsymbol{\Sigma}_{12})^{-1})^{-1},$$

$$\dot{\lambda}_0^* = \lambda_0^* + (\boldsymbol{\lambda}_2^{*\prime} - \boldsymbol{\gamma}_2'\mathbf{B}_{22})\boldsymbol{\Sigma}_{21}\boldsymbol{\Sigma}_{11}^{-1}x_1, \ \lambda_0^* = \lambda_0 + (\boldsymbol{\lambda}_1' - \boldsymbol{\gamma}_1'\mathbf{B}_{11} - \boldsymbol{\gamma}_2'\mathbf{B}_{21})x_1,$$

$\dot{\lambda}_2^{*'} = \lambda_2^{*'}\Sigma_{22}^{\frac{1}{2}}$ and $\dot{\gamma}_2 = \Sigma_{22}^{-\frac{1}{2}}\gamma_2$.

Proof of 1. To simplify the calculations, let $x_{12} = x_2 - \Sigma_{21}\Sigma_{11}^{-1}x_1$, and write $-\frac{1}{2}x'\Sigma^{-1}x$

as

$$-\frac{1}{2}x'\Sigma^{-1}x = -\frac{1}{2}(x_1'\Sigma_{11}^{-1}x_1 + x_{12}'\mathbf{B}_{22}x_{12}).$$

Also $\lambda_0 + (\lambda' - \gamma'\Sigma^{-1})x$ can be written as

$$\lambda_0 + (\lambda' - \gamma'\Sigma^{-1})x = \lambda_0^* + (\lambda_2^{*\prime} - \gamma_2'\mathbf{B}_{22})x_2,$$

where

$$\lambda_0^* = \lambda_0 + (\lambda_1' - \gamma_1'\mathbf{B}_{11} - \gamma_2'\mathbf{B}_{21})x_1 \text{ and } \lambda_2^{*\prime} = \lambda_2' - \gamma_1'\mathbf{B}_{12}.$$

and

$$|\Sigma|^{\frac{1}{2}} = |\Sigma_{11}|^{\frac{1}{2}}|\Sigma_{22} - \Sigma_{21}\Sigma_{11}^{-1}\Sigma_{12}|^{\frac{1}{2}},$$

$$= |\Sigma_{11}|^{\frac{1}{2}}|\mathbf{B}_{22}^{-1}|^{\frac{1}{2}}.$$

So

$$f(x_1) = \frac{exp\left(-\frac{1}{2}x_1'\Sigma_{11}^{-1}x_1\right)}{(2\pi)^{\frac{m}{2}}\Phi(\delta_0)|\Sigma_{11}|^{\frac{1}{2}}}\int \Phi\left(\frac{\lambda_0^* + (\lambda_2^{*\prime} - \gamma_2'\mathbf{B}_{22})x_2}{\sqrt{\sigma^{*2} - \gamma_2'\mathbf{B}_{22}\gamma_2}}\right)\frac{exp\left(-\frac{1}{2}x_{12}'\mathbf{B}_{22}x_{12}\right)}{(2\pi)^{\frac{n-m}{2}}|\mathbf{B}_{22}^{-1}|^{\frac{1}{2}}}dx_2,$$

$$= \frac{exp\left(-\frac{1}{2}x_1'\Sigma_{11}^{-1}x_1\right)}{(2\pi)^{\frac{m}{2}}\Phi(\delta_0)|\Sigma_{11}|^{\frac{1}{2}}}\int \frac{exp\left(-\frac{1}{2}x_{12}'\mathbf{B}_{22}x_{12}\right)}{(2\pi)^{\frac{n-m}{2}}|\mathbf{B}_{22}^{-1}|^{\frac{1}{2}}} \times$$

$$\Phi\left(\frac{\lambda_0^* + (\lambda_2^{*\prime} - \gamma_2'\mathbf{B}_{22})\Sigma_{21}\Sigma_{11}^{-1}x_1 + (\lambda_2^{*\prime} - \gamma_2'\mathbf{B}_{22})x_{12}}{\sqrt{\sigma^{*2} - \gamma_2'\mathbf{B}_{22}\gamma_2}}\right)dx_{12},$$

$$= \frac{exp\left(-\frac{1}{2}x_1'\boldsymbol{\Sigma}_{11}^{-1}x_1\right)}{(2\pi)^{\frac{m}{2}}\Phi(\delta_0)|\boldsymbol{\Sigma}_{11}|^{\frac{1}{2}}}\Phi\left(\frac{\lambda_0^* + (\boldsymbol{\lambda}_2^{*\prime} - \boldsymbol{\gamma}_2'\mathbf{B}_{22})\boldsymbol{\Sigma}_{21}\boldsymbol{\Sigma}_{11}^{-1}x_1}{\sqrt{\sigma^{*2} - \boldsymbol{\gamma}_2'\mathbf{B}_{22}\boldsymbol{\gamma}_2}}\right).$$

Substituting the value of λ_0^*, then

$$f(x_1)$$

$$= \frac{exp\left(-\frac{1}{2}x_1'\boldsymbol{\Sigma}_{11}^{-1}x_1\right)}{(2\pi)^{\frac{m}{2}}\Phi(\delta_0)|\boldsymbol{\Sigma}_{11}|^{\frac{1}{2}}}\Phi\left(\frac{\lambda_0 + (\boldsymbol{\lambda}_1' - \boldsymbol{\gamma}_1'\mathbf{B}_{11} - \boldsymbol{\gamma}_2'\mathbf{B}_{21} + (\boldsymbol{\lambda}_2^{*\prime} - \boldsymbol{\gamma}_2'\mathbf{B}_{22})\boldsymbol{\Sigma}_{21}\boldsymbol{\Sigma}_{11}^{-1})x_1}{\sqrt{\sigma^{*2} - \boldsymbol{\gamma}_2'\mathbf{B}_{22}\boldsymbol{\gamma}_2}}\right).$$

More simplification, we get

$$f(x_1) = \frac{exp\left(-\frac{1}{2}x_1'\boldsymbol{\Sigma}_{11}^{-1}x_1\right)}{(2\pi)^{\frac{m}{2}}\Phi(\delta_0)|\boldsymbol{\Sigma}_{11}|^{\frac{1}{2}}}\Phi\left(\frac{\lambda_0 + (\boldsymbol{\lambda}_*' - \boldsymbol{\gamma}_*'\boldsymbol{\Sigma}_{11}^{-1})x_1}{\sqrt{\dot{\sigma}_*^2 - \boldsymbol{\gamma}_*'\boldsymbol{\Sigma}_{11}^{-1}\boldsymbol{\gamma}_*}}\right),$$

where

$$\dot{\sigma}_*^2 = \sigma^{*2} - \boldsymbol{\gamma}_2'\mathbf{B}_{22}\boldsymbol{\gamma}_2 + \boldsymbol{\gamma}_*'\boldsymbol{\Sigma}_{11}^{-1}\boldsymbol{\gamma}_*,\ \boldsymbol{\lambda}_*' = \boldsymbol{\lambda}_1' + \boldsymbol{\lambda}_2^{*\prime}\boldsymbol{\Sigma}_{21}\boldsymbol{\Sigma}_{11}^{-1} \text{ and } \boldsymbol{\gamma}_* = \boldsymbol{\Sigma}_{11}\mathbf{B}_{11}\boldsymbol{\gamma}_1.$$

Proof of 2. The pdf of $\boldsymbol{X}$ can be written as

$$f(x_1, x_2) = \frac{exp\left(-\frac{1}{2}(x_1'\boldsymbol{\Sigma}_{11}^{-1}x_1 + x_{12}'\mathbf{B}_{22}x_{12})\right)}{(2\pi)^{\frac{n}{2}}|\boldsymbol{\Sigma}|^{\frac{1}{2}}\Phi(\delta_0)}\Phi\left(\frac{\lambda_0 + (\boldsymbol{\lambda}' - \boldsymbol{\gamma}'\boldsymbol{\Sigma}^{-1})\boldsymbol{x}}{\sqrt{\sigma^2 - \boldsymbol{\gamma}'\boldsymbol{\Sigma}^{-1}\boldsymbol{\gamma}}}\right).$$

Hence

$$f(x_2|x_1) \propto exp\left(-\frac{1}{2}(x_{12}'\mathbf{B}_{22}x_{12})\right)\Phi\left(\frac{\lambda_0 + (\boldsymbol{\lambda}' - \boldsymbol{\gamma}'\boldsymbol{\Sigma}^{-1})\boldsymbol{x}}{\sqrt{\sigma^2 - \boldsymbol{\gamma}'\boldsymbol{\Sigma}^{-1}\boldsymbol{\gamma}}}\right).$$

To complete the proof, rewrite $\lambda_0 + (\boldsymbol{\lambda}' - \boldsymbol{\gamma}'\boldsymbol{\Sigma}^{-1})\boldsymbol{x}$ as

$$\lambda_0 + (\lambda' - \gamma'\boldsymbol{\Sigma}^{-1})x = \lambda_0 + \left((\lambda_1' \quad \lambda_2') - (\gamma_1' \quad \gamma_2')\Sigma^{-1}\right)\binom{x_1}{x_2},$$

$$= \lambda_0 + \left((\lambda_1' \quad \lambda_2') - (\gamma_1'\mathbf{B}_{11} + \gamma_2'\mathbf{B}_{21} \quad \gamma_1'\mathbf{B}_{12} + \gamma_2'\mathbf{B}_{22})\right)\binom{x_1}{x_2},$$

$$= \lambda_0 + (\lambda_1' - \gamma_1'\mathbf{B}_{11} - \gamma_2'\mathbf{B}_{21} \quad \lambda_2' - \gamma_1'\mathbf{B}_{12} - \gamma_2'\mathbf{B}_{22})\binom{x_1}{x_2},$$

$$= \lambda_0 + (\lambda_1' - \gamma_1'\mathbf{B}_{11} - \gamma_2'\mathbf{B}_{21})x_1 + (\lambda_2' - \gamma_1'\mathbf{B}_{12} - \gamma_2'\mathbf{B}_{22})x_2,$$

$$= \lambda_0^* + (\lambda_2^{*\prime} - \gamma_2'\mathbf{B}_{22})x_2,$$

$$= \dot{\lambda}_0^* + (\lambda_2^{*\prime} - \gamma_2'\mathbf{B}_{22})(x_2 - \boldsymbol{\Sigma}_{21}\boldsymbol{\Sigma}_{11}^{-1}x_1),$$

where

$\dot{\lambda}_0^* = \lambda_0^* + (\lambda_2^{*\prime} - \gamma_2'\mathbf{B}_{22})\boldsymbol{\Sigma}_{21}\boldsymbol{\Sigma}_{11}^{-1}x_1,\ \lambda_0^* = \lambda_0 + (\lambda_1' - \gamma_1'\mathbf{B}_{11} - \gamma_2'\mathbf{B}_{21})x_1$ and

$\lambda_2^{*'} = \lambda_2' - \gamma_1'\mathbf{B}_{12}$.

Similarly, rewrite $\sigma^2 - \gamma'\boldsymbol{\Sigma}^{-1}\gamma$ as follows:

$$\sigma^2 - \gamma'\boldsymbol{\Sigma}^{-1}\gamma = \sigma^2 - (\gamma_1' \quad \gamma_2')\mathbf{B}\binom{\gamma_1}{\gamma_2},$$

$$= \sigma^2 - \gamma_1'\mathbf{B}_{11}\gamma_1 - 2\gamma_1'\mathbf{B}_{12}\gamma_2 - \gamma_1'\mathbf{B}_{22}\gamma_2,$$

$$= \sigma^{*2} - \gamma_2'\mathbf{B}_{22}\gamma_2,$$

where

$$\sigma^{*2} = \sigma^2 - \gamma_1'\mathbf{B}_{11}\gamma_1 - 2\gamma_1'\mathbf{B}_{12}\gamma_2.$$

To complete, rewrite $|\boldsymbol{\Sigma}|^{\frac{1}{2}}$ as

$$|\boldsymbol{\Sigma}|^{\frac{1}{2}} = |\boldsymbol{\Sigma}_{11}(\boldsymbol{\Sigma}_{22} - \boldsymbol{\Sigma}_{21}\boldsymbol{\Sigma}_{11}^{-1}\boldsymbol{\Sigma}_{12})|^{\frac{1}{2}},$$

$$= |\boldsymbol{\Sigma}_{11}|^{\frac{1}{2}}|\mathbf{B}_{22}^{-1}|^{\frac{1}{2}}.$$

If we use the matrices identity

$$(\boldsymbol{\Sigma}_{22}-\boldsymbol{\Sigma}_{21}\boldsymbol{\Sigma}_{11}^{-1}\boldsymbol{\Sigma}_{12})^{-1}=\boldsymbol{\Sigma}_{22}^{-1}+\boldsymbol{\Sigma}_{22}^{-1}(\boldsymbol{\Sigma}_{22}^{-1}-(\boldsymbol{\Sigma}_{21}\boldsymbol{\Sigma}_{11}^{-1}\boldsymbol{\Sigma}_{12})^{-1})^{-1}\boldsymbol{\Sigma}_{22}^{-1},$$

then $\mathbf{B}_{22}=(\boldsymbol{\Sigma}_{22}-\boldsymbol{\Sigma}_{21}\boldsymbol{\Sigma}_{11}^{-1}\boldsymbol{\Sigma}_{12})^{-1}$ can be written as

$$\begin{aligned}\mathbf{B}_{22}&=\boldsymbol{\Sigma}_{22}^{-1}+\boldsymbol{\Sigma}_{22}^{-1}(\boldsymbol{\Sigma}_{22}^{-1}-(\boldsymbol{\Sigma}_{21}\boldsymbol{\Sigma}_{11}^{-1}\boldsymbol{\Sigma}_{12})^{-1})^{-1}\boldsymbol{\Sigma}_{22}^{-1},\\&=\boldsymbol{\Sigma}_{22}^{-1}(\boldsymbol{\Sigma}_{22}+(\boldsymbol{\Sigma}_{22}^{-1}-(\boldsymbol{\Sigma}_{21}\boldsymbol{\Sigma}_{11}^{-1}\boldsymbol{\Sigma}_{12})^{-1})^{-1})\boldsymbol{\Sigma}_{22}^{-1},\\&=\boldsymbol{\Sigma}_{22}^{-1}\mathbf{A}^{-1}\boldsymbol{\Sigma}_{22}^{-1}=\boldsymbol{\Sigma}_{22}^{-\frac{1}{2}}\boldsymbol{\Sigma}_{22}^{-\frac{1}{2}}\mathbf{A}^{-1}\boldsymbol{\Sigma}_{22}^{-\frac{1}{2}}\boldsymbol{\Sigma}_{22}^{-\frac{1}{2}},\\&=\boldsymbol{\Sigma}_{22}^{-\frac{1}{2}}\mathbf{M}^{-1}\boldsymbol{\Sigma}_{22}^{-\frac{1}{2}},\end{aligned}$$

where

$$\mathbf{M}=\boldsymbol{\Sigma}_{22}^{\frac{1}{2}}\mathbf{A}\boldsymbol{\Sigma}_{22}^{\frac{1}{2}}\text{ and }\mathbf{A}^{-1}=\boldsymbol{\Sigma}_{22}+(\boldsymbol{\Sigma}_{22}^{-1}-(\boldsymbol{\Sigma}_{21}\boldsymbol{\Sigma}_{11}^{-1}\boldsymbol{\Sigma}_{12})^{-1})^{-1}$$

So

$$f(x_2|x_1)\propto exp\left(-\frac{1}{2}(x_{12}'\mathbf{B}_{22}x_{12})\right)\Phi\left(\frac{\lambda_0+(\lambda'-\gamma'\Sigma^{-1})x}{\sqrt{\sigma^2-\gamma'\Sigma^{-1}\gamma}}\right),$$

$$\propto exp\left(-\frac{1}{2}(x_{12}'\mathbf{B}_{22}x_{12})\right)\Phi\left(\frac{\lambda_0^*+(\lambda_2^{*}{}'-\gamma_2'\mathbf{B}_{22})x_2}{\sqrt{\sigma^{*2}-\gamma_2'\mathbf{B}_{22}\gamma_2}}\right),$$

$$\propto exp\left(-\frac{1}{2}((x_2-\boldsymbol{\Sigma}_{21}\boldsymbol{\Sigma}_{11}^{-1}x_1)'\mathbf{B}_{22}(x_2-\boldsymbol{\Sigma}_{21}\boldsymbol{\Sigma}_{11}^{-1}x_1))\right)\times$$

$$\Phi\left(\frac{\lambda_0^*+(\lambda_2^{*}{}'-\gamma_2'\mathbf{B}_{22})x_2}{\sqrt{\sigma^{*2}-\gamma_2'\mathbf{B}_{22}\gamma_2}}\right),$$

$$\propto exp\left(-\frac{1}{2}\left((x_2-\mu_{x_1})'\left(\boldsymbol{\Sigma}_{22}^{\frac{1}{2}}\mathbf{M}\boldsymbol{\Sigma}_{22}^{\frac{1}{2}}\right)^{-1}(x_2-\mu_{x_1})\right)\right)\times$$

$$\Phi\left(\frac{\dot{\lambda}_0^* + \left(\lambda_2^{*\prime} - \gamma_2'\Sigma_{22}^{-\frac{1}{2}}\mathbf{M}^{-1}\Sigma_{22}^{-\frac{1}{2}}\right)(x_2 - \mu_{x_1})}{\sqrt{\sigma^{*2} - \gamma_2'\left(\Sigma_{22}^{\frac{1}{2}}\mathbf{M}\Sigma_{22}^{\frac{1}{2}}\right)^{-1}\gamma_2}}\right),$$

$$\propto exp\left(-\frac{1}{2}\left((x_2 - \mu_{x_1})'\left(\Sigma_{22}^{\frac{1}{2}}\mathbf{M}\Sigma_{22}^{\frac{1}{2}}\right)^{-1}(x_2 - \mu_{x_1})\right)\right)\times$$

$$\Phi\left(\frac{\dot{\lambda}_0^* + \left(\dot{\lambda}_2^{*\prime} - \gamma_2'\Sigma_{22}^{-\frac{1}{2}}\mathbf{M}^{-1}\right)\Sigma_{22}^{-\frac{1}{2}}(x_2 - \mu_{x_1})}{\sqrt{\sigma^{*2} - \gamma_2'\left(\Sigma_{22}^{\frac{1}{2}}\mathbf{M}\Sigma_{22}^{\frac{1}{2}}\right)^{-1}\gamma_2}}\right),$$

$$\propto exp\left(-\frac{1}{2}\left((x_2 - \mu_{x_1})'\left(\Sigma_{22}^{\frac{1}{2}}\mathbf{M}\Sigma_{22}^{\frac{1}{2}}\right)^{-1}(x_2 - \mu_{x_1})\right)\right)\times$$

$$\Phi\left(\frac{\dot{\lambda}_0^* + (\dot{\lambda}_2^{*\prime} - \dot{\gamma}_2'\mathbf{M}^{-1})\Sigma_{22}^{-\frac{1}{2}}(x_2 - \mu_{x_1})}{\sqrt{\sigma^{*2} - \dot{\gamma}_2'\mathbf{M}^{-1}\dot{\gamma}_2}}\right).$$

Finally, the conditional distribution function of X_2 given $X_1 = x_1$ is

$$f(x_2|x_1) = \frac{exp\left(-\frac{1}{2}\left((x_2 - \mu_{x_1})'\left(\Sigma_{22}^{\frac{1}{2}}\mathbf{M}\Sigma_{22}^{\frac{1}{2}}\right)^{-1}(x_2 - \mu_{x_1})\right)\right)}{(2\pi)^{\frac{n-m}{2}}\Phi(\delta_0^{**})|\mathbf{M}\Sigma_{22}|^{\frac{1}{2}}}\times$$

$$\Phi\left(\frac{\dot{\lambda}_0^* + (\dot{\lambda}_2^{*\prime} - \dot{\gamma}_2'\mathbf{M}^{-1})\Sigma_{22}^{-\frac{1}{2}}(x_2 - \mu_{x_1})}{\sqrt{\sigma^{*2} - \dot{\gamma}_2'\mathbf{M}^{-1}\dot{\gamma}_2}}\right),$$

where

$$\delta_0^{**} = \frac{\dot{\lambda}_0^*}{\sqrt{\sigma^{*2} - 2\dot{\lambda}_2^{*'}\dot{\gamma}_2 + \dot{\lambda}_2^{*'}\mathbf{M}^{-1}\dot{\lambda}_2^*}}.$$

Now, we are able to find the moment generating function by the following theorem

Theorem 3.3.2. The moment generating function of $\boldsymbol{X} \sim GSN_n^1(\lambda_0, \sigma^2, \boldsymbol{\lambda}, \boldsymbol{\gamma}, \boldsymbol{\Sigma})$ is

$$M_{\boldsymbol{X}}(\boldsymbol{t}) = \frac{exp\left(\frac{1}{2}\boldsymbol{t}'\boldsymbol{\Sigma}\boldsymbol{t}\right)}{\Phi(\delta_0)}\Phi\left(\frac{\lambda_0 + (\boldsymbol{\lambda}' - \boldsymbol{\gamma}'\boldsymbol{\Sigma}^{-1})\boldsymbol{\Sigma}\boldsymbol{t}}{\sqrt{\sigma^2 - 2\boldsymbol{\lambda}'\boldsymbol{\gamma} + \boldsymbol{\lambda}'\boldsymbol{\Sigma}\boldsymbol{\lambda}}}\right), \quad \boldsymbol{t} \in \mathbb{R}^n \qquad (3.2)$$

Proof: By the definition of the mgf, we have

$$M_{\boldsymbol{X}}(\boldsymbol{t}) = \int \frac{1}{(2\pi)^{\frac{n}{2}}\Phi(\delta_0)|\boldsymbol{\Sigma}|^{\frac{1}{2}}} exp\left(-\frac{1}{2}\boldsymbol{x}'\boldsymbol{\Sigma}^{-1}\boldsymbol{x}\right)\Phi\left(\frac{\lambda_0 + (\boldsymbol{\lambda}' - \boldsymbol{\gamma}'\boldsymbol{\Sigma}^{-1})\boldsymbol{x}}{\sqrt{\sigma^2 - \boldsymbol{\gamma}'\boldsymbol{\Sigma}^{-1}\boldsymbol{\gamma}}}\right) exp(\boldsymbol{t}'\boldsymbol{x})d\boldsymbol{x},$$

where the integration is taken over $\mathbb{R}^n$. If we set

$$C = \frac{1}{(2\pi)^{\frac{n}{2}}\Phi(\delta_0)|\boldsymbol{\Sigma}|^{\frac{1}{2}}},$$

then

$$M_{\boldsymbol{X}}(\boldsymbol{t}) = C\int \Phi\left(\frac{\lambda_0 + (\boldsymbol{\lambda}' - \boldsymbol{\gamma}'\boldsymbol{\Sigma}^{-1})\boldsymbol{x}}{\sqrt{\sigma^2 - \boldsymbol{\gamma}'\boldsymbol{\Sigma}^{-1}\boldsymbol{\gamma}}}\right) exp\left(-\frac{1}{2}\boldsymbol{x}'\boldsymbol{\Sigma}^{-1}\boldsymbol{x} + \boldsymbol{t}'\boldsymbol{x}\right)d\boldsymbol{x}.$$

Applying the transformation $\boldsymbol{x} = \boldsymbol{y} + \boldsymbol{\Sigma}\boldsymbol{t}$, we get

$$M_{X}(t) = C\int \Phi\left(\frac{\lambda_0 + (\lambda' - \gamma'\Sigma^{-1})(y + \Sigma t)}{\sqrt{\sigma^2 - \gamma'\Sigma^{-1}\gamma}}\right)\times$$

$$exp\left(-\frac{1}{2}(y + \Sigma t)'\Sigma^{-1}(y + \Sigma t) + t'(y + \Sigma t)\right)\ dy,$$

$$= C\int \Phi\left(\frac{\lambda_0 + (\lambda' - \gamma'\Sigma^{-1})\Sigma t + (\lambda' - \gamma'\Sigma^{-1})y}{\sqrt{\sigma^2 - \gamma'\Sigma^{-1}\gamma}}\right)\times$$

$$exp\left(-\frac{1}{2}(y'\Sigma^{-1} + t')(y + \Sigma t) + t'(y + \Sigma t)\right)dy,$$

$$= C\int \Phi\left(\frac{\lambda_0 + (\lambda' - \gamma'\Sigma^{-1})\Sigma t + (\lambda' - \gamma'\Sigma^{-1})y}{\sqrt{\sigma^2 - \gamma'\Sigma^{-1}\gamma}}\right)\times$$

$$exp\left(-\frac{1}{2}y'\Sigma^{-1}y - t'y - \frac{1}{2}t'\Sigma y + t'y + \frac{1}{2}t'\Sigma t\right)dy,$$

$$= Cexp\left(\frac{1}{2}t'\Sigma t\right)\int exp\left(\frac{-1}{2}y'\Sigma^{-1}y\right)\times$$

$$\Phi\left(\frac{\lambda_0 + (\lambda' - \gamma'\Sigma^{-1})\Sigma t + (\lambda' - \gamma'\Sigma^{-1})y}{\sqrt{\sigma^2 - \gamma'\Sigma^{-1}\gamma}}\right)dy,$$

$$= \frac{exp\left(\frac{1}{2}t'\Sigma t\right)}{(2\pi)^{\frac{n}{2}}\Phi(\delta_0)}\int \frac{exp\left(-\frac{1}{2}y'\Sigma^{-1}y\right)}{|\Sigma|^{\frac{1}{2}}}\times$$

$$\Phi\left(\frac{\lambda_0 + (\lambda' - \gamma'\Sigma^{-1})\Sigma t + (\lambda' - \gamma'\Sigma^{-1})y}{\sqrt{\sigma^2 - \gamma'\Sigma^{-1}\gamma}}\right)\times$$

$$\frac{\Phi\left(\frac{\lambda_0 + (\boldsymbol{\lambda}' - \boldsymbol{\gamma}'\boldsymbol{\Sigma}^{-1})\boldsymbol{\Sigma t}}{\sqrt{\sigma^2 - 2\boldsymbol{\lambda}'\boldsymbol{\gamma} + \boldsymbol{\lambda}'\boldsymbol{\Sigma}\boldsymbol{\lambda}}}\right)}{\Phi\left(\frac{\lambda_0 + (\boldsymbol{\lambda}' - \boldsymbol{\gamma}'\boldsymbol{\Sigma}^{-1})\boldsymbol{\Sigma t}}{\sqrt{\sigma^2 - 2\boldsymbol{\lambda}'\boldsymbol{\gamma} + \boldsymbol{\lambda}'\boldsymbol{\Sigma}\boldsymbol{\lambda}}}\right)} dy,$$

$$= \frac{exp\left(\frac{1}{2}\boldsymbol{t}'\boldsymbol{\Sigma t}\right)}{\Phi(\delta_0)} \Phi\left(\frac{\lambda_0 + (\boldsymbol{\lambda}' - \boldsymbol{\gamma}'\boldsymbol{\Sigma}^{-1})\boldsymbol{\Sigma t}}{\sqrt{\sigma^2 - 2\boldsymbol{\lambda}'\boldsymbol{\gamma} + \boldsymbol{\lambda}'\boldsymbol{\Sigma}\boldsymbol{\lambda}}}\right).$$

We find the expectation and the covariance matrix by the following theorem.

Theorem 3.3.3. If $\boldsymbol{X} \sim GSN_n^1(\lambda_0, \sigma^2, \boldsymbol{\lambda}, \boldsymbol{\gamma}, \boldsymbol{\Sigma})$, then

1. $E(\boldsymbol{X}) = \frac{\varphi(\delta_0)}{\Phi(\delta_0)} \frac{\boldsymbol{\Sigma}(\boldsymbol{\lambda} - \boldsymbol{\Sigma}^{-1}\boldsymbol{\gamma})}{\sqrt{\sigma^2 - 2\boldsymbol{\lambda}'\boldsymbol{\gamma} + \boldsymbol{\lambda}'\boldsymbol{\Sigma}\boldsymbol{\lambda}}}.$

2. $Cov(\boldsymbol{X}) = \boldsymbol{\Sigma} - h\boldsymbol{\Sigma}^2.$

where

$$h = \left(\left(\frac{\varphi(\delta_0)}{\sqrt{k}\Phi(\delta_0)}\right)^2 + \frac{\delta_0\varphi(\delta_0)}{k\Phi(\delta_0)}\right)(\boldsymbol{\lambda}' - \boldsymbol{\gamma}'\boldsymbol{\Sigma}^{-1})(\boldsymbol{\lambda} - \boldsymbol{\Sigma}^{-1}\boldsymbol{\gamma}),$$

and

$$k = \sigma^2 - 2\boldsymbol{\lambda}'\boldsymbol{\gamma} + \boldsymbol{\lambda}'\boldsymbol{\Sigma}\boldsymbol{\lambda}.$$

Proof of 1. Taking the first derivative of $M_{\mathrm{X}}(\boldsymbol{t})$ with respect to $\boldsymbol{t}$ yields

$$\frac{\partial M_{\mathrm{X}}(\boldsymbol{t})}{\partial \boldsymbol{t}} = \frac{1}{\Phi(\delta_0)}\left[\Phi\left(\frac{\lambda_0 + (\boldsymbol{\lambda}' - \boldsymbol{\gamma}'\boldsymbol{\Sigma}^{-1})\boldsymbol{\Sigma t}}{\sqrt{\sigma^2 - 2\boldsymbol{\lambda}'\boldsymbol{\gamma} + \boldsymbol{\lambda}'\boldsymbol{\Sigma}\boldsymbol{\lambda}}}\right)(\boldsymbol{\Sigma t})exp\left(\frac{1}{2}\boldsymbol{t}'\boldsymbol{\Sigma t}\right)\right] +$$

$$exp\left(\frac{1}{2}\boldsymbol{t}'\Sigma\boldsymbol{t}\right)\left(\frac{(\lambda'-\gamma'\Sigma^{-1})\Sigma}{\sqrt{\sigma^2-2\lambda'\gamma+\lambda'\Sigma\lambda}}\right)'\varphi\left(\frac{\lambda_0+(\lambda'-\gamma'\Sigma^{-1})\Sigma\boldsymbol{t}}{\sqrt{\sigma^2-2\lambda'\gamma+\lambda'\Sigma\lambda}}\right) \qquad (3.3)$$

Setting $\boldsymbol{t}=\boldsymbol{0}$ in (3.3) yields

$$E(X)=\frac{\partial M_X(\boldsymbol{t})}{\partial \boldsymbol{t}}\Big|_{\boldsymbol{t}=\boldsymbol{0}}$$

$$=\frac{1}{\Phi(\delta_0)}\left[\Phi\left(\frac{\lambda_0+(\lambda'-\gamma'\Sigma^{-1})\Sigma\boldsymbol{0}}{\sqrt{\sigma^2-2\lambda'\gamma+\lambda'\Sigma\lambda}}\right)(\Sigma\boldsymbol{0})exp\left(\frac{1}{2}\boldsymbol{0}'\Sigma\boldsymbol{0}\right)\right]+$$

$$exp\left(\frac{1}{2}\boldsymbol{0}'\Sigma\boldsymbol{0}\right)\left(\frac{(\lambda'-\gamma'\Sigma^{-1})\Sigma}{\sqrt{\sigma^2-2\lambda'\gamma+\lambda'\Sigma\lambda}}\right)'\varphi\left(\frac{\lambda_0+(\lambda'-\gamma'\Sigma^{-1})\Sigma\boldsymbol{0}}{\sqrt{\sigma^2-2\lambda'\gamma+\lambda'\Sigma\lambda}}\right),$$

$$=\frac{\varphi\left(\frac{\lambda_0}{\sqrt{\sigma^2-2\lambda'\gamma+\lambda'\Sigma\lambda}}\right)}{\Phi(\delta_0)}\frac{\Sigma'(\lambda'-\gamma'\Sigma^{-1})'}{\sqrt{\sigma^2-2\lambda'\gamma+\lambda'\Sigma\lambda}},$$

Since Σ is a symmetric matrix, then

$$E(X)=\frac{\varphi\left(\frac{\lambda_0}{\sqrt{\sigma^2-2\lambda'\gamma+\lambda'\Sigma\lambda}}\right)}{\Phi(\delta_0)}\frac{\Sigma(\lambda-\Sigma^{-1}\gamma)}{\sqrt{\sigma^2-2\lambda'\gamma+\lambda'\Sigma\lambda}}.$$

Finally, we get

$$E(X)=\frac{\varphi(\delta_0)}{\Phi(\delta_0)}\frac{\Sigma(\lambda-\Sigma^{-1}\gamma)}{\sqrt{\sigma^2-2\lambda'\gamma+\lambda'\Sigma\lambda}}.$$

Proof of 2. The covariance matrix is given by

$$Cov(X)=E(XX')-E(X)E(X'),$$

where

$$E(\boldsymbol{XX'}) = \frac{\partial^2 M_{\mathrm{X}}(\boldsymbol{t})}{\partial \boldsymbol{t} \partial \boldsymbol{t}'} \Big|_{\boldsymbol{t} = \boldsymbol{0}}.$$

Firstly we need to find $E(\boldsymbol{XX'})$. Using (3.3) with $\boldsymbol{S} = \boldsymbol{\Sigma}(\boldsymbol{\lambda} - \boldsymbol{\Sigma}^{-1}\boldsymbol{\gamma})$, we get

$$\frac{\partial^2 M_{\mathrm{X}}(\boldsymbol{t})}{\partial \boldsymbol{t} \partial \boldsymbol{t}'} = \frac{1}{\Phi(\delta_0)}\left[-\frac{1}{\sqrt{k}} \boldsymbol{S} exp\left(\frac{1}{2}\boldsymbol{t}'\boldsymbol{\Sigma t}\right)\left(\frac{\lambda_0 + \boldsymbol{S}'\boldsymbol{t}}{\sqrt{k}}\right)\left(\frac{1}{\sqrt{k}}\boldsymbol{S}'\right)\varphi\left(\frac{\lambda_0 + \boldsymbol{S}'\boldsymbol{t}}{\sqrt{k}}\right)\right] +$$

$$\varphi\left(\frac{\lambda_0 + \boldsymbol{S}'\boldsymbol{t}}{\sqrt{k}}\right)\frac{1}{\sqrt{k}}\boldsymbol{S t}'\boldsymbol{\Sigma} exp\left(\frac{1}{2}\boldsymbol{t}'\boldsymbol{\Sigma t}\right) + \boldsymbol{\Sigma t} exp\left(\frac{1}{2}\boldsymbol{t}'\boldsymbol{\Sigma t}\right)\frac{1}{\sqrt{k}}\boldsymbol{S}'\varphi\left(\frac{\lambda_0 + \boldsymbol{S}'\boldsymbol{t}}{\sqrt{k}}\right) +$$

$$\Phi\left(\frac{\lambda_0 + \boldsymbol{S}'\boldsymbol{t}}{\sqrt{k}}\right)\left(\boldsymbol{\Sigma t t}'\boldsymbol{\Sigma} exp\left(\frac{1}{2}\boldsymbol{t}'\boldsymbol{\Sigma t}\right) + \boldsymbol{\Sigma} exp\left(\frac{1}{2}\boldsymbol{t}'\boldsymbol{\Sigma t}\right)\right).$$

Hence

$$E(\boldsymbol{XX'}) = \frac{\partial^2 M_{\mathrm{X}}(\boldsymbol{0})}{\partial \boldsymbol{t} \partial \boldsymbol{t}'} = \frac{1}{\Phi(\delta_0)}\left[-\frac{1}{\sqrt{k}}\left(\frac{\lambda_0}{\sqrt{k}\sqrt{k}}\right)\varphi\left(\frac{\lambda_o}{\sqrt{k}}\right)\boldsymbol{SS}' + \Phi\left(\frac{\lambda_0}{\sqrt{k}}\right)\boldsymbol{\Sigma}\right].$$

Since $\delta_0 = \lambda_0 / \sqrt{k}$, then we get

$$E(\boldsymbol{XX'}) = \boldsymbol{\Sigma} - \frac{\delta_o \varphi(\delta_o)}{k\Phi(\delta_o)}\boldsymbol{SS}'.$$

To complete the proof, we use part one of Theorem (3.3.3) to write

$$E(\boldsymbol{X}) = \frac{\varphi(\delta_0)}{\Phi(\delta_0)}\left(\frac{\boldsymbol{\Sigma}(\boldsymbol{\lambda} - \boldsymbol{\Sigma}^{-1}\boldsymbol{\gamma})}{\sqrt{\sigma^2 - 2\boldsymbol{\lambda}'\boldsymbol{\gamma} + \boldsymbol{\lambda}'\boldsymbol{\Sigma\lambda}}}\right),$$

$$= \frac{\varphi(\delta_0)}{\sqrt{k}\Phi(\delta_0)}\boldsymbol{S}.$$

Therefore

$$E(X') = \frac{\varphi(\delta_0)}{\sqrt{k}\Phi(\delta_0)} S'.$$

Hence

$$Cov(X) = E(XX') - E(X)E(X'),$$

$$= \Sigma - \frac{\delta_0\varphi(\delta_0)}{k\Phi(\delta_0)} SS' - \frac{\varphi(\delta_0)}{\sqrt{k}\Phi(\delta_0)} S \frac{\varphi(\delta_0)}{\sqrt{k}\Phi(\delta_0)} S',$$

$$= \Sigma - \frac{\delta_0\varphi(\delta_0)}{k\Phi(\delta_0)} SS' - \left(\frac{\varphi(\delta_0)}{\sqrt{k}\Phi(\delta_0)}\right)^2 SS'.$$

More simplification leads to

$$Cov(X) = \Sigma - h\Sigma^2.$$

where

$$h = \left(\left(\frac{\varphi(\delta_0)}{\sqrt{k}\Phi(\delta_0)}\right)^2 + \frac{\delta_0\varphi(\delta_0)}{k\Phi(\delta_0)}\right)(\lambda' - \gamma'\Sigma^{-1})(\lambda - \Sigma^{-1}\gamma).$$

The two results in Theorem (3.3.2) and Theorem (3.3.3) are extended to the location-scale generalized skew-normal distribution.

Theorem 3.3.4. The mgf of $Y \sim GSN_n^2\left(\mu, \Omega^{\frac{1}{2}}\Sigma\Omega^{\frac{1}{2}}, \lambda_0, \sigma^2, \lambda, \gamma, \Sigma\right)$ is

$$M_Y(t) = \frac{exp\left(t'\mu + \frac{1}{2}t'\Omega^{\frac{1}{2}}\Sigma\Omega^{\frac{1}{2}}t\right)}{\Phi(\delta_0)}\Phi\left(\frac{\lambda_0 + (\lambda' - \gamma'\Sigma^{-1})\Sigma\Omega^{\frac{1}{2}}t}{\sqrt{\sigma^2 - 2\lambda'\gamma + \lambda'\Sigma\lambda}}\right).$$

Proof: Since $Y = \mu + \Omega^{\frac{1}{2}}X$, where $X \sim GSN_n^1(\lambda_0, \sigma^2, \lambda, \gamma, \Sigma)$, then

$$M_Y(t) = E\big(exp(t'Y)\big) = exp(t'\mu)M_X\left(\Omega^{\frac{1}{2}}t\right).$$

Since

$$M_X(t) = \frac{exp\left(\frac{1}{2}t'\Sigma t\right)}{\Phi(\delta_0)}\Phi\left(\frac{\lambda_0 + (\lambda' - \gamma'\Sigma^{-1})\Sigma t}{\sqrt{\sigma^2 - 2\lambda'\gamma + \lambda'\Sigma\lambda}}\right),$$

then

$$M_X\left(\Omega^{\frac{1}{2}}t\right) = \frac{exp\left(\frac{1}{2}t'\Omega^{\frac{1}{2}}\Sigma\Omega^{\frac{1}{2}}t\right)}{\Phi(\delta_0)}\Phi\left(\frac{\lambda_0 + (\lambda' - \gamma'\Sigma^{-1})\Sigma\Omega^{\frac{1}{2}}t}{\sqrt{\sigma^2 - 2\lambda'\gamma + \lambda'\Sigma\lambda}}\right).$$

Finally, we get

$$M_Y(t) = \frac{exp\left(t'\mu + \frac{1}{2}t'\Omega^{\frac{1}{2}}\Sigma\Omega^{\frac{1}{2}}t\right)}{\Phi(\delta_0)}\Phi\left(\frac{\lambda_0 + (\lambda' - \gamma'\Sigma^{-1})\Sigma\Omega^{\frac{1}{2}}t}{\sqrt{\sigma^2 - 2\lambda'\gamma + \lambda'\Sigma\lambda}}\right).$$

Theorem 3.3.5: If $Y \sim GSN_n^2\left(\mu, \Omega^{\frac{1}{2}}\Sigma\Omega^{\frac{1}{2}}, \lambda_0, \sigma^2, \lambda, \gamma, \Sigma\right)$, then

1. $E(Y) = \mu + \frac{\varphi(\delta_0)}{\Phi(\delta_0)}\left(\frac{\Omega^{\frac{1}{2}}\Sigma(\lambda - \Sigma^{-1}\gamma)}{\sqrt{\sigma^2 - 2\lambda'\gamma + \lambda'\Sigma\lambda}}\right).$

2. $Cov(Y) = \Omega^{\frac{1}{2}}(\Sigma - h\Sigma^2)\Omega^{\frac{1}{2}}$

where h and k are as given in Theorem (3.3.3).

Proof of 1: Since $Y = \mu + \Omega^{\frac{1}{2}}X$, where $X \sim GSN_n^1(\lambda_0, \sigma^2, \lambda, \gamma, \Sigma)$, then

$$E(Y) = E\left(\mu + \Omega^{\frac{1}{2}}X\right) = \mu + \Omega^{\frac{1}{2}}E(X),$$

$$= \mu + \Omega^{\frac{1}{2}}\frac{\varphi(\delta_0)}{\Phi(\delta_0)}\left(\frac{\Sigma(\lambda - \Sigma^{-1}\gamma)}{\sqrt{\sigma^2 - 2\lambda'\gamma + \lambda'\Sigma\lambda}}\right),$$

$$= \mu + \frac{\varphi(\delta_0)}{\Phi(\delta_0)}\left(\frac{\Omega^{\frac{1}{2}}\Sigma(\lambda - \Sigma^{-1}\gamma)}{\sqrt{\sigma^2 - 2\lambda'\gamma + \lambda'\Sigma\lambda}}\right).$$

Proof of 2: Since $Y = \mu + \Omega^{\frac{1}{2}}X$, where $X \sim GSN_n^1(\lambda_0, \sigma^2, \lambda, \gamma, \Sigma)$, then

$$Cov(Y) = Cov\left(\mu + \Omega^{\frac{1}{2}}X\right) = \Omega^{\frac{1}{2}\prime}Cov(X)\Omega^{\frac{1}{2}}.$$

Since $\Omega^{\frac{1}{2}}$ was considered as a symmetric matrix, i.e. $\Omega^{\frac{1}{2}\prime} = \Omega^{\frac{1}{2}}$, then

$$Cov(Y) = \Omega^{\frac{1}{2}}Cov(X)\Omega^{\frac{1}{2}}.$$

Using Theorem (3.3.3), we conclude

$$Cov(Y) = \Omega^{\frac{1}{2}}(\Sigma - h\Sigma^2)\Omega^{\frac{1}{2}}.$$

3.4 Elliptical Distribution for the GSN

In this section, we introduce a new parameter $\boldsymbol{\alpha}$; which is derived from the skew-elliptical distributions where the generalized skew-normal distribution is a particular case. Moreover, we use that parameter to extend the result of Vernic (2006); some of properties for the elliptical generalized skew normal distribution will be shown.

Theorem 3.4.1. Let $Y \sim GSN_n^2\left(\mu, \Omega^{\frac{1}{2}}\Sigma\Omega^{\frac{1}{2}}, \lambda_0, \sigma^2, \lambda, \gamma, \Sigma\right)$, if we define the parameter $\boldsymbol{\alpha}$ as

$$\alpha = \frac{\Omega^{\frac{1}{2}}\Sigma(\lambda - \Sigma^{-1}\gamma)}{\sqrt{\sigma^2 - 2\lambda'\gamma + \lambda'\Sigma\lambda}},$$

then, the pdf of Y can be written using $\boldsymbol{\alpha}$ as:

$$f_Y(y) = \frac{\varphi_n(y; \mu, \Theta)}{\Phi(\delta_0)} \Phi\left(\frac{\delta_0 + \alpha'\Theta^{-1}(y - \mu)}{\sqrt{1 - \alpha'\Theta^{-1}\alpha}}\right),$$

where $\Theta = \Omega^{\frac{1}{2}}\Sigma\Omega^{\frac{1}{2}}$.

Proof. Using the definition of $\boldsymbol{\alpha}$ and write it as follows:

$$\alpha = \frac{\Omega^{\frac{1}{2}}\Sigma^{\frac{1}{2}}\left(\Sigma^{\frac{1}{2}}\lambda - \Sigma^{-\frac{1}{2}}\gamma\right)}{\sqrt{\sigma^2 - 2\lambda'\gamma + \lambda'\Sigma\lambda}}.$$

If we set $\theta = \Sigma^{\frac{1}{2}}\lambda - \Sigma^{-\frac{1}{2}}\gamma$, then

$$\sigma^2 - 2\lambda'\gamma + \lambda'\Sigma\lambda = \theta'\theta + \sigma^{*2},$$

where

$$\sigma^{*2} = \sigma^2 - \gamma'\Sigma^{-1}\gamma.$$

Hence

$$\alpha = \frac{\Omega^{\frac{1}{2}}\Sigma^{\frac{1}{2}}\theta}{\sqrt{\theta'\theta + \sigma^{*2}}}$$

Multiplying both sides by $\Sigma^{-\frac{1}{2}}\Omega^{-\frac{1}{2}}$ yields

$$\Sigma^{-\frac{1}{2}}\Omega^{-\frac{1}{2}}\alpha = \frac{\theta}{\sqrt{\theta'\theta + \sigma^{*2}}},$$

Squaring both sides yields

$$\left(\alpha'\Omega^{-\frac{1}{2}}\Sigma^{-\frac{1}{2}}\right)\left(\Sigma^{-\frac{1}{2}}\Omega^{-\frac{1}{2}}\alpha\right) = \frac{\theta'\theta}{\theta'\theta + \sigma^{*2}}$$

By association property of matrix multiplication, we get

$$\alpha'\Omega^{-\frac{1}{2}}\Sigma^{-1}\Omega^{-\frac{1}{2}}\alpha = \frac{\theta'\theta}{\theta'\theta + \sigma^{*2}}.$$

So we have

$$\alpha'\left(\Omega^{\frac{1}{2}}\Sigma\Omega^{\frac{1}{2}}\right)^{-1}\alpha = \frac{\theta'\theta}{\theta'\theta + \sigma^{*2}}.$$

Multiplying both sides by $\theta'\theta + \sigma^{*2}$ yields

$$\theta'\theta\alpha'\left(\Omega^{\frac{1}{2}}\Sigma\Omega^{\frac{1}{2}}\right)^{-1}\alpha + \sigma^{*2}\alpha'\left(\Omega^{\frac{1}{2}}\Sigma\Omega^{\frac{1}{2}}\right)^{-1}\alpha = \theta'\theta$$

More simplification leads to

$$\sigma^{*2}\alpha'\left(\Omega^{\frac{1}{2}}\Sigma\Omega^{\frac{1}{2}}\right)^{-1}\alpha = \left(1-\alpha'\left(\Omega^{\frac{1}{2}}\Sigma\Omega^{\frac{1}{2}}\right)^{-1}\alpha\right)\theta'\theta.$$

Finally we have

$$\theta'\theta = \frac{\sigma^{*2}\alpha'\left(\Omega^{\frac{1}{2}}\Sigma\Omega^{\frac{1}{2}}\right)^{-1}\alpha}{1-\alpha'\left(\Omega^{\frac{1}{2}}\Sigma\Omega^{\frac{1}{2}}\right)^{-1}\alpha} \qquad (3.4)$$

By noting that

$$\lambda = \Sigma^{-1}\Omega^{-\frac{1}{2}}\alpha\sqrt{\theta'\theta+\sigma^{*2}}+\Sigma^{-1}\gamma,$$

and substituting the value of $\theta'\theta$ in (3.4) , we reach to the following value of λ

$$\lambda = \Sigma^{-1}\Omega^{-\frac{1}{2}}\alpha\sqrt{\sigma^{*2}+\frac{\sigma^{*2}\alpha'\left(\Omega^{\frac{1}{2}}\Sigma\Omega^{\frac{1}{2}}\right)^{-1}\alpha}{1-\alpha'\left(\Omega^{\frac{1}{2}}\Sigma\Omega^{\frac{1}{2}}\right)^{-1}\alpha}}+\Sigma^{-1}\gamma,$$

$$= \Sigma^{-1}\Omega^{-\frac{1}{2}}\alpha\frac{\sigma^{*}}{\sqrt{1-\alpha'\left(\Omega^{\frac{1}{2}}\Sigma\Omega^{\frac{1}{2}}\right)^{-1}\alpha}}+\Sigma^{-1}\gamma.$$

Now we write λ_o in term of α. The equation

$$\delta_0 = \frac{\lambda_0}{\sqrt{\sigma^2-2\lambda'\gamma+\lambda'\Sigma\lambda}},$$

implies

$$\lambda_0 = \delta_0\sqrt{\boldsymbol{\theta}'\boldsymbol{\theta} + \sigma^{*2}},$$

$$= \delta_0 \frac{\sigma^*}{\sqrt{1 - \boldsymbol{\alpha}'\left(\boldsymbol{\Omega}^{\frac{1}{2}}\boldsymbol{\Sigma}\boldsymbol{\Omega}^{\frac{1}{2}}\right)^{-1}\boldsymbol{\alpha}}}.$$

So we have

$$\boldsymbol{\lambda} = \sigma^* \frac{\boldsymbol{\Sigma}^{-1}\boldsymbol{\Omega}^{-\frac{1}{2}}\boldsymbol{\alpha}}{\sqrt{1 - \boldsymbol{\alpha}'\left(\boldsymbol{\Omega}^{\frac{1}{2}}\boldsymbol{\Sigma}\boldsymbol{\Omega}^{\frac{1}{2}}\right)^{-1}\boldsymbol{\alpha}}} + \boldsymbol{\Sigma}^{-1}\boldsymbol{\gamma} \text{ and } \lambda_0 = \frac{\delta_0\sigma^*}{\sqrt{1 - \boldsymbol{\alpha}'\left(\boldsymbol{\Omega}^{\frac{1}{2}}\boldsymbol{\Sigma}\boldsymbol{\Omega}^{\frac{1}{2}}\right)^{-1}\boldsymbol{\alpha}}}.$$

By substituting the values of $\boldsymbol{\lambda}$ and λ_0 in the term

$$\frac{\lambda_0 + (\boldsymbol{\lambda}' - \boldsymbol{\gamma}'\boldsymbol{\Sigma}^{-1})\boldsymbol{\Omega}^{-\frac{1}{2}}(\boldsymbol{y} - \boldsymbol{\mu})}{\sqrt{\sigma^2 - \boldsymbol{\gamma}'\boldsymbol{\Sigma}^{-1}\boldsymbol{\gamma}}},$$

then

$$\frac{\lambda_0 + (\boldsymbol{\lambda}' - \boldsymbol{\gamma}'\boldsymbol{\Sigma}^{-1})\boldsymbol{\Omega}^{-\frac{1}{2}}(\boldsymbol{y} - \boldsymbol{\mu})}{\sqrt{\sigma^2 - \boldsymbol{\gamma}'\boldsymbol{\Sigma}^{-1}\boldsymbol{\gamma}}} = \frac{\delta_0}{\sigma^*}\frac{\sigma^*}{\sqrt{1 - \boldsymbol{\alpha}'\left(\boldsymbol{\Omega}^{\frac{1}{2}}\boldsymbol{\Sigma}\boldsymbol{\Omega}^{\frac{1}{2}}\right)^{-1}\boldsymbol{\alpha}}} +$$

$$\frac{1}{\sigma^*}\left(\boldsymbol{\Sigma}^{-1}\boldsymbol{\Omega}^{-\frac{1}{2}}\boldsymbol{\alpha}\sqrt{\boldsymbol{\theta}'\boldsymbol{\theta} + \sigma^{*2}}\right)'\boldsymbol{\Omega}^{-\frac{1}{2}}(\boldsymbol{y} - \boldsymbol{\mu}),$$

$$= \frac{\delta_0 + \alpha'\left(\Omega^{\frac{1}{2}}\Sigma\Omega^{\frac{1}{2}}\right)^{-1}(y-\mu)}{\sqrt{1-\alpha'\left(\Omega^{\frac{1}{2}}\Sigma\Omega^{\frac{1}{2}}\right)^{-1}\alpha}}.$$

Now we write the pdf of $\mathbf{Y}$ using the new parameterization, i.e., using $\boldsymbol{\alpha}$ as:

$$f_Y(y)$$

$$= \frac{exp\left(-\frac{1}{2}\left(\Omega^{-\frac{1}{2}}(y-\mu)\right)'\Sigma^{-1}\left(\Omega^{-\frac{1}{2}}(y-\mu)\right)\right)}{(2\pi)^{\frac{n}{2}}\Phi(\delta_0)|\Sigma\Omega|^{\frac{1}{2}}}\Phi\left(\frac{\delta_0 + \alpha'\left(\Omega^{\frac{1}{2}}\Sigma\Omega^{\frac{1}{2}}\right)^{-1}(y-\mu)}{\sqrt{1-\alpha'\left(\Omega^{\frac{1}{2}}\Sigma\Omega^{\frac{1}{2}}\right)^{-1}\alpha}}\right),$$

$$= \frac{exp\left(-\frac{1}{2}(y-\mu)'\Omega^{-\frac{1}{2}}\Sigma^{-1}\Omega^{-\frac{1}{2}}(y-\mu)\right)}{(2\pi)^{\frac{n}{2}}\Phi(\delta_0)|\Sigma\Omega|^{\frac{1}{2}}}\Phi\left(\frac{\delta_0 + \alpha'\left(\Omega^{\frac{1}{2}}\Sigma\Omega^{\frac{1}{2}}\right)^{-1}(y-\mu)}{\sqrt{1-\alpha'\left(\Omega^{\frac{1}{2}}\Sigma\Omega^{\frac{1}{2}}\right)^{-1}\alpha}}\right).$$

If $\Omega^{\frac{1}{2}}\Sigma\Omega^{\frac{1}{2}} = \Theta$, then we write $f_Y(y)$ as follows:

$$f_Y(y) = \frac{exp\left(-\frac{1}{2}(y-\mu)'\Theta^{-1}(y-\mu)\right)}{(2\pi)^{\frac{n}{2}}\Phi(\delta_0)|\Theta|^{\frac{1}{2}}}\Phi\left(\frac{\delta_0 + \alpha'\Theta^{-1}(y-\mu)}{\sqrt{1-\alpha'\Theta^{-1}\alpha}}\right),$$

$$= \frac{\varphi_n(y;\mu,\Theta)}{\Phi(\delta_0)}\Phi\left(\frac{\delta_0 + \alpha'\Theta^{-1}(y-\mu)}{\sqrt{1-\alpha'\Theta^{-1}\alpha}}\right).$$

We use the notation $Y \sim GSN_n^3(\mu, \delta_0, \alpha, \Theta)$ to mean that Y has elliptical generalized skew-normal distribution with parameters μ, δ_0, α and Θ.

Remark: It is easy to see that if $\Sigma^{-1}\gamma = \lambda$, which is equivalent to $\alpha = 0$, then we obtain the density of the multivariate normal distribution i.e. $N_n(\mu, \Theta)$.

The elliptical generalized skew-normal distribution has some properties, which are presented in the following theorems.

Theorem 3.4.2. The moment generating function (mgf) of $Y \sim GSN_n^3(\mu, \delta_0, \alpha, \Theta)$ is

$$M_Y(t) = \frac{\Phi(\delta_0 + \alpha' t)}{\Phi(\delta_0)} exp\left(t'\mu + \frac{t'\Theta t}{2}\right).$$

Proof. Using $Y = \mu + \Omega^{\frac{1}{2}} X$, where $X \sim GSN_n^1(\lambda_0, \sigma^2, \lambda, \gamma, \Sigma)$, then

$$M_Y(t) = exp(t'\mu) M_X\left(\Omega^{\frac{1}{2}} t\right),$$

Since

$$M_X(t) = \frac{exp\left(\frac{1}{2} t'\Sigma t\right)}{\Phi(\delta_0)} \Phi\left(\frac{\lambda_0 + (\lambda' - \gamma'\Sigma^{-1})\Sigma t}{\sqrt{\sigma^2 - 2\lambda'\gamma + \lambda'\Sigma\lambda}}\right),$$

then

$$M_X\left(\Omega^{\frac{1}{2}} t\right) = \frac{exp\left(\frac{1}{2} t'\Omega^{\frac{1}{2}}\Sigma\Omega^{\frac{1}{2}} t\right)}{\Phi(\delta_0)} \Phi\left(\frac{\lambda_0 + (\lambda' - \gamma'\Sigma^{-1})\Sigma\Omega^{\frac{1}{2}} t}{\sqrt{\sigma^2 - 2\lambda'\gamma + \lambda'\Sigma\lambda}}\right),$$

$$= \frac{exp\left(\frac{1}{2} t'\Omega^{\frac{1}{2}}\Sigma\Omega^{\frac{1}{2}} t\right)}{\Phi(\delta_0)} \Phi\left(\frac{\lambda_0}{\sqrt{\sigma^2 - 2\lambda'\gamma + \lambda'\Sigma\lambda}} + \frac{(\lambda' - \gamma'\Sigma^{-1})\Sigma\Omega^{\frac{1}{2}} t}{\sqrt{\sigma^2 - 2\lambda'\gamma + \lambda'\Sigma\lambda}}\right),$$

$$= \Phi(\delta_0 + \boldsymbol{\alpha}'\boldsymbol{t}) \frac{exp\left(\frac{1}{2}\boldsymbol{t}'\boldsymbol{\Omega}^{\frac{1}{2}}\boldsymbol{\Sigma}\boldsymbol{\Omega}^{\frac{1}{2}}\boldsymbol{t}\right)}{\Phi(\delta_0)}.$$

So

$$M_Y(\boldsymbol{t}) = \frac{\Phi(\delta_0 + \boldsymbol{\alpha}'\boldsymbol{t})}{\Phi(\delta_0)} exp\left(\boldsymbol{t}'\boldsymbol{\mu} + \frac{\boldsymbol{t}'\boldsymbol{\Omega}^{\frac{1}{2}}\boldsymbol{\Sigma}\boldsymbol{\Omega}^{\frac{1}{2}}\mathbf{t}}{2}\right),$$

$$= \frac{\Phi(\delta_0 + \boldsymbol{\alpha}'\boldsymbol{t})}{\Phi(\delta_0)} exp\left(\boldsymbol{t}'\boldsymbol{\mu} + \frac{\boldsymbol{t}'\boldsymbol{\Theta}\mathbf{t}}{2}\right).$$

Theorem 3.4.3. If $Y \sim GSN_n^3(\boldsymbol{\mu}, \delta_0, \boldsymbol{\alpha}, \boldsymbol{\Theta})$, then

$$E(Y) = \boldsymbol{\mu} + \boldsymbol{\alpha}\frac{\varphi(\delta_0)}{\Phi(\delta_0)}.$$

Proof: Since $\boldsymbol{Y} = \boldsymbol{\mu} + \boldsymbol{\Omega}^{\frac{1}{2}}\mathbf{X}$, then

$$E(Y) = \boldsymbol{\mu} + \boldsymbol{\Omega}^{\frac{1}{2}}E(\mathbf{X}) = \boldsymbol{\mu} + \boldsymbol{\Omega}^{\frac{1}{2}} \frac{\boldsymbol{\Sigma}(\boldsymbol{\lambda} - \boldsymbol{\Sigma}^{-1}\boldsymbol{\gamma})}{\sqrt{\sigma^2 - 2\boldsymbol{\lambda}'\boldsymbol{\gamma} + \boldsymbol{\lambda}'\boldsymbol{\Sigma}\boldsymbol{\lambda}}} \frac{\varphi(\delta_0)}{\Phi(\delta_0)},$$

$$= \boldsymbol{\mu} + \boldsymbol{\alpha}\frac{\varphi(\delta_0)}{\Phi(\delta_0)}.$$

Theorem 3.4.4. Let $\boldsymbol{b}$ be an $m \times 1$ real vector and $\mathbf{C}$ is $m \times n$ matrix of rank m, where $m < n$. If $Y \sim GSN_n^3(\boldsymbol{\mu}, \delta_0, \boldsymbol{\alpha}, \boldsymbol{\Theta})$, then

$$\boldsymbol{b} + \mathbf{C}Y \sim GSN_m^3(\boldsymbol{b} + \mathbf{C}\boldsymbol{\mu}, \delta_0, \mathbf{C}\boldsymbol{\alpha}, \mathbf{C}\boldsymbol{\Theta}\mathbf{C}').$$

Proof. We will find the mgf of $\boldsymbol{b} + \mathbf{C}Y$. Since $Y \sim GSN_n^3(\boldsymbol{\mu}, \delta_0, \boldsymbol{\alpha}, \boldsymbol{\Theta})$, then

$$M_{b+\mathrm{C}Y}(t) = exp(\boldsymbol{t}'\boldsymbol{b})M_Y(\mathbf{C}'\boldsymbol{t}),$$

$$= exp(\boldsymbol{t}'\boldsymbol{b})exp\left(\boldsymbol{t}'\boldsymbol{C}\boldsymbol{\mu} + \frac{\boldsymbol{t}'\mathbf{C}\boldsymbol{\Theta}\mathbf{C}'\boldsymbol{t}}{2}\right)\frac{\Phi(\delta_0 + \boldsymbol{\alpha}'\mathbf{C}'\boldsymbol{t})}{\Phi(\delta_0)},$$

$$= exp\left(\boldsymbol{t}'(\boldsymbol{b} + \boldsymbol{C}\boldsymbol{\mu}) + \frac{\boldsymbol{t}'\mathbf{C}\boldsymbol{\Theta}\mathbf{C}'\boldsymbol{t}}{2}\right)\frac{\Phi(\delta_0 + (\mathbf{C}\boldsymbol{\alpha})'\boldsymbol{t})}{\Phi(\delta_0)},$$

this leads to

$$\boldsymbol{b} + \mathbf{C}\boldsymbol{Y} \sim GSN_m^3(\boldsymbol{b} + \mathbf{C}\boldsymbol{\mu}, \delta_0, \mathbf{C}\boldsymbol{\alpha}, \mathbf{C}\boldsymbol{\Theta}\mathbf{C}').$$

In addition to these results, the elliptical generalized skew-normal distribution can be characterized via the one dimensional elliptical generalized skew-normal distributions.

Theorem 3.4.5. (Characterization of the GSN Distribution)

The vector $\boldsymbol{Y} \sim GSN_n^3(\boldsymbol{\mu}, \delta_0, \boldsymbol{\alpha}, \boldsymbol{\Theta})$ if, and only if, $\boldsymbol{a}'\boldsymbol{Y} \sim GSN_1^3(\mu_a, \delta_0, \boldsymbol{a}'\boldsymbol{\alpha}, \Theta_a)$, for every $\boldsymbol{a}$, a non zero vector in $\mathbb{R}^n$, where $\mu_a = \boldsymbol{a}'\boldsymbol{\mu}$ and $\Theta_a = \boldsymbol{a}'\boldsymbol{\Theta}\boldsymbol{a}$.

Proof. We will use the method of mgf. Since

$$M_Y(t) = \frac{\Phi(\delta_0 + \boldsymbol{\alpha}'\boldsymbol{t})}{\Phi(\delta_0)} exp\left(\boldsymbol{t}'\boldsymbol{\mu} + \frac{1}{2}\boldsymbol{t}'\boldsymbol{\Theta}\boldsymbol{t}\right),$$

then

$$M_{a'Y}(t) = E\big(exp(ta'Y)\big) = E\left(exp\big((at)'Y\big)\right)$$

$$= M_Y(\boldsymbol{a}\boldsymbol{t}),$$

$$= \frac{\Phi(\delta_0 + \alpha' a t)}{\Phi(\delta_0)} exp\left(t a' \boldsymbol{\mu} + \frac{1}{2} t^2 \boldsymbol{a}' \boldsymbol{\Theta} \boldsymbol{a}\right).$$

If $\mu_a = \boldsymbol{a}'\boldsymbol{\mu}$ and $\Theta_a = \boldsymbol{a}'\boldsymbol{\Theta}\boldsymbol{a}$, then

$$\boldsymbol{a}'\boldsymbol{Y} \sim GSN_1^3(\mu_a, \delta_0, \boldsymbol{a}'\boldsymbol{\alpha}, \Theta_a).$$

Conversely, suppose that $\boldsymbol{a}'\boldsymbol{Y} \sim GSN_1^3(\mu_a, \delta_0, \boldsymbol{a}'\boldsymbol{\alpha}, \Theta_a)$, we will show that $\boldsymbol{Y}$ follows

$GSN_n^3(\boldsymbol{\mu}, \delta_0, \boldsymbol{\alpha}, \boldsymbol{\Theta})$.

The m.g.f of $\boldsymbol{a}'\boldsymbol{Y}$ equal

$$M_{\boldsymbol{a}'\boldsymbol{Y}}(t) = M_{\boldsymbol{Y}}(\boldsymbol{a}t).$$

Setting $t = 1$, we get

$$M_{\boldsymbol{Y}}(\boldsymbol{a}t) = M_{\boldsymbol{Y}}(\boldsymbol{a})$$

$$= \frac{\Phi(\delta_0 + \boldsymbol{\alpha}'\boldsymbol{a})}{\Phi(\delta_0)} exp\left(\boldsymbol{a}'\boldsymbol{\mu} + \frac{1}{2}\boldsymbol{a}'\boldsymbol{\Theta}\boldsymbol{a}\right).$$

So, the theorem is proved.

The following theorem gives the distributions of Y_j and $\sum_{i=1}^{n} Y_i$, where

$(Y_1, \ldots, Y_n)' \sim GSN_n^3(\boldsymbol{\mu}, \delta_0, \boldsymbol{\alpha}, \boldsymbol{\Theta})$.

Corollary 3.4.1. Let $\boldsymbol{Y} = (Y_1, \ldots, Y_n)' \sim GSN_n^3(\boldsymbol{\mu}, \delta_0, \boldsymbol{\alpha}, \boldsymbol{\Theta})$, where $\boldsymbol{\Theta} = \boldsymbol{\Omega}^{\frac{1}{2}}\boldsymbol{\Sigma}\boldsymbol{\Omega}^{\frac{1}{2}}$, then

i. $Y_j \sim GSN_1^3(\mu_j, \delta_0, \alpha_j, \sigma_{jj}), \quad j = 1,2, \ldots, n$

where $Y_j = e_j'Y$, $\mu_j = e_j'\boldsymbol{\mu}$, e_j is the j^{th} unit vector in the standard bass of $\mathbb{R}^n$ and $\alpha_j = e_j'\boldsymbol{\alpha}$ and $\sigma_{jj} = e_j'\Theta e_j$.

ii. $S = \sum_{i=1}^{n} Y_i \sim GSN_1^3(\mu_S, \delta_0, \alpha_S, \sigma_S{}^2)$,

where

$\mu_S = \boldsymbol{e}'\boldsymbol{\mu} = \sum_{j=1}^{n} \mu_j$, $\sigma_S{}^2 = \boldsymbol{e}'\Theta\boldsymbol{e} = \sum_{i=1}^{n}\sum_{j=1}^{n} \sigma_{ij}$, σ_{ij} is the (i,j) element of $\boldsymbol{e}'\Theta\boldsymbol{e}$, $\alpha_S = \boldsymbol{e}'\boldsymbol{\alpha}$ and $\boldsymbol{e}$ is the $n \times 1$ vector of one's.

iii. $(Y_i, S) \sim GSN_2^3\left(\begin{pmatrix}\mu_i\\ \mu_S\end{pmatrix}, \delta_0, \begin{pmatrix}\alpha_i\\ \alpha_S\end{pmatrix}, \begin{pmatrix}\sigma_i{}^2 & \sigma_{S,i}\\ \sigma_{S,i} & \sigma_S{}^2\end{pmatrix}\right)$,

where

$$\begin{pmatrix}\sigma_i{}^2 & \sigma_{S,i}\\ \sigma_{S,i} & \sigma_S{}^2\end{pmatrix} = \begin{pmatrix}e_i'\Theta e_i & e_i'\Theta \boldsymbol{e}\\ \boldsymbol{e}'\Theta e_i & \boldsymbol{e}'\Theta\boldsymbol{e}\end{pmatrix} \text{ and } \begin{pmatrix}\alpha_i\\ \alpha_S\end{pmatrix} = \begin{pmatrix}e_i'\boldsymbol{\alpha}\\ \sum_{j=1}^{n} \alpha_j\end{pmatrix}.$$

Proof of (i): We will find the mgf of Y_j. Using the mgf of $\boldsymbol{Y}$ and write it as follows

$$M_Y(t) = exp\left(\sum_{i=1}^{n} t_i\mu_i + \frac{(t_1, t_2 \dots, t_n)\Theta\begin{pmatrix}t_1\\ t_2\\ \vdots\\ t_n\end{pmatrix}}{2}\right)\frac{\Phi\left(\delta_0 + (\alpha_1, \alpha_2 \dots, \alpha_n)\begin{pmatrix}t_1\\ t_2\\ \vdots\\ t_n\end{pmatrix}\right)}{\Phi(\delta_0)},$$

If we set $t_i = 0$ for every $i \neq j$, then

$$M_{Y_j}(t_j) = exp\left(t_j\mu_j + \frac{t_j^2\sigma_{jj}}{2}\right)\frac{\Phi(\delta_0 + \alpha_j t_j)}{\Phi(\delta_0)},$$

hence

$$Y_j \sim GSN_1^3(\mu_j, \delta_0, \alpha_j, \sigma_{jj}).$$

Proof of (ii): We will find the mgf of S.

Not that: $S = \sum_{i=1}^{n} Y_i$, which is equivalent to $S = \boldsymbol{e}'\boldsymbol{Y}$. So

$$M_S(t) = E(exp(tS))$$

$$= E\left(exp\left(t\sum_{i=1}^{n} Y_i\right)\right).$$

Since $\sum_{i=1}^{n} Y_i = \boldsymbol{e}'\boldsymbol{Y}$, then

$$M_S(t) = E\big(exp(t\boldsymbol{e}'\boldsymbol{Y})\big) = M_{\boldsymbol{Y}}(t\boldsymbol{e}),$$

$$= exp\left(t\boldsymbol{e}'\boldsymbol{\mu} + \frac{t'\boldsymbol{e}'\boldsymbol{\Theta}\boldsymbol{e}t}{2}\right)\frac{\Phi(\delta_0 + \boldsymbol{e}'\boldsymbol{\alpha}t)}{\Phi(\delta_0)},$$

$$= exp\left(t\mu_S + \frac{t\sigma_S{}^2 t}{2}\right)\frac{\Phi(\delta_0 + \alpha_S t)}{\Phi(\delta_0)}.$$

Proof of (iii): If we use the mgf method, then

$$M_{Y_i,S}(t) = E\big(exp(t_1Y_i + t_2S)\big),$$

$$= E\left(exp\big(t_1Y_i + t_2(Y_1 + Y_2 + \cdots + Y_{i-1} + Y_i + Y_{i+1} + \cdots + Y_n)\big)\right),$$

$$= E\left(exp\left((t_1+t_2)Y_i + t_2Y_1 + t_2Y_2 + \cdots + t_2Y_{i-1} + t_2Y_{i+1} + \cdots + t_2Y_n\right)\right),$$

$$= M_Y(t_2, t_2, \dots, (t_1+t_2), t_2, \dots, t_2),$$

Let $\boldsymbol{t}^{*\prime} = (t_2, t_2, \dots, (t_1+t_2), t_2, \dots, t_2)$, and $\boldsymbol{\mu}^{*\prime} = (\mu_1, \mu_2, \dots, \mu_n)$, then

$$M_{Y_i,S}(t) = exp\left(\boldsymbol{t}^{*\prime}\begin{pmatrix}\mu_1\\ \mu_2\\ \vdots\\ \mu_n\end{pmatrix} + \frac{\boldsymbol{t}^{*\prime}\Theta\boldsymbol{t}^{*}}{2}\right)\frac{\Phi(\delta_0 + \alpha'\boldsymbol{t}^{*})}{\Phi(\delta_0)},$$

$$= exp\left((t_2\boldsymbol{e}' + t_1\boldsymbol{e}_i')\boldsymbol{\mu} + \frac{(t_2\boldsymbol{e}' + t_1\boldsymbol{e}_i')\Theta(t_2\boldsymbol{e} + t_1\boldsymbol{e}_i)}{2}\right)\times$$

$$\frac{\Phi\left(\delta_0 + (\alpha_1, \alpha_2, \dots, \alpha_n)(t_2\boldsymbol{e} + t_1\boldsymbol{e}_i)\right)}{\Phi(\delta_0)},$$

$$= exp\left(t_2\sum_{j=1}^{n}\mu_j + t_1\mu_i + \frac{\begin{pmatrix}t_1 & t_2\end{pmatrix}\begin{pmatrix}\boldsymbol{e}_i'\Theta\boldsymbol{e}_i & \boldsymbol{e}_i'\Theta\boldsymbol{e}\\ \boldsymbol{e}'\Theta\boldsymbol{e}_i & \boldsymbol{e}'\Theta\boldsymbol{e}\end{pmatrix}\begin{pmatrix}t_1\\ t_2\end{pmatrix}}{2}\right)\times$$

$$\frac{\Phi\left(\delta_0 + t_1\alpha_i + t_2\sum_{j=1}^{n}\alpha_j\right)}{\Phi(\delta_0)}.$$

3.5 Marginal and Conditional Distribution of Elliptical GSN

In this section, we derive the marginal and conditional distributions of a random vector which is distributed according to elliptical generalized skew-normal distribution. The following theorem shows that the elliptical generalized skew-normal distribution is closed under marginal and conditional distributions.

Theorem 3.5.1. Let $\boldsymbol{Y} \sim GSN_n^3(\boldsymbol{\mu}, \delta_0, \boldsymbol{\alpha}, \boldsymbol{\Theta})$ and partition $\boldsymbol{Y}$ into two subvectors $\boldsymbol{Y}_1$ and $\boldsymbol{Y}_2$ with dimensions m and $n-m$, respectively. Also consider the following corresponding partitions

$$\boldsymbol{\Theta} = \begin{array}{cc} \begin{array}{cc} m & n-m \end{array} & \\ \begin{pmatrix} \boldsymbol{\Theta}_{11} & \boldsymbol{\Theta}_{12} \\ \boldsymbol{\Theta}_{21} & \boldsymbol{\Theta}_{22} \end{pmatrix} & \begin{array}{l} m \\ n-m \end{array} \end{array}, \boldsymbol{\alpha} = \begin{pmatrix} \boldsymbol{\alpha}_1 \\ \boldsymbol{\alpha}_2 \end{pmatrix} \begin{array}{l} m \\ n-m \end{array} \text{ and } \boldsymbol{\mu} = \begin{pmatrix} \boldsymbol{\mu}_1 \\ \boldsymbol{\mu}_2 \end{pmatrix} \begin{array}{l} m \\ n-m \end{array}.$$

Then

(i) $\boldsymbol{Y}_1 \sim GSN_m^3(\boldsymbol{\mu}_1, \delta_0, \boldsymbol{\alpha}_1, \boldsymbol{\Theta}_{11})$,

(ii) $(\boldsymbol{Y}_2 | \boldsymbol{Y}_1 = \boldsymbol{y}_1) \sim GSN_{n-m}^3(\boldsymbol{\mu}^*, \delta_0^{**}, \dot{\boldsymbol{\alpha}}, \boldsymbol{\Psi}_{22})$,

where

$$\delta_0^{**} = \frac{\delta_0^*}{\sigma_1} + \frac{\dot{\boldsymbol{\alpha}}'}{\sigma_1} \boldsymbol{\Psi}_{22}^{-1}(\boldsymbol{\mu}^* - \boldsymbol{\mu}_2), \delta_0^* = \delta_0 + (\boldsymbol{\alpha}_1' - \boldsymbol{\alpha}_2' \boldsymbol{\Theta}_{22}^{-1} \boldsymbol{\Theta}_{21}) \boldsymbol{\Psi}_{11}^{-1} \dot{\boldsymbol{y}}_1$$

$$\boldsymbol{\mu}^* = \boldsymbol{\mu}_2 + \boldsymbol{\Theta}_{21} \boldsymbol{\Theta}_{11}^{-1} \dot{\boldsymbol{y}}_1, {\sigma_1}^2 = 1 - \boldsymbol{\alpha}_1' \boldsymbol{\Theta}_{11}^{-1} \boldsymbol{\alpha}_1, \dot{\boldsymbol{y}}_1 = \boldsymbol{y}_1 - \boldsymbol{\mu}_1,$$

$$\boldsymbol{\Psi}_{11} = \boldsymbol{\Theta}_{11} - \boldsymbol{\Theta}_{12} \boldsymbol{\Theta}_{22}^{-1} \boldsymbol{\Theta}_{21} \text{ and } \boldsymbol{\Psi}_{22} = \boldsymbol{\Theta}_{22} - \boldsymbol{\Theta}_{21} \boldsymbol{\Theta}_{11}^{-1} \boldsymbol{\Theta}_{12}.$$

Proof of (i). The density function of $\boldsymbol{Y}$ is given as follows:

$$f(\boldsymbol{y}_1,\boldsymbol{y}_2)=\frac{exp\left(-\frac{1}{2}(\boldsymbol{y}-\boldsymbol{\mu})'\boldsymbol{\Theta}^{-1}(\boldsymbol{y}-\boldsymbol{\mu})\right)}{(2\pi)^{\frac{n}{2}}\Phi(\delta_0)|\boldsymbol{\Theta}|^{\frac{1}{2}}}\Phi\left(\frac{\delta_0+\boldsymbol{\alpha}'\boldsymbol{\Theta}^{-1}(\boldsymbol{y}-\boldsymbol{\mu})}{\sqrt{1-\boldsymbol{\alpha}'\boldsymbol{\Theta}^{-1}\boldsymbol{\alpha}}}\right).$$

To find the marginal pdf of $\boldsymbol{Y}_1$, we integrate $f(\boldsymbol{y}_1,\boldsymbol{y}_2)$ with respect to $\boldsymbol{y}_2$. To this end,

$$f(\boldsymbol{y}_1)=\int f(\boldsymbol{y}_1,\boldsymbol{y}_2)\,d\boldsymbol{y}_2,$$

$$=\int\frac{exp\left(-\frac{1}{2}(\boldsymbol{y}-\boldsymbol{\mu})'\boldsymbol{\Theta}^{-1}(\boldsymbol{y}-\boldsymbol{\mu})\right)}{(2\pi)^{\frac{n}{2}}\Phi(\delta_0)|\boldsymbol{\Theta}|^{\frac{1}{2}}}\Phi\left(\frac{\delta_0+\boldsymbol{\alpha}'\boldsymbol{\Theta}^{-1}(\boldsymbol{y}-\boldsymbol{\mu})}{\sqrt{1-\boldsymbol{\alpha}'\boldsymbol{\Theta}^{-1}\boldsymbol{\alpha}}}\right)d\boldsymbol{y}_2.$$

Using the following

$$\dot{\boldsymbol{y}}_1=\boldsymbol{y}_1-\boldsymbol{\mu}_1,\ \boldsymbol{\Psi}_{11}=\boldsymbol{\Theta}_{11}-\boldsymbol{\Theta}_{12}\boldsymbol{\Theta}_{22}^{-1}\boldsymbol{\Theta}_{21},\ \boldsymbol{\Psi}_{22}=\boldsymbol{\Theta}_{22}-\boldsymbol{\Theta}_{21}\boldsymbol{\Theta}_{11}^{-1}\boldsymbol{\Theta}_{12},$$

$$\dot{\boldsymbol{\alpha}}_2'=\boldsymbol{\alpha}_2'-\boldsymbol{\alpha}_1'\boldsymbol{\Theta}_{11}^{-1}\boldsymbol{\Theta}_{12},\ \delta_0^*=\delta_0+(\boldsymbol{\alpha}_1'-\boldsymbol{\alpha}_2'\boldsymbol{\Theta}_{22}^{-1}\boldsymbol{\Theta}_{21})\boldsymbol{\Psi}_{11}^{-1}\dot{\boldsymbol{y}}_1,\ \sigma_1{}^2=1-\boldsymbol{\alpha}_1'\boldsymbol{\Theta}_{11}^{-1}\boldsymbol{\alpha}_1$$

and using the transformation $\dot{\boldsymbol{y}}_2=\boldsymbol{y}_2-\boldsymbol{\mu}_2$, yields

$$f(\boldsymbol{y}_1)=\frac{exp\left(-\frac{1}{2}\dot{\boldsymbol{y}}_1'\boldsymbol{\Theta}_{11}^{-1}\dot{\boldsymbol{y}}_1\right)}{(2\pi)^{\frac{m}{2}}\Phi(\delta_0)|\boldsymbol{\Theta}|^{\frac{1}{2}}}\int\frac{exp\left(-\frac{1}{2}(\dot{\boldsymbol{y}}_2-\boldsymbol{\Theta}_{21}\boldsymbol{\Theta}_{11}^{-1}\dot{\boldsymbol{y}}_1)'\boldsymbol{\Psi}_{22}^{-1}(\dot{\boldsymbol{y}}_2-\boldsymbol{\Theta}_{21}\boldsymbol{\Theta}_{11}^{-1}\dot{\boldsymbol{y}}_1)\right)}{(2\pi)^{\frac{n-m}{2}}|\boldsymbol{\Psi}_{22}|^{\frac{1}{2}}}\times$$

$$\Phi\left(\frac{\delta_0^*+\dot{\boldsymbol{\alpha}}_2'\boldsymbol{\Psi}_{22}^{-1}\dot{\boldsymbol{y}}_2}{\sqrt{\sigma_1{}^2-\dot{\boldsymbol{\alpha}}_2'\boldsymbol{\Psi}_{22}^{-1}\dot{\boldsymbol{\alpha}}_2}}\right)d\boldsymbol{y}_2,$$

$$=\frac{exp\left(-\frac{1}{2}\dot{\boldsymbol{y}}_1'\boldsymbol{\Theta}_{11}^{-1}\dot{\boldsymbol{y}}_1\right)}{(2\pi)^{\frac{m}{2}}\Phi(\delta_0)|\boldsymbol{\Theta}_{11}|^{\frac{1}{2}}}\int\frac{exp\left(-\frac{1}{2}(\dot{\boldsymbol{y}}_2-\boldsymbol{\Theta}_{21}\boldsymbol{\Theta}_{11}^{-1}\dot{\boldsymbol{y}}_1)'\boldsymbol{\Psi}_{22}^{-1}(\dot{\boldsymbol{y}}_2-\boldsymbol{\Theta}_{21}\boldsymbol{\Theta}_{11}^{-1}\dot{\boldsymbol{y}}_1)\right)}{(2\pi)^{\frac{n-m}{2}}|\boldsymbol{\Psi}_{22}|^{\frac{1}{2}}}\times$$

$$\Phi\left(\frac{\frac{\delta_0^*}{\sigma_1}+\frac{\dot{\boldsymbol{\alpha}}_2'}{\sigma_1}\boldsymbol{\Psi}_{22}^{-1}(\dot{\boldsymbol{y}}_2-\boldsymbol{\Theta}_{21}\boldsymbol{\Theta}_{11}^{-1}\dot{\boldsymbol{y}}_1+\boldsymbol{\Theta}_{21}\boldsymbol{\Theta}_{11}^{-1}\dot{\boldsymbol{y}}_1)}{\sqrt{1-\frac{\dot{\boldsymbol{\alpha}}_2'}{\sigma_1}\boldsymbol{\Psi}_{22}^{-1}\frac{\dot{\boldsymbol{\alpha}}_2}{\sigma_1}}}\right)d\dot{\boldsymbol{y}}_2,$$

$$=\frac{exp\left(-\frac{1}{2}\dot{\boldsymbol{y}}_1'\boldsymbol{\Theta}_{11}^{-1}\dot{\boldsymbol{y}}_1\right)}{(2\pi)^{\frac{m}{2}}\Phi(\delta_0)|\boldsymbol{\Theta}_{11}|^{\frac{1}{2}}}\int\frac{exp\left(-\frac{1}{2}(\dot{\boldsymbol{y}}_2-\boldsymbol{\Theta}_{21}\boldsymbol{\Theta}_{11}^{-1}\dot{\boldsymbol{y}}_1)'\boldsymbol{\Psi}_{22}^{-1}(\dot{\boldsymbol{y}}_2-\boldsymbol{\Theta}_{21}\boldsymbol{\Theta}_{11}^{-1}\dot{\boldsymbol{y}}_1)\right)}{(2\pi)^{\frac{n-m}{2}}|\boldsymbol{\Psi}_{22}|^{\frac{1}{2}}}\times$$

$$\Phi\left(\frac{\delta_0^{**}+\frac{\dot{\boldsymbol{\alpha}}_2'}{\sigma_1}\boldsymbol{\Psi}_{22}^{-1}(\dot{\boldsymbol{y}}_2-\boldsymbol{\Theta}_{21}\boldsymbol{\Theta}_{11}^{-1}\dot{\boldsymbol{y}}_1)}{\sqrt{1-\frac{\dot{\boldsymbol{\alpha}}_2'}{\sigma_1}\boldsymbol{\Psi}_{22}^{-1}\frac{\dot{\boldsymbol{\alpha}}_2}{\sigma_1}}}\right)d\dot{\boldsymbol{y}}_2$$

where

$$\delta_0^{**}=\frac{\delta_0^*}{\sigma_1}+\frac{\dot{\boldsymbol{\alpha}}_2'}{\sigma_1}\boldsymbol{\Psi}_{22}^{-1}\boldsymbol{\Theta}_{21}\boldsymbol{\Theta}_{11}^{-1}\dot{\boldsymbol{y}}_1.$$

Thus

$$f(\boldsymbol{y}_1)=\frac{exp\left(-\frac{1}{2}\dot{\boldsymbol{y}}_1'\boldsymbol{\Theta}_{11}^{-1}\dot{\boldsymbol{y}}_1\right)}{(2\pi)^{\frac{m}{2}}\Phi(\delta_0)|\boldsymbol{\Theta}_{11}|^{\frac{1}{2}}}\Phi(\delta_0^{**}),$$

$$=\frac{exp\left(-\frac{1}{2}\dot{\boldsymbol{y}}_1'\boldsymbol{\Theta}_{11}^{-1}\dot{\boldsymbol{y}}_1\right)}{(2\pi)^{\frac{m}{2}}\Phi(\delta_0)|\boldsymbol{\Theta}_{11}|^{\frac{1}{2}}}\Phi\left(\frac{\delta_0^*}{\sigma_1}+\frac{\dot{\boldsymbol{\alpha}}_2'}{\sigma_1}\boldsymbol{\Psi}_{22}^{-1}\boldsymbol{\Theta}_{21}\boldsymbol{\Theta}_{11}^{-1}\dot{\boldsymbol{y}}_1\right),$$

$$=\frac{exp\left(-\frac{1}{2}(\boldsymbol{y}_1-\boldsymbol{\mu}_1)'\boldsymbol{\Theta}_{11}^{-1}(\boldsymbol{y}_1-\boldsymbol{\mu}_1)\right)}{(2\pi)^{\frac{m}{2}}\Phi(\delta_0)|\boldsymbol{\Theta}_{11}|^{\frac{1}{2}}}\times$$

$$\Phi\left(\frac{\delta_0}{\sigma_1}+\frac{((\boldsymbol{\alpha}_1'-\boldsymbol{\alpha}_2'\boldsymbol{\Theta}_{22}^{-1}\boldsymbol{\Theta}_{21})\boldsymbol{\Psi}_{11}^{-1}\dot{\boldsymbol{y}}_1)}{\sigma_1}+\frac{\dot{\boldsymbol{\alpha}}_2'}{\sigma_1}\boldsymbol{\Psi}_{22}^{-1}\boldsymbol{\Theta}_{21}\boldsymbol{\Theta}_{11}^{-1}\dot{\boldsymbol{y}}_1\right),$$

$$= \frac{exp\left(-\frac{1}{2}(y_1-\mu_1)'\Theta_{11}^{-1}(y_1-\mu_1)\right)}{(2\pi)^{\frac{m}{2}}\Phi(\delta_0)|\Theta_{11}|^{\frac{1}{2}}} \times$$

$$\Phi\left(\frac{\delta_0+\left((\alpha_1'-\alpha_2'\Theta_{22}^{-1}\Theta_{21})\Psi_{11}^{-1}+\alpha_2'\Psi_{22}^{-1}\Theta_{21}\Theta_{11}^{-1}\right)(y_1-\mu_1)}{\sqrt{1-\alpha_1'\Theta_{11}^{-1}\alpha_1}}\right).$$

Note that

$(-\Theta_{22}^{-1}\Theta_{21}\Psi_{11}^{-1})' = -\Psi_{22}^{-1}\Theta_{21}\Theta_{11}^{-1}$, (since Θ^{-1} is symmetric), *implies that*

$$(\alpha_1'-\alpha_2'\Theta_{22}^{-1}\Theta_{21})\Psi_{11}^{-1}+\alpha_2'\Psi_{22}^{-1}\Theta_{21}\Theta_{11}^{-1} = \alpha_1'\Psi_{11}^{-1}-\alpha_2'\Theta_{22}^{-1}\Theta_{21}\Psi_{11}^{-1}+$$

$$(\alpha_2'\Psi_{22}^{-1}\Theta_{21}\Theta_{11}^{-1}-\alpha_1'\Theta_{11}^{-1}\Theta_{12}\Psi_{22}^{-1}\Theta_{21}\Theta_{11}^{-1})$$

$$= \alpha_1'(\Psi_{11}^{-1}-\Theta_{11}^{-1}\Theta_{12}\Psi_{22}^{-1}\Theta_{21}\Theta_{11}^{-1})+$$

$$\alpha_2'(\Psi_{22}^{-1}\Theta_{21}\Theta_{11}^{-1}-\Theta_{22}^{-1}\Theta_{21}\Psi_{11}^{-1}).$$

Since

$(-\Theta_{22}^{-1}\Theta_{21}\Psi_{11}^{-1})' = -\Psi_{22}^{-1}\Theta_{21}\Theta_{11}^{-1}$ and $\Psi_{11}^{-1}-\Theta_{11}^{-1}\Theta_{12}\Psi_{22}^{-1}\Theta_{21}\Theta_{11}^{-1} = \Theta_{11}^{-1}$,

then

$$(\alpha_1'-\alpha_2'\Theta_{22}^{-1}\Theta_{21})\Psi_{11}^{-1}+\alpha_2'\Psi_{22}^{-1}\Theta_{21}\Theta_{11}^{-1} = \alpha_1'\Theta_{11}^{-1}.$$

Hence

$$f(y_1) = \frac{exp\left(-\frac{1}{2}(y_1-\mu_1)'\Theta_{11}^{-1}(y_1-\mu_1)\right)}{(2\pi)^{\frac{m}{2}}\Phi(\delta_0)|\Theta_{11}|^{\frac{1}{2}}}\Phi\left(\frac{\delta_0+\alpha_1'\Theta_{11}^{-1}(y_1-\mu_1)}{\sqrt{1-\alpha_1'\Theta_{11}^{-1}\alpha_1}}\right)$$

Proof of ii. The pdf of $Y_2|Y_1=y_1$ is derived as follows. Since

$$f(y_1, y_2) = \frac{exp\left(-\frac{1}{2}\dot{y}_1'\boldsymbol{\Theta}_{11}^{-1}\dot{y}_1\right)}{(2\pi)^{\frac{n}{2}}\Phi(\delta_0)|\boldsymbol{\Theta}_{11}|^{\frac{1}{2}}} \frac{exp\left(-\frac{1}{2}(\dot{y}_2 - \boldsymbol{\Theta}_{21}\boldsymbol{\Theta}_{11}^{-1}\dot{y}_1)'\boldsymbol{\Psi}_{22}^{-1}(\dot{y}_2 - \boldsymbol{\Theta}_{21}\boldsymbol{\Theta}_{11}^{-1}\dot{y}_1)\right)}{|\boldsymbol{\Psi}_{22}|^{\frac{1}{2}}} \times$$

$$\Phi\left(\frac{\frac{\delta_0^*}{\sigma_1} + \frac{\dot{\alpha}'}{\sigma_1}\boldsymbol{\Psi}_{22}^{-1}\dot{y}_2}{\sqrt{1 - \frac{\dot{\alpha}'}{\sigma_1}\boldsymbol{\Psi}_{22}^{-1}\frac{\dot{\alpha}}{\sigma_1}}}\right),$$

then

$$f(y_2|y_1) \propto \frac{exp\left(-\frac{1}{2}(\dot{y}_2 - \boldsymbol{\Theta}_{21}\boldsymbol{\Theta}_{11}^{-1}\dot{y}_1)'\boldsymbol{\Psi}_{22}^{-1}(\dot{y}_2 - \boldsymbol{\Theta}_{21}\boldsymbol{\Theta}_{11}^{-1}\dot{y}_1)\right)}{|\boldsymbol{\Psi}_{22}|^{\frac{1}{2}}} \Phi\left(\frac{\frac{\delta_0^*}{\sigma_1} + \frac{\dot{\alpha}'}{\sigma_1}\boldsymbol{\Psi}_{22}^{-1}\dot{y}_2}{\sqrt{1 - \frac{\dot{\alpha}'}{\sigma_1}\boldsymbol{\Psi}_{22}^{-1}\frac{\dot{\alpha}}{\sigma_1}}}\right).$$

To do more simplification, let

$$\mu^* = \mu_2 + \boldsymbol{\Theta}_{21}\boldsymbol{\Theta}_{11}^{-1}\dot{y}_1,$$

then

$$\dot{y}_2 - \boldsymbol{\Theta}_{21}\boldsymbol{\Theta}_{11}^{-1}\dot{y}_1 = y_2 - \mu^*.$$

So

$$f(y_2|y_1) \propto exp\left(-\frac{1}{2}(y_2 - \mu^*)'\boldsymbol{\Psi}_{22}^{-1}(y_2 - \mu^*)\right) \Phi\left(\frac{\frac{\delta_0^*}{\sigma_1} + \frac{\dot{\alpha}'}{\sigma_1}\boldsymbol{\Psi}_{22}^{-1}\dot{y}_2}{\sqrt{1 - \frac{\dot{\alpha}'}{\sigma_1}\boldsymbol{\Psi}_{22}^{-1}\frac{\dot{\alpha}}{\sigma_1}}}\right),$$

$$\propto exp\left(-\frac{1}{2}(y_2-\mu^*)'\Psi_{22}^{-1}(y_2-\mu^*)\right)\times$$

$$\Phi\left(\frac{\frac{\delta_0^*}{\sigma_1}+\frac{\dot{\alpha}'}{\sigma_1}\Psi_{22}^{-1}(y_2-\mu_2-\mu^*+\mu^*)}{\sqrt{1-\frac{\dot{\alpha}'}{\sigma_1}\Psi_{22}^{-1}\frac{\dot{\alpha}}{\sigma_1}}}\right),$$

$$\propto exp\left(-\frac{1}{2}(y_2-\mu^*)'\Psi_{22}^{-1}(y_2-\mu^*)\right)\Phi\left(\frac{\delta_0^{**}+\frac{\dot{\alpha}'}{\sigma_1}\Psi_{22}^{-1}(y_2-\mu^*)}{\sqrt{1-\frac{\dot{\alpha}'}{\sigma_1}\Psi_{22}^{-1}\frac{\dot{\alpha}}{\sigma_1}}}\right),$$

where

$$\delta_0^{**}=\frac{\delta_0^*}{\sigma_1}+\frac{\dot{\alpha}'}{\sigma_1}\Psi_{22}^{-1}(\mu^*-\mu_2).$$

So

$$f(y_2|y_1)=\frac{exp\left(-\frac{1}{2}(y_2-\mu^*)'\Psi_{22}^{-1}(y_2-\mu^*)\right)}{(2\pi)^{\frac{n-m}{2}}|\Psi_{22}|^{\frac{1}{2}}\Phi(\delta_0^{**})}\Phi\left(\frac{\delta_0^{**}+\frac{\dot{\alpha}'}{\sigma_1}\Psi_{22}^{-1}(y_2-\mu^*)}{\sqrt{1-\frac{\dot{\alpha}'}{\sigma_1}\Psi_{22}^{-1}\frac{\dot{\alpha}}{\sigma_1}}}\right).$$

Chapter Four

Generalized Skew Gaussian Process

In this chapter, we discuss the theory of the Generalized Skew-Gaussian Process (GSGP). Here we define the generalized multivariate skew-normal random process or field as a generalization to the multivariate Gaussian random process. Moreover, we define the concept of third order stationary GSGP. Moreover, we will show that the GSGP admits a best linear unbiased predictor BLUP, and we derive that BLUP and its confidence prediction using the ordinary Kriging method.

Definition 5. A random process $Y(\boldsymbol{t}), \boldsymbol{t} \in D \subseteq \mathbb{R}^n$, is said to be a (GSGP) if for every n and $\boldsymbol{t}_1, \dots, \boldsymbol{t}_n \in D$, the vector $\left(Y(\boldsymbol{t}_1), \dots, Y(\boldsymbol{t}_n)\right)'$ has an n −dimensional GSN distribution.

Definition 6. A GSGP is weak stationary of the third order if for every n −finite dimensional distribution of the GSGP satisfies the following assumptions

i. $\boldsymbol{\mu} = \mu \mathbf{1}_n,\ \mu \in \mathbb{R}$.

ii. The elements of $\boldsymbol{\Sigma}$ and $\boldsymbol{\Omega}$ are functions of $\boldsymbol{t}_i - \boldsymbol{s}_j$, for all $\boldsymbol{t}_i, \boldsymbol{s}_j \in D$.

iii. $\boldsymbol{\lambda} = \lambda \mathbf{1}_n$ and $\boldsymbol{\gamma} = \gamma \mathbf{1}_n$, for some $\lambda, \gamma \in \mathbb{R}$, i.e. skewness is fixed in each direction.

4.1 Ordinary kriging for GSNRP

In this section, we assume that for a random process $Y(\boldsymbol{t}), \boldsymbol{t} \in D$, $\left(Y(\boldsymbol{t}_1), \ldots, Y(\boldsymbol{t}_n)\right)'$ has a GSN distribution for any choice of n and $\boldsymbol{t}_1, \ldots, \boldsymbol{t}_n \in D$. Our objective is to show that $Y(\boldsymbol{t})$ admits a best linear unbiased predictor BLUP if the sampling points are chosen on a regular polygon. Also we derive the variance and a prediction interval for this predictor. To this end, the following preliminary definitions and theorems are needed.

Definition 7. (Schott, 1997) An $n \times n$ matrix $\mathbf{A}$ is said to be a circulant matrix if each row of $\mathbf{A}$ can be obtained from the previous row by a circular rotation of its elements. We denote a circulant matrix $\mathbf{A}$ by $\mathbf{A} = \text{circ}(a_1, a_2, \ldots, a_n)$. One special circulant matrix is $\Pi_n = circ(0,1,0, \ldots, 0)$.

It is easy to show that the vector of one's, i.e., $\mathbf{1}_n$ is an eigenvector of every circulant matrix of dimension n. (See Schott, 1997)

Theorem 4.1.1. (Schott, 1997) The $n \times n$ matrix $\boldsymbol{\Omega}$ is circulant matrix if and only if

$$\boldsymbol{\Omega} = \Pi_n \boldsymbol{\Omega} \Pi_n'.$$

Note that $\boldsymbol{\Omega} = \mathbf{I}_n$ implies that $\Pi_n' = \Pi_n^{-1}$.

We prove the following lemma which will be used later.

Lemma 4.1.1. If $\boldsymbol{\Omega}$ is a circulant matrix, then $\boldsymbol{\Omega}^{\frac{1}{2}}$ is a circulant matrix.

Proof. Since $\mathbf{\Omega}$ is positive definite matrix, then there exist an orthogonal matrix **P** such that $\mathbf{P}'\mathbf{\Omega}\mathbf{P} = \mathbf{D} = diag(d_1, \ldots, d_n)\, , d_i > 0 \, , \forall\, i$, where $d_1, \ldots, d_n$ are the eigenvalues of $\mathbf{\Omega}$. Also since $\mathbf{\Omega}$ is circulant matrix, then

$$\mathbf{\Omega} = \Pi_n \mathbf{\Omega} \Pi'_n,$$

using the two facts that $\Pi'_n = \Pi_n^{-1}$and $\mathbf{P}' = \mathbf{P}^{-1}$, then

$$\mathbf{PDP}^{-1} = \Pi_n \mathbf{PDP}^{-1} \Pi'_n.$$

Taking the square root of both sides, we get

$$(\mathbf{PDP}^{-1})^{\frac{1}{2}} = (\Pi_n \mathbf{PDP}^{-1} \Pi'_n)^{\frac{1}{2}},$$

$$\mathbf{PD}^{\frac{1}{2}}\mathbf{P}^{-1} = \left((\Pi_n \mathbf{P})\mathbf{D}(\Pi_n \mathbf{P})^{-1}\right)^{\frac{1}{2}},$$

$$\mathbf{PD}^{\frac{1}{2}}\mathbf{P}^{-1} = (\Pi_n \mathbf{P})\mathbf{D}^{\frac{1}{2}}(\Pi_n \mathbf{P})^{-1}.$$

Using the fact that $\Pi'_n = \Pi_n^{-1}$, we get

$$\mathbf{\Omega}^{\frac{1}{2}} = \Pi_n \mathbf{PD}^{\frac{1}{2}}\mathbf{P}^{-1}\Pi_n^{-1},$$

$$= \Pi_n \mathbf{PD}^{\frac{1}{2}}\mathbf{P}^{-1}\Pi'_n$$

$$= \Pi_n \mathbf{\Omega}^{\frac{1}{2}} \Pi'_n.$$

This completes the proof of the lemma.

In this thesis, we consider only isotropic generalized skew-Gaussian process, i.e., for every n –finite dimensional distribution, the elements of $\mathbf{\Sigma}$ and $\mathbf{\Omega}$ are functions of $\|\boldsymbol{t}_i - \boldsymbol{s}_j\|$ for all $\boldsymbol{t}_i, \boldsymbol{s}_j \in D$. Moreover, we assume that the elements of $\mathbf{\Sigma}$ and $\mathbf{\Omega}$ are given in terms of isotropic covariance functions K_1 and K_2, i.e.

$$\mathbf{\Sigma} = \left(K_1\left(\left\|\boldsymbol{t}_i - \boldsymbol{t}_j\right\|\right)\right)_{i,j=1}^{n} \text{ and } \mathbf{\Omega} = \left(K_2\left(\left\|\boldsymbol{t}_i - \boldsymbol{t}_j\right\|\right)\right)_{i,j=1}^{n}.$$

In definition 2, "stationary" is replaced by isotropic if the elements of $\mathbf{\Sigma}$ and $\mathbf{\Omega}$ are functions of $\left\|\boldsymbol{t}_i - \boldsymbol{s}_j\right\|$, for all $\boldsymbol{t}_i, \boldsymbol{s}_j \in D$.

The following Theorem guarantees the existence of a BLUP under a certain condition.

Theorem 4.1.2. Let $Y(\boldsymbol{t})$ be a weak third order isotropic (GSGP), $\boldsymbol{t}_1, \dots, \boldsymbol{t}_n$ and $\boldsymbol{t}_0$ are chosen in D such that they are the vertices of a regular polygon of $n+1$ sides. Let $\boldsymbol{Y} = \left(Y(\boldsymbol{t}_1), \dots, Y(\boldsymbol{t}_n)\right)'$ be the observed values of $Y(\boldsymbol{t})$ at $\boldsymbol{t}_1, \dots, \boldsymbol{t}_n$. Then the process $Y(\boldsymbol{t})$ admits a BLUP for $Y(\boldsymbol{t}_0)$ using the data $\boldsymbol{Y}$.

Proof. Let $\mathbf{\Sigma}$ and $\mathbf{\Omega}$ be the matrices corresponding to the vector $(Y(\boldsymbol{t}_0), \boldsymbol{Y}')'$. Then $\mathbf{\Sigma}\mathbf{1}_{n+1} = \mathrm{c}\mathbf{1}_{n+1}$ and $\mathbf{\Omega}\mathbf{1}_{n+1} = h^*\mathbf{1}_{n+1}$, where c and h^* are, respectively, the corresponding eigen values of $\mathbf{\Sigma}$ and $\mathbf{\Omega}$. Using Theorem (3.3.5), we find that

$$E\begin{pmatrix} Y(\boldsymbol{t}_0) \\ \boldsymbol{Y} \end{pmatrix} = \mu\mathbf{1}_{n+1} + \frac{\varphi(\delta_0)}{\Phi(\delta_0)} \frac{\mathbf{\Omega}^{\frac{1}{2}}\mathbf{\Sigma}(\lambda\mathbf{1}_{n+1} - \mathbf{\Sigma}^{-1}\gamma\mathbf{1}_{n+1})}{\sqrt{\sigma^2 - 2\boldsymbol{\lambda}'\boldsymbol{\gamma} + \boldsymbol{\lambda}'\mathbf{\Sigma}\boldsymbol{\lambda}}},$$

$$= \mu\mathbf{1}_{n+1} + \frac{\varphi(\delta_0)}{\Phi(\delta_0)} \frac{\mathbf{\Omega}^{\frac{1}{2}}(\lambda\mathbf{\Sigma}\mathbf{1}_{n+1} - \gamma\mathbf{1}_{n+1})}{\sqrt{\sigma^2 - 2\boldsymbol{\lambda}'\boldsymbol{\gamma} + \boldsymbol{\lambda}'\mathbf{\Sigma}\boldsymbol{\lambda}}},$$

$$= \mu\mathbf{1}_{n+1} + \mathbf{\Omega}^{\frac{1}{2}} \frac{\varphi(\delta_0)}{\Phi(\delta_0)} \frac{(\lambda\mathrm{c} - \gamma)}{\sqrt{\sigma^2 - 2\boldsymbol{\lambda}'\boldsymbol{\gamma} + \boldsymbol{\lambda}'\mathbf{\Sigma}\boldsymbol{\lambda}}} \mathbf{1}_{n+1},$$

$$= \mu\mathbf{1}_{n+1} + \mathbf{\Omega}^{\frac{1}{2}}\mathrm{c}^*\mathbf{1}_{n+1},$$

where

$$c^* = \frac{\varphi(\delta_0)}{\Phi(\delta_0)} \frac{(\lambda c - \gamma)}{\sqrt{\sigma^2 - 2\boldsymbol{\lambda}'\boldsymbol{\gamma} + \boldsymbol{\lambda}'\Sigma\boldsymbol{\lambda}}}.$$

By Lemma (4.1.1), $\boldsymbol{\Omega}^{\frac{1}{2}}$ is circulant matrix. Hence

$$E\begin{pmatrix} Y(\boldsymbol{t}_0) \\ \boldsymbol{Y} \end{pmatrix} = \mu\mathbf{1}_{n+1} + h^* c^* \mathbf{1}_{n+1} = c^{**}\mathbf{1}_{n+1},$$

where

$$c^{**} = \mu + h^* c^*, \quad h^* \in \mathbb{R}.$$

This means that there exists a BLUP of $Y(\mathbf{t}_0)$. To simplify the calculation, when finding the ordinary kriging predictor $\hat{P}(\boldsymbol{Y}; \boldsymbol{t}_0)$, we assume that $\boldsymbol{\Omega} = \mathbf{I}_n$. If $\boldsymbol{K}$ is the $n \times n$ covariance matrix of $\boldsymbol{Y}$ whose elements $K_{ij} = K_1(\|\boldsymbol{t}_i - \boldsymbol{t}_j\|)$ and $\boldsymbol{m} = m\mathbf{1}_n$, then following Cressie (1993),

$$\hat{P}(Y; \boldsymbol{t}_0) = \hat{\boldsymbol{\eta}}' \boldsymbol{Y},$$

where

$$\hat{\boldsymbol{\eta}} = \mathbf{K}^{-1}(\boldsymbol{k} + \boldsymbol{m}).$$

Using Theorem (3.3.3), we find that

$$\boldsymbol{k}' = \left(K_1(\|\boldsymbol{t}_0 - \boldsymbol{t}_1\|) - h\sum_{j=1}^{n} K_1^2(\|\boldsymbol{t}_0 - \boldsymbol{t}_j\|), \dots, K_1(\|\boldsymbol{t}_0 - \boldsymbol{t}_n\|) - h\sum_{j=1}^{n} K_1^2(\|\boldsymbol{t}_n - \boldsymbol{t}_j\|) \right),$$

and

$$m = \frac{1 - \mathbf{1}_n'(\boldsymbol{\Sigma} - h\boldsymbol{\Sigma}^2)^{-1}\boldsymbol{k}}{\mathbf{1}_n'(\boldsymbol{\Sigma} - h\boldsymbol{\Sigma}^2)^{-1}\mathbf{1}_n}.$$

So

$$\hat{P}(\boldsymbol{Y};\boldsymbol{t}_0) = \left((\boldsymbol{\Sigma} - h\boldsymbol{\Sigma}^2)^{-1}(\boldsymbol{k} + m\mathbf{1}_n)\right)'\boldsymbol{Y}.$$

The variance of the predictor is given as follows

$$\sigma_k^2 = \boldsymbol{k}'(\boldsymbol{\Sigma} - h\boldsymbol{\Sigma}^2)^{-1}\boldsymbol{k} - \frac{(\mathbf{1}_n'(\boldsymbol{\Sigma} - h\boldsymbol{\Sigma}^2)^{-1}\boldsymbol{k} - 1)^2}{\mathbf{1}_n'(\boldsymbol{\Sigma} - h\boldsymbol{\Sigma}^2)^{-1}\mathbf{1}_n}.$$

A $100(1-\alpha)\%$ prediction interval for $Y(\boldsymbol{t}_0)$ can be obtained as follows. We need to find the distribution of $\hat{\boldsymbol{\eta}}'\boldsymbol{Y} - Y(\boldsymbol{t}_0)$. Since

$$\begin{pmatrix} Y(\boldsymbol{t}_0) \\ \boldsymbol{Y} \end{pmatrix} \sim GSN_{n+1}^2(\mu\mathbf{1}_{n+1}, \boldsymbol{\Sigma}, \lambda_0, \sigma^2, \lambda\mathbf{1}_{n+1}, \gamma\mathbf{1}_{n+1}, \boldsymbol{\Sigma}),$$

then

$$\begin{pmatrix} Y(\boldsymbol{t}_0) \\ \boldsymbol{Y} \end{pmatrix} \sim GSN_{n+1}^3(\mu\mathbf{1}_{n+1}, \delta_0, \boldsymbol{\alpha}, \boldsymbol{\Theta}),$$

where

$$\boldsymbol{\alpha} = \frac{\boldsymbol{\Sigma}(\lambda\mathbf{I}_{n+1} - \boldsymbol{\Sigma}^{-1}\gamma)\mathbf{1}_{n+1}}{\sqrt{\sigma^2 - 2\lambda\gamma\mathbf{1}_{n+1}'\mathbf{1}_{n+1} + \lambda^2\mathbf{1}_{n+1}'\mathbf{1}_{n+1}}} \text{ and } \boldsymbol{\Theta} = \boldsymbol{\Sigma}.$$

So, $\hat{\boldsymbol{\eta}}'\boldsymbol{Y} - Y(\boldsymbol{t}_0)$ can be written as follows

$$\hat{\boldsymbol{\eta}}'\boldsymbol{Y} - Y(\boldsymbol{t}_0) = \boldsymbol{w}'\boldsymbol{Y}_*,$$

where

$$\boldsymbol{w}' = (-1 \quad \hat{\boldsymbol{\eta}}') \text{ and } \boldsymbol{Y}_* = \begin{pmatrix} Y(\boldsymbol{t}_0) \\ \boldsymbol{Y} \end{pmatrix}.$$

Using Theorem (3.4.4), it follows that

$$\boldsymbol{w}'Y_* \sim GSN_1^3(0, \delta_0, \alpha_w, \Theta_w),$$

where $\Theta_w = \boldsymbol{w}'\boldsymbol{\Theta}\boldsymbol{w}$ and $\alpha_w = \boldsymbol{w}'\boldsymbol{\alpha}$. Assuming q_α denotes the $100(1-\alpha)$ quantile of $GSN_1^3(0, \delta_0, \alpha_w, \Theta_w)$, then a $(1-\alpha)\%$ prediction interval of $Y(\boldsymbol{t}_0)$ can be found as follows

$$P\left(q_{1-\frac{\alpha}{2}} \leq \widehat{\boldsymbol{\eta}}'\boldsymbol{Y} - Y(\boldsymbol{t}_0) \leq q_{\frac{\alpha}{2}}\right) = 1-\alpha,$$

Solving the above inequality, with respect to $Y(\boldsymbol{t}_0)$, we get

$$P\left(\widehat{\boldsymbol{\eta}}'\boldsymbol{Y} - q_{\frac{\alpha}{2}} \leq Y(\boldsymbol{t}_0) \leq \widehat{\boldsymbol{\eta}}'\boldsymbol{Y} - q_{1-\frac{\alpha}{2}}\right) = 1-\alpha.$$

Hence a $100(1-\alpha)\%$ prediction interval of $Y(\boldsymbol{t}_0)$ is

$$\left(\widehat{\boldsymbol{\eta}}'\boldsymbol{Y} - q_{\frac{\alpha}{2}},\ \widehat{\boldsymbol{\eta}}'\boldsymbol{Y} - q_{1-\frac{\alpha}{2}}\right).$$

4.2 Convolution of Generalized Skew Gaussian Distribution with Multivariate Normal Distribution

In this section, we will show that the generalized skew normal distribution is closed under convolution with the multivariate normal distribution. The results of this section will be used in chapter five to generalize the Gaussian process for regression.

Theorem 4.2.1 Let $\mathbf{X}\sim GSN_n^2(\mathbf{0},\boldsymbol{\Sigma},\lambda_0,1,\boldsymbol{\lambda}_1,\mathbf{0},\mathbf{I}_n)$ and $\epsilon\sim N_n(\mathbf{0},\tau^2\mathbf{I}_n)$, where $\tau>0$. If $Y=X+\epsilon$, then $Y\sim GSN_n^2(\mathbf{0},\boldsymbol{\Sigma}+\tau^2\mathbf{I}_n,\lambda_0^*,1,\tilde{\boldsymbol{\lambda}}_1,\mathbf{0},\mathbf{I}_n)$, where

$$\lambda_0^*=\lambda_0/\sqrt{1+\boldsymbol{\lambda}_1'\boldsymbol{\Sigma}^{-\frac{1}{2}}\left(\boldsymbol{\Sigma}^{-1}+\frac{1}{\tau^2}\mathbf{I}_n\right)^{-1}\boldsymbol{\Sigma}^{-\frac{1}{2}}\boldsymbol{\lambda}_1}$$

and

$$\tilde{\boldsymbol{\lambda}}_1'=\boldsymbol{\lambda}_1'\boldsymbol{\Sigma}^{\frac{1}{2}}(\boldsymbol{\Sigma}+\tau^2\mathbf{I}_n)^{-\frac{1}{2}}\Big/\sqrt{1+\boldsymbol{\lambda}_1'\boldsymbol{\Sigma}^{-\frac{1}{2}}\left(\boldsymbol{\Sigma}^{-1}+\frac{1}{\tau^2}\mathbf{I}_n\right)^{-1}\boldsymbol{\Sigma}^{-\frac{1}{2}}\boldsymbol{\lambda}_1}.$$

Proof. Let $Z\sim GSN_n^2(\mathbf{0},\ \mathbf{I}_n,1,\lambda_0,\boldsymbol{\lambda}_1,\mathbf{0},\mathbf{I}_n)$, i.e., Z has the pdf

$$f(z)=\frac{1}{(2\pi)^{\frac{n}{2}}\Phi\left(\lambda_0/\sqrt{1+\boldsymbol{\lambda}_1'\boldsymbol{\lambda}_1}\right)}exp\left(-\frac{1}{2}\mathbf{z}'\mathbf{z}\right)\Phi(\lambda_0+\boldsymbol{\lambda}_1'\mathbf{z}).$$

If $X=\boldsymbol{\Sigma}^{\frac{1}{2}}Z$, then $Z=\boldsymbol{\Sigma}^{-\frac{1}{2}}X$ and the Jacobian is $|J|=|\boldsymbol{\Sigma}|^{-\frac{1}{2}}$. Then

$$f_X(x)=\frac{exp\left(-\frac{1}{2}\boldsymbol{x}'\boldsymbol{\Sigma}^{-1}\boldsymbol{x}\right)}{(2\pi)^{\frac{n}{2}}|\boldsymbol{\Sigma}|^{\frac{1}{2}}\Phi(\delta_0)}\Phi\left(\lambda_0+\boldsymbol{\lambda}_1'\boldsymbol{\Sigma}^{-\frac{1}{2}}\boldsymbol{x}\right).$$

i.e.

$$\mathbf{X}\sim GSN_n^2(\mathbf{0},\boldsymbol{\Sigma},\lambda_0,1,\boldsymbol{\lambda}_1,\mathbf{0},\mathbf{I}_n).$$

When the integration is taken over $\mathbb{R}^n$, we get

$$f_Y(y) = \int f(y|\epsilon)\, f(\epsilon) d\epsilon,$$

$$= \frac{1}{(2\pi)^{\frac{n}{2}}|\boldsymbol{\Sigma}|^{\frac{1}{2}}\Phi(\delta_0)} \int exp\left(-\frac{1}{2}(\boldsymbol{y}-\epsilon)'\boldsymbol{\Sigma}^{-1}(\boldsymbol{y}-\epsilon) - \frac{1}{2\tau^2}\epsilon'\epsilon\right) \times$$

$$\frac{1}{(2\pi)^{\frac{n}{2}}\tau^n}\Phi\left(\lambda_0 + \boldsymbol{\lambda}_1'\boldsymbol{\Sigma}^{-\frac{1}{2}}(\boldsymbol{y}-\epsilon)\right) d\epsilon,$$

$$= \frac{exp\left(-\frac{1}{2}\boldsymbol{y}'\boldsymbol{\Sigma}^{-1}\boldsymbol{y}\right)}{(2\pi)^n|\boldsymbol{\Sigma}|^{\frac{1}{2}}\Phi(\delta_0)\tau^n} \int exp\left(-\frac{1}{2}\left(\epsilon'\left(\boldsymbol{\Sigma}^{-1} + \frac{1}{\tau^2}\mathbf{I}_n\right)\epsilon - 2\epsilon'\boldsymbol{\Sigma}^{-1}\boldsymbol{y}\right)\right) \times$$

$$\Phi\left(\lambda_0 - \boldsymbol{\lambda}_1'\boldsymbol{\Sigma}^{-\frac{1}{2}}(\epsilon - \boldsymbol{y})\right) d\epsilon,$$

Let $\boldsymbol{w} = \mathbf{A}\epsilon$, where $\mathbf{A}$ be such that $\mathbf{A}^2 = \boldsymbol{\Sigma}^{-1} + \frac{1}{\tau^2}\mathbf{I}_n$. The Jacobian of this transformation is

$$|J| = |\mathbf{A}^{-1}| = \left|\boldsymbol{\Sigma}^{-1} + \frac{1}{\tau^2}\mathbf{I}_n\right|^{-\frac{1}{2}}.$$

Therefore

$$f_Y(y) = \frac{exp\left(-\frac{1}{2}\boldsymbol{y}'\boldsymbol{\Sigma}^{-1}\boldsymbol{y}\right)}{(2\pi)^n|\boldsymbol{\Sigma}|^{\frac{1}{2}}\Phi(\delta_0)\tau^n} \int exp\left(-\frac{1}{2}(\boldsymbol{w}'\boldsymbol{w} - 2\boldsymbol{w}'\mathbf{A}^{-1}\boldsymbol{\Sigma}^{-1}\boldsymbol{y})\right) \times$$

$$\frac{\Phi\left(\lambda_0 - \boldsymbol{\lambda}_1'\boldsymbol{\Sigma}^{-\frac{1}{2}}\mathbf{A}^{-1}\boldsymbol{w} + \boldsymbol{\lambda}_1'\boldsymbol{\Sigma}^{-\frac{1}{2}}\boldsymbol{y}\right)}{\left|\boldsymbol{\Sigma}^{-1} + \frac{1}{\tau^2}\mathbf{I}_n\right|^{\frac{1}{2}}} d\boldsymbol{w}.$$

To simplify the calculation, let $\boldsymbol{q} = \mathbf{A}^{-1}\boldsymbol{\Sigma}^{-1}\boldsymbol{y}$. Then

$$f_Y(y) = \frac{exp\left(-\frac{1}{2}y'\boldsymbol{\Sigma}^{-1}y\right)}{(2\pi)^n\Phi(\delta_0)\tau^n|\boldsymbol{\Sigma}|^{\frac{1}{2}}\left|\boldsymbol{\Sigma}^{-1}+\frac{1}{\tau^2}\mathbf{I}_n\right|^{\frac{1}{2}}}\int exp\left(-\frac{1}{2}(w'w-2w'q+q'q-q'q)\right)\times$$

$$\Phi\left(\lambda_0+\boldsymbol{\lambda}_1'\boldsymbol{\Sigma}^{-\frac{1}{2}}\boldsymbol{y}-\boldsymbol{\lambda}_1'\boldsymbol{\Sigma}^{-\frac{1}{2}}\mathbf{A}^{-1}\boldsymbol{w}\right)dw,$$

$$= \frac{exp\left(-\frac{1}{2}y'\boldsymbol{\Sigma}^{-1}y+\frac{1}{2}q'q\right)}{(2\pi)^n\Phi(\delta_0)\tau^n|\boldsymbol{\Sigma}|^{\frac{1}{2}}\left|\boldsymbol{\Sigma}^{-1}+\frac{1}{\tau^2}\mathbf{I}_n\right|^{\frac{1}{2}}}\int exp\left(-\frac{1}{2}((\boldsymbol{w}-\boldsymbol{q})'(\boldsymbol{w}-\boldsymbol{q}))\right)\times$$

$$\Phi\left(\lambda_0+\boldsymbol{\lambda}_1'\boldsymbol{\Sigma}^{-\frac{1}{2}}\boldsymbol{y}-\boldsymbol{\lambda}_1'\boldsymbol{\Sigma}^{-\frac{1}{2}}\mathbf{A}^{-1}\boldsymbol{q}-\boldsymbol{\lambda}_1'\boldsymbol{\Sigma}^{-\frac{1}{2}}\mathbf{A}^{-1}(w-\boldsymbol{q})\right)dw,$$

$$= \frac{exp\left(-\frac{1}{2}y'\left(\boldsymbol{\Sigma}^{-1}-\boldsymbol{\Sigma}^{-1}\left(\boldsymbol{\Sigma}^{-1}+\frac{1}{\tau^2}\mathbf{I}_n\right)^{-1}\boldsymbol{\Sigma}^{-1}\right)y\right)}{(2\pi)^n\Phi(\delta_0)|\boldsymbol{\Sigma}+\tau^2\mathbf{I}_n|^{\frac{1}{2}}}\times$$

$$\int exp\left(-\frac{1}{2}((\boldsymbol{w}-\boldsymbol{q})'(\boldsymbol{w}-\boldsymbol{q}))\right)\times$$

$$\Phi\left(\lambda_0+\boldsymbol{\lambda}_1'\boldsymbol{\Sigma}^{-\frac{1}{2}}\left(\mathbf{I}_n-\left(\boldsymbol{\Sigma}^{-1}+\frac{1}{\tau^2}\mathbf{I}_n\right)^{-1}\boldsymbol{\Sigma}^{-1}\right)\boldsymbol{y}-\boldsymbol{\lambda}_1'\boldsymbol{\Sigma}^{-\frac{1}{2}}\left(\boldsymbol{\Sigma}^{-1}+\frac{1}{\tau^2}\mathbf{I}_n\right)^{-\frac{1}{2}}(w-\boldsymbol{q})\right)dw,$$

$$= \frac{exp\left(-\frac{1}{2}\boldsymbol{y}'(\boldsymbol{\Sigma}+\tau^2\mathbf{I}_n)^{-1}\boldsymbol{y}\right)}{(2\pi)^n\Phi(\delta_0)|\boldsymbol{\Sigma}+\tau^2\mathbf{I}_n|^{\frac{1}{2}}}\int exp\left(-\frac{1}{2}((w-\boldsymbol{q})'\mathbf{I}_n^{-1}(w-\boldsymbol{q}))\right)\times$$

$$\Phi\left(\lambda_0+\boldsymbol{\lambda}_1'\boldsymbol{\Sigma}^{\frac{1}{2}}(\boldsymbol{\Sigma}+\tau^2\mathbf{I}_n)^{-1}\boldsymbol{y}-\boldsymbol{\lambda}_1'\boldsymbol{\Sigma}^{-\frac{1}{2}}\left(\boldsymbol{\Sigma}^{-1}+\frac{1}{\tau^2}\mathbf{I}_n\right)^{-\frac{1}{2}}\mathbf{I}_n^{-\frac{1}{2}}(w-\boldsymbol{q})\right)dw,$$

Let $\boldsymbol{\lambda}_1^{*\prime} = -\boldsymbol{\lambda}_1' \boldsymbol{\Sigma}^{-\frac{1}{2}} \left(\boldsymbol{\Sigma}^{-1} + \frac{1}{\tau^2} \mathbf{I}_n\right)^{-\frac{1}{2}}$. Then

$$f_Y(y) = \frac{exp\left(-\frac{1}{2}\boldsymbol{y}'(\boldsymbol{\Sigma} + \tau^2 \mathbf{I}_n)^{-1}\boldsymbol{y}\right)}{(2\pi)^n \Phi(\delta_0)|\boldsymbol{\Sigma} + \tau^2 \mathbf{I}_n|^{\frac{1}{2}}} \int exp\left(-\frac{1}{2}\left((\boldsymbol{w} - \boldsymbol{q})' \mathbf{I}_n^{-1} (\boldsymbol{w} - \boldsymbol{q})\right)\right) \times$$

$$\Phi\left(\lambda_0 + \boldsymbol{\lambda}_1' \boldsymbol{\Sigma}^{\frac{1}{2}} (\boldsymbol{\Sigma} + \tau^2 \mathbf{I}_n)^{-1} \boldsymbol{y} + \boldsymbol{\lambda}_1^{*\prime} \mathbf{I}_n^{-1} (\boldsymbol{w} - \boldsymbol{q})\right) d\boldsymbol{w},$$

$$= \frac{exp\left(-\frac{1}{2}\boldsymbol{y}'(\boldsymbol{\Sigma} + \tau^2 \mathbf{I}_n)^{-1}\boldsymbol{y}\right)}{(2\pi)^n \Phi(\delta_0)|\boldsymbol{\Sigma} + \tau^2 \mathbf{I}_n|^{\frac{1}{2}}} \Phi\left(\frac{\lambda_0 + \boldsymbol{\lambda}_1' \boldsymbol{\Sigma}^{\frac{1}{2}} (\boldsymbol{\Sigma} + \tau^2 \mathbf{I}_n)^{-1} \boldsymbol{y}}{\sqrt{1 + \boldsymbol{\lambda}_1' \boldsymbol{\Sigma}^{-\frac{1}{2}} \left(\boldsymbol{\Sigma}^{-1} + \frac{1}{\tau^2} \mathbf{I}_n\right)^{-1} \boldsymbol{\Sigma}^{-\frac{1}{2}} \boldsymbol{\lambda}_1}}\right) (2\pi)^{\frac{n}{2}},$$

$$= \frac{exp\left(-\frac{1}{2}\boldsymbol{y}'(\boldsymbol{\Sigma} + \tau^2 \mathbf{I}_n)^{-1}\boldsymbol{y}\right)}{(2\pi)^{\frac{n}{2}} \Phi(\delta_0)|\boldsymbol{\Sigma} + \tau^2 \mathbf{I}_n|^{\frac{1}{2}}} \Phi\left(\frac{\lambda_0 + \boldsymbol{\lambda}_1' \boldsymbol{\Sigma}^{\frac{1}{2}} (\boldsymbol{\Sigma} + \tau^2 \mathbf{I}_n)^{-\frac{1}{2}} (\boldsymbol{\Sigma} + \tau^2 \mathbf{I}_n)^{-\frac{1}{2}} \boldsymbol{y}}{\sqrt{1 + \boldsymbol{\lambda}_1' \boldsymbol{\Sigma}^{-\frac{1}{2}} \left(\boldsymbol{\Sigma}^{-1} + \frac{1}{\tau^2} \mathbf{I}_n\right)^{-1} \boldsymbol{\Sigma}^{-\frac{1}{2}} \boldsymbol{\lambda}_1}}\right).$$

Finally, we get

$$f_Y(y) = \frac{exp\left(-\frac{1}{2}\boldsymbol{y}'(\boldsymbol{\Sigma} + \tau^2 \mathbf{I}_n)^{-1}\boldsymbol{y}\right)}{(2\pi)^{\frac{n}{2}} \Phi(\delta_0^*)|\boldsymbol{\Sigma} + \tau^2 \mathbf{I}_n|^{\frac{1}{2}}} \Phi\left(\lambda_0^* + \tilde{\boldsymbol{\lambda}}_1' (\boldsymbol{\Sigma} + \tau^2 \mathbf{I}_n)^{-\frac{1}{2}} \boldsymbol{y}\right),$$

where

$$\lambda_0^* = \lambda_0 / \sqrt{1 + \boldsymbol{\lambda}_1' \boldsymbol{\Sigma}^{-\frac{1}{2}} \left(\boldsymbol{\Sigma}^{-1} + \frac{1}{\tau^2} \mathbf{I}_n\right)^{-1} \boldsymbol{\Sigma}^{-\frac{1}{2}} \boldsymbol{\lambda}_1},$$

$$\tilde{\boldsymbol{\lambda}}_1' = \boldsymbol{\lambda}_1' \boldsymbol{\Sigma}^{\frac{1}{2}} (\boldsymbol{\Sigma} + \tau^2 \mathbf{I}_n)^{-\frac{1}{2}} / \sqrt{1 + \boldsymbol{\lambda}_1' \boldsymbol{\Sigma}^{-\frac{1}{2}} \left(\boldsymbol{\Sigma}^{-1} + \frac{1}{\tau^2} \mathbf{I}_n\right)^{-1} \boldsymbol{\Sigma}^{-\frac{1}{2}} \boldsymbol{\lambda}_1}$$

and

$$\delta_0^* = \frac{\lambda_0}{\sqrt{1+\boldsymbol{\lambda}_1'\boldsymbol{\lambda}_1}} = \frac{\lambda_0^*}{\sqrt{1+\tilde{\boldsymbol{\lambda}}_1'\tilde{\boldsymbol{\lambda}}_1}}.$$

So we have

$$\boldsymbol{Y} \sim GSN_n^2\left(\mathbf{0}, \boldsymbol{\Sigma}+\tau^2\mathbf{I}_n, \lambda_0^*, 1, \tilde{\boldsymbol{\lambda}}_1, \mathbf{0}, \mathbf{I}_n\right).$$

To complete the proof, we show that $\delta_0^* = \delta_0$. Since

$$\delta_0^* = \frac{\lambda_0^*}{\sqrt{1+\tilde{\boldsymbol{\lambda}}_1'\tilde{\boldsymbol{\lambda}}_1}},$$

then

$$\delta_0^* = \frac{\lambda_0 / \sqrt{1+\boldsymbol{\lambda}_1'\boldsymbol{\Sigma}^{-\frac{1}{2}}\left(\boldsymbol{\Sigma}^{-1}+\frac{1}{\tau^2}\mathbf{I}_n\right)^{-1}\boldsymbol{\Sigma}^{-\frac{1}{2}}\boldsymbol{\lambda}_1}}{\sqrt{1+\boldsymbol{\lambda}_1'\boldsymbol{\Sigma}^{\frac{1}{2}}(\boldsymbol{\Sigma}+\tau^2\mathbf{I}_n)^{-1}\boldsymbol{\Sigma}^{\frac{1}{2}}\boldsymbol{\lambda}_1 / 1+\boldsymbol{\lambda}_1'\boldsymbol{\Sigma}^{-\frac{1}{2}}\left(\boldsymbol{\Sigma}^{-1}+\frac{1}{\tau^2}\mathbf{I}_n\right)^{-1}\boldsymbol{\Sigma}^{-\frac{1}{2}}\boldsymbol{\lambda}_1}},$$

$$= \frac{\lambda_0}{\sqrt{1+\boldsymbol{\lambda}_1'\boldsymbol{\Sigma}^{-\frac{1}{2}}\left(\boldsymbol{\Sigma}^{-1}+\frac{1}{\tau^2}\mathbf{I}_n\right)^{-1}\boldsymbol{\Sigma}^{-\frac{1}{2}}\boldsymbol{\lambda}_1+\boldsymbol{\lambda}_1'\boldsymbol{\Sigma}^{\frac{1}{2}}(\boldsymbol{\Sigma}+\tau^2\mathbf{I}_n)^{-1}\boldsymbol{\Sigma}^{\frac{1}{2}}\boldsymbol{\lambda}_1}}.$$

To complete, we use the matrix identity

$$(\boldsymbol{\Sigma}+\tau^2\mathbf{I}_n)^{-1} = \boldsymbol{\Sigma}^{-1}-\boldsymbol{\Sigma}^{-1}\left(\boldsymbol{\Sigma}^{-1}+\frac{1}{\tau^2}\mathbf{I}_n\right)^{-1}\boldsymbol{\Sigma}^{-1},$$

we simplify δ_0^* to

$$\boldsymbol{\lambda}_1'\boldsymbol{\Sigma}^{\frac{1}{2}}(\boldsymbol{\Sigma}+\tau^2\mathbf{I}_n)^{-1}\boldsymbol{\Sigma}^{\frac{1}{2}}\boldsymbol{\lambda}_1 = \boldsymbol{\lambda}_1'\boldsymbol{\Sigma}^{\frac{1}{2}}\left[\boldsymbol{\Sigma}^{-1}-\boldsymbol{\Sigma}^{-1}\left(\boldsymbol{\Sigma}^{-1}+\frac{1}{\tau^2}\mathbf{I}_n\right)^{-1}\boldsymbol{\Sigma}^{-1}\right]\boldsymbol{\Sigma}^{\frac{1}{2}}\boldsymbol{\lambda}_1,$$

$$= \boldsymbol{\lambda}_1'\boldsymbol{\lambda}_1 - \boldsymbol{\lambda}_1'\boldsymbol{\Sigma}^{-\frac{1}{2}}\left(\boldsymbol{\Sigma}^{-1} + \frac{1}{\tau^2}\mathbf{I}_n\right)^{-1}\boldsymbol{\Sigma}^{-\frac{1}{2}}\boldsymbol{\lambda}_1.$$

Thus

$$\boldsymbol{\lambda}_1'\boldsymbol{\Sigma}^{-\frac{1}{2}}\left(\boldsymbol{\Sigma}^{-1} + \frac{1}{\tau^2}\mathbf{I}_n\right)^{-1}\boldsymbol{\Sigma}^{-\frac{1}{2}}\boldsymbol{\lambda}_1 + \boldsymbol{\lambda}_1'\boldsymbol{\lambda}_1 - \boldsymbol{\lambda}_1'\boldsymbol{\Sigma}^{-\frac{1}{2}}\left(\boldsymbol{\Sigma}^{-1} + \frac{1}{\tau^2}\mathbf{I}_n\right)^{-1}\boldsymbol{\Sigma}^{-\frac{1}{2}}\boldsymbol{\lambda}_1 = \boldsymbol{\lambda}_1'\boldsymbol{\lambda}_1.$$

Hence

$$\delta_0^* = \frac{\lambda_0}{\sqrt{1 + +\boldsymbol{\lambda}_1'\boldsymbol{\lambda}_1}} = \delta_0.$$

Chapter Five

Generalization of Skew Gaussian Process for Regression

In this chapter, we give two generalizations to the Gaussian process for regression. The first one comes from assuming a GSGP on the output function $f(x)$, while a Gaussian process is assumed on the noise. The second comes from assuming a GP on the output $f(x)$ and a GSGP on the noise.

5.1 Generalized Skew Gaussian Process for Regression (GSGPR)

In this section, we generalize the GPR to the generalized skew Gaussian process regression (GSGPR) by assuming a generalized skew Gaussian process prior on the output function $f(x)$. In this generalization, the input function $f(x)$ will vary over the class of all sample path functions of a GSGP which contains the class of all sample path functions of a GP.

To generalize the GPR into GSGPR, we start by considering the following regression model

$$y_j = f(x_j) + \epsilon(x_j), \quad j = 1,2,\dots,n \qquad (5.1)$$

where $f(x)$ is assumed to have a GSGP, i.e., for every set of inputs $x_1,\dots,x_n$, the vector $\boldsymbol{f} = (f(x_1),\dots,f(x_n))'$ has the distribution $GSN_n^2\left(\mathbf{0},\boldsymbol{\Omega}^{\frac{1}{2}}\boldsymbol{\Sigma}\boldsymbol{\Omega}^{\frac{1}{2}},\lambda_0,\sigma^2,\lambda\mathbf{1}_n,\gamma\mathbf{1}_n,\boldsymbol{\Sigma}\right)$, where $\boldsymbol{\Omega} = \left(K_1(x_i,x_j)\right)_{i,j=1}^{n}$ and $\boldsymbol{\Sigma} = \left(K_2(x_i,x_j)\right)_{i,j=1}^{n}$, K_1 and K_2 are covariance functions and

$\epsilon = (\epsilon(x_1), \dots, \epsilon(x_n))' \sim N_n(0, \tau^2 \mathbf{I}_n)$, independent of f. As a special case, we may assume that $\mathbf{\Omega} = \mathbf{I}_n$, i.e., $K_1(x_i, x_j) = \delta_{ij}$, where δ_{ij} the Kroneker function defined as

$$\delta_{ij} = \begin{cases} 1, & i = j \\ 0, & i \neq j \end{cases}$$

It can be noted that the Gaussian process regression model is a special case of the model (5.1) when $\lambda_0 = 0, \boldsymbol{\lambda} = \mathbf{0}, \boldsymbol{\gamma} = \mathbf{0}, \sigma^2 = 1, \mathbf{\Omega} = \mathbf{I}_n$ and $\mathbf{\Sigma} = \mathbf{I}_n$.

5.2 Prediction at fixed input

In this section, we used the GSGP for predicting the output function at a fixed input and deriving the predicting distribution, along with it is mean and variance.

Let $\boldsymbol{x}^*$ be a fixed input vector in the domain of $\boldsymbol{x}$. We are interested in predicting the value of $f(\boldsymbol{x})$ at $\boldsymbol{x}^*$. Let $f^* = f(\boldsymbol{x}^*)$ and $S = \{(\boldsymbol{x}_1, y_1), \dots, (\boldsymbol{x}_n, y_n)\}$ be the observed data according to the model (5.1). Since $f(\boldsymbol{x})$ follows a GSGP, then

$$\begin{pmatrix} \boldsymbol{f} \\ f^* \end{pmatrix} = \begin{pmatrix} f(\boldsymbol{x}_1) \\ \vdots \\ f(\boldsymbol{x}_n) \\ f(\boldsymbol{x}^*) \end{pmatrix} \sim GSN_{n+1}^2(\mathbf{0}, \Sigma, \lambda_0, \sigma^2, \lambda \mathbf{1}_{n+1}, \gamma \mathbf{1}_{n+1}, \Sigma)$$

and

$$\boldsymbol{y} = \begin{pmatrix} y_1 \\ \vdots \\ y_n \end{pmatrix} | \boldsymbol{f} \sim N_n(\boldsymbol{f}, \tau^2 \mathbf{I}_n),$$

where

$$\boldsymbol{\Sigma} = \begin{pmatrix} \Sigma_{11} & \Sigma_{12} \\ \Sigma_{21} & \Sigma_{22} \end{pmatrix}, \Sigma_{11} = \left(K_2(x_i, x_j)\right)_{i,j=1}^{n}, \boldsymbol{\Sigma}_{12} = \left(K_2(x_1, \boldsymbol{x}^*), \dots, K_2(\boldsymbol{x}_n, \boldsymbol{x}^*)\right)', \boldsymbol{\Sigma}_{21} = \boldsymbol{\Sigma}_{12}'$$

and $\boldsymbol{\Sigma}_{22} = K_2(\boldsymbol{x}^*, \boldsymbol{x}^*)$.

To obtain the predictive distribution of f^* given $\boldsymbol{x}^*$ and $\boldsymbol{y}$, we need the following theorem.

Theorem. 5.2.1 Consider the model $y_i = f(\boldsymbol{x}_i) + \epsilon(\boldsymbol{x}_i)$, $\boldsymbol{x}_i \in D$, $i = 1,2,\dots,n$, where $f(\boldsymbol{x})$ and $\epsilon(\boldsymbol{x})$ are the same as in (5.1). Moreover, assume that $f(\boldsymbol{x})$ satisfies the weak stationarity of the third order conditions with $\boldsymbol{\gamma} = \mathbf{0}$, $\boldsymbol{\mu} = \mathbf{0}$ and $\sigma^2 = 1$. If $\boldsymbol{y}' = (y_1, \dots, y_n)$, $\boldsymbol{x}_i \in D$ and

$$\mathbf{V} = \begin{pmatrix} \mathbf{V}_{11} & \mathbf{V}_{12} \\ \mathbf{V}_{12}' & \mathbf{V}_{22} \end{pmatrix} = \begin{pmatrix} \left(K_1(x_i, x_j)\right)_{i,j=1}^{n} & \left(K_1(x^*, x_j)\right)_{j=1}^{n} \\ \left(K_1(x^*, x_j)\right)_{j=1}^{n} & K_1(x^*, x^*) \end{pmatrix},$$

then

$$(\boldsymbol{y}', f^*)' \sim GSN_{n+1}^{2}\left(\mathbf{0}, \mathbf{V} + \tau^2 \tilde{\mathbf{I}}_{n+1}, \lambda_0^+, 1, \tilde{\boldsymbol{\lambda}}^+, \mathbf{0}, \mathbf{I}_{n+1}\right)$$

where

$$\lambda_0^+ = \frac{\lambda_0}{\sqrt{1+Q}}, \qquad \tilde{\lambda}^+ = \frac{\tilde{\lambda}}{\sqrt{1+Q}}, \qquad \check{\mathbf{I}}_{n+1} = \begin{pmatrix} \mathbf{I}_n & \mathbf{0} \\ \mathbf{0}' & 0 \end{pmatrix},$$

$$Q = \lambda^2 \mathbf{1}_{n+1}' \mathbf{V}^{-\frac{1}{2}} \begin{pmatrix} \mathbf{I}_n - \left(\mathbf{I}_n + \tau^2 \widetilde{\mathbf{V}}^{-1}\right)^{-1} & \mathbf{0} \\ -\tau^2 \mathbf{w}_{\mathbf{v}} \widetilde{\mathbf{V}}^{-1}\left(\mathbf{I}_n + \tau^2 \widetilde{\mathbf{V}}^{-1}\right)^{-1} & \mathbf{0} \end{pmatrix} \mathbf{V}^{-\frac{1}{2}} \mathbf{1}_{n+1}, \text{ and}$$

$$\widetilde{\mathbf{V}} = \mathbf{V}_{11} - \mathbf{V}_{12} \mathbf{V}_{22}^{-1} \mathbf{V}_{21}$$

Proof: We will use the mgf. Since $f(x)$ is a GSGPR, then

$$(\boldsymbol{f}', f^*)' \sim GSN^2_{n+1}(\boldsymbol{\mu}, \mathbf{V}, \lambda_0, 1, \lambda \mathbf{1}'_{n+1}, \mathbf{0}, \mathbf{I}_{n+1}),$$

and $\boldsymbol{y} = \boldsymbol{f} + \boldsymbol{\epsilon},$ where $\boldsymbol{\epsilon} = \left(\epsilon(x_1), \ldots, \epsilon(x_n)\right)'$. This implies

$$M_{\boldsymbol{y}, f^*}(\boldsymbol{t}, s) = E(\exp(\boldsymbol{t}'\boldsymbol{y} + s f^*))$$

$$= E(\exp\,(\boldsymbol{t}'(\boldsymbol{f} + \boldsymbol{\epsilon}) + s f^*)),$$

$$= E(\exp\,(\boldsymbol{t}'\boldsymbol{f} + s f^* + \boldsymbol{t}'\boldsymbol{\epsilon})),$$

$$= M_{\boldsymbol{f}, f^*}(\boldsymbol{t}, s) M_{\boldsymbol{\epsilon}}(\boldsymbol{t}),$$

since

$$M_{\boldsymbol{\epsilon}}(\boldsymbol{t}) = exp\left(\frac{1}{2}\tau^2 \boldsymbol{t}'\boldsymbol{t}\right),$$

and

$$M_{\boldsymbol{f}, f^*}(\boldsymbol{t}, s) = \frac{exp\left(\frac{1}{2}(\boldsymbol{t}' \quad s)\mathbf{V}\begin{pmatrix}\boldsymbol{t}\\ s\end{pmatrix}\right)}{\Phi(\delta_0)} \Phi\left(\frac{\lambda_0 + \lambda \mathbf{1}'_{n+1}\mathbf{V}^{\frac{1}{2}}\begin{pmatrix}\boldsymbol{t}\\ s\end{pmatrix}}{\sqrt{1 + \lambda^2(n+1)}}\right),$$

then

$$M_{\boldsymbol{y}, f^*}(\boldsymbol{t}, s) = \frac{1}{\Phi(\delta_0)} \Phi\left(\frac{\lambda_0 + \lambda \mathbf{1}'_{n+1}\mathbf{V}^{\frac{1}{2}}\begin{pmatrix}\boldsymbol{t}\\ s\end{pmatrix}}{\sqrt{1 + \lambda^2(n+1)}}\right) exp\left(\frac{1}{2}(\boldsymbol{t}' \quad s)\mathbf{V}\begin{pmatrix}\boldsymbol{t}\\ s\end{pmatrix} + \frac{1}{2}\tau^2 \boldsymbol{t}'\boldsymbol{t}\right)$$

$$= \frac{1}{\Phi(\delta_0)} \Phi\left(\frac{\lambda_0 + \lambda \mathbf{1}'_{n+1} \mathbf{V}^{\frac{1}{2}} \binom{t}{s}}{\sqrt{1+\lambda^2(n+1)}} \right) exp\left(\frac{1}{2} (\boldsymbol{t}' \quad s) \left(\mathbf{V} + \tau^2 \breve{\mathbf{I}}_{n+1} \right) \binom{t}{s} \right).$$

Note that

$$\lambda_0 + \lambda \mathbf{1}'_{n+1} \mathbf{V}^{\frac{1}{2}} \binom{t}{s} = \lambda_0 + \lambda \mathbf{1}'_{n+1} \mathbf{V}^{\frac{1}{2}} \left(\mathbf{V} + \tau^2 \breve{\mathbf{I}}_{n+1} \right)^{-\frac{1}{2}} \left(\mathbf{V} + \tau^2 \breve{\mathbf{I}}_{n+1} \right)^{\frac{1}{2}} \binom{t}{s},$$

$$= \lambda_0 + \tilde{\boldsymbol{\lambda}}' \left(\mathbf{V} + \tau^2 \breve{\mathbf{I}}_{n+1} \right)^{\frac{1}{2}} \binom{t}{s},$$

where

$$\tilde{\boldsymbol{\lambda}}' = \lambda \mathbf{1}'_{n+1} \mathbf{V}^{\frac{1}{2}} \left(\mathbf{V} + \tau^2 \breve{\mathbf{I}}_{n+1} \right)^{-\frac{1}{2}}.$$

Hence

$$M_{y,f^*}(\boldsymbol{t}, s) = \frac{1}{\Phi(\delta_o)} \Phi\left(\frac{\lambda_0}{\sqrt{1+\lambda^2(n+1)}} + \frac{\tilde{\boldsymbol{\lambda}}'}{\sqrt{1+\lambda^2(n+1)}} \left(\mathbf{V} + \tau^2 \breve{\mathbf{I}}_{n+1} \right)^{\frac{1}{2}} \binom{t}{s} \right) \times$$

$$exp\left(\frac{1}{2} (\boldsymbol{t}' \quad s) \left(\mathbf{V} + \tau^2 \breve{\mathbf{I}}_{n+1} \right) \binom{t}{s} \right),$$

and

$$\tilde{\boldsymbol{\lambda}}' \tilde{\boldsymbol{\lambda}} = \lambda \mathbf{1}'_{n+1} \mathbf{V}^{\frac{1}{2}} \left(\mathbf{V} + \tau^2 \breve{\mathbf{I}}_{n+1} \right)^{-\frac{1}{2}} \lambda \left(\mathbf{V} + \tau^2 \breve{\mathbf{I}}_{n+1} \right)^{-\frac{1}{2}} \mathbf{V}^{\frac{1}{2}} \mathbf{1}_{n+1},$$

$$= \lambda^2 \mathbf{1}'_{n+1} \mathbf{V}^{\frac{1}{2}} \left(\mathbf{V} + \tau^2 \breve{\mathbf{I}}_{n+1} \right)^{-1} \mathbf{V}^{\frac{1}{2}} \mathbf{1}_{n+1},$$

$$= \lambda^2 \mathbf{1}'_{n+1} \mathbf{V}^{\frac{1}{2}} \left(\mathbf{I}_{n+1} + \tau^2 \mathbf{V}^{-1} \breve{\mathbf{I}}_{n+1} \right)^{-1} \mathbf{V}^{-\frac{1}{2}} \mathbf{1}_{n+1}.$$

To complete the proof, let

$$\mathbf{I}_{n+1} + \tau^2 \mathbf{V}^{-1} \breve{\mathbf{I}}_{n+1} = \begin{pmatrix} \mathbf{I}_n + \tau^2 (\mathbf{V}_{11} - \mathbf{V}_{12}\mathbf{V}_{22}^{-1}\mathbf{V}_{21})^{-1} & \mathbf{0} \\ -\tau^2 \mathbf{V}_{22}^{-1}\mathbf{V}_{21}(\mathbf{V}_{11} - \mathbf{V}_{12}\mathbf{V}_{22}^{-1}\mathbf{V}_{21})^{-1} & 1 \end{pmatrix}.$$

To simplify the notation, let $\widetilde{\mathbf{V}} = \mathbf{V}_{11} - \mathbf{V}_{12}\mathbf{V}_{22}^{-1}\mathbf{V}_{21}$ and $\boldsymbol{w}_{\mathrm{v}} = \mathbf{V}_{22}^{-1}\mathbf{V}_{21}$, then

$$\mathbf{I}_{n+1} + \tau^2 \mathbf{V}^{-1} \breve{\mathbf{I}}_{n+1} = \begin{pmatrix} \mathbf{I}_n + \tau^2 \widetilde{\mathbf{V}}^{-1} & \mathbf{0} \\ -\tau^2 \boldsymbol{w}_{\mathrm{v}} \widetilde{\mathbf{V}}^{-1} & 1 \end{pmatrix}.$$

So

$$\left(\mathbf{I}_{n+1} + \tau^2 \mathbf{V}^{-1} \breve{\mathbf{I}}_{n+1}\right)^{-1} = \begin{pmatrix} \left(\mathbf{I}_n + \tau^2 \widetilde{\mathbf{V}}^{-1}\right)^{-1} & \mathbf{0} \\ \tau^2 \boldsymbol{w}_{\mathrm{v}} \widetilde{\mathbf{V}}^{-1} \left(\mathbf{I}_n + \tau^2 \widetilde{\mathbf{V}}^{-1}\right)^{-1} & 1 \end{pmatrix},$$

$$= \begin{pmatrix} \mathbf{I}_n - \left(\mathbf{I}_n + \tau^{-2} \widetilde{\mathbf{V}}\right)^{-1} & \mathbf{0} \\ \tau^2 \boldsymbol{w}_v \widetilde{\mathbf{V}}^{-1} \left(\mathbf{I}_n + \tau^2 \widetilde{\mathbf{V}}^{-1}\right)^{-1} & 1 \end{pmatrix},$$

$$= \mathbf{I}_{n+1} - \begin{pmatrix} \left(\mathbf{I}_n + \tau^{-2} \widetilde{\mathbf{V}}\right)^{-1} & \mathbf{0} \\ -\tau^2 \boldsymbol{w}_{\mathrm{v}} \widetilde{\mathbf{V}}^{-1} \left(\mathbf{I}_n + \tau^2 \widetilde{\mathbf{V}}^{-1}\right)^{-1} & 0 \end{pmatrix},$$

$$= \mathbf{I}_{n+1} - \begin{pmatrix} \mathbf{I}_n - \left(\mathbf{I}_n + \tau^2 \widetilde{\mathbf{V}}^{-1}\right)^{-1} & \mathbf{0} \\ -\tau^2 \boldsymbol{w}_{\mathrm{v}} \widetilde{\mathbf{V}}^{-1} \left(\mathbf{I}_n + \tau^2 \widetilde{\mathbf{V}}^{-1}\right)^{-1} & 0 \end{pmatrix},$$

So

$$\tilde{\lambda}'\tilde{\lambda} = \lambda^2(n+1) - \lambda^2 \mathbf{1}_{n+1}' \mathbf{V}^{\frac{1}{2}} \begin{pmatrix} \mathbf{I}_n - \left(\mathbf{I}_n + \tau^2 \widetilde{\mathbf{V}}^{-1}\right)^{-1} & \mathbf{0} \\ -\tau^2 \boldsymbol{w}_{\mathrm{v}} \widetilde{\mathbf{V}}^{-1} \left(\mathbf{I}_n + \tau^2 \widetilde{\mathbf{V}}^{-1}\right)^{-1} & 0 \end{pmatrix} \mathbf{V}^{-\frac{1}{2}} \mathbf{1}_{n+1}.$$

Therefore,

$$1 + \lambda^2(n+1) = 1 + \tilde{\lambda}'\tilde{\lambda} + Q,$$

where

$$Q = \lambda^2 \mathbf{1}'_{n+1}\mathbf{V}^{\frac{1}{2}}\begin{pmatrix} \mathbf{I}_n - \left(\mathbf{I}_n + \tau^2\widetilde{\mathbf{V}}^{-1}\right)^{-1} & \mathbf{0} \\ -\tau^2 w_{\mathbf{v}}\widetilde{\mathbf{V}}^{-1}\left(\mathbf{I}_n + \tau^2\widetilde{\mathbf{V}}^{-1}\right)^{-1} & 0 \end{pmatrix}\mathbf{V}^{-\frac{1}{2}}\mathbf{1}_{n+1}.$$

Now

$$M_{y,f^*}(\boldsymbol{t},s) = \frac{1}{\Phi(\delta_0)}\Phi\left(\frac{\lambda_0/\sqrt{1+Q} + \tilde{\lambda}'/\sqrt{1+Q}}{\sqrt{1+\tilde{\lambda}'/\sqrt{1+Q}\tilde{\lambda}/\sqrt{1+Q}}}(\mathbf{V} + \tau^2\tilde{\mathbf{I}}_{n+1})^{\frac{1}{2}}\binom{\boldsymbol{t}}{s}\right) \times$$

$$exp\left(\frac{1}{2}(\boldsymbol{t}' \quad s)(\mathbf{V} + \tau^2\tilde{\mathbf{I}}_{n+1})\binom{\boldsymbol{t}}{s}\right).$$

If we set $\tilde{\boldsymbol{\lambda}}^+ = \tilde{\boldsymbol{\lambda}}/\sqrt{1+Q}$, and $\lambda_0^+ = \lambda_0 / \sqrt{1+Q}$, then

$$M_{y,f^*}(\boldsymbol{t},s) = \frac{1}{\Phi(\delta_0)}\Phi\left(\frac{\lambda_0^+ + \tilde{\boldsymbol{\lambda}}^{+'}}{\sqrt{1+\tilde{\boldsymbol{\lambda}}^{+'}\tilde{\boldsymbol{\lambda}}^+}}(\mathbf{V} + \tau^2\tilde{\mathbf{I}}_{n+1})^{\frac{1}{2}}\binom{\boldsymbol{t}}{s}\right) \times$$

$$exp\left(\frac{1}{2}(\boldsymbol{t}' \quad s)(\mathbf{V} + \tau^2\tilde{\mathbf{I}}_{n+1})\binom{\boldsymbol{t}}{s}\right),$$

so, we get

$$g_{Y,f^*}(y,f^*) = \frac{1}{(2\pi)^{\frac{n+1}{2}}\Phi(\delta_0)\left|\mathbf{V} + \tau^2\tilde{\mathbf{I}}_{n+1}\right|^{\frac{1}{2}}}\Phi\left(\lambda_0^+ + \tilde{\boldsymbol{\lambda}}^{+'}(\mathbf{V} + \tau^2\tilde{\mathbf{I}}_{n+1})^{-\frac{1}{2}}\binom{\boldsymbol{y}}{f^*}\right) \times$$

$$exp\left(-\frac{1}{2}(\boldsymbol{y}' \quad f^*)(\mathbf{V}+\tau^2\tilde{\mathbf{I}}_{n+1})^{-1}\binom{\boldsymbol{y}}{f^*}\right),$$

that is

$$(\boldsymbol{y}', f^*)' \sim GSN_{n+1}^2\left(\mathbf{0}, \mathbf{V}+\tau^2\tilde{\mathbf{I}}_{n+1}, \lambda_0^+, 1, \tilde{\lambda}^+, \mathbf{0}, \mathbf{I}_{n+1}\right).$$

Lemma. 5.2.1 Consider the arguments in previous theorem. Then

(i). The conditional distribution of f^* given $\boldsymbol{Y}=\boldsymbol{y}, \boldsymbol{X}^*=\boldsymbol{x}^*$ is

$$(f^*|\boldsymbol{y}, \boldsymbol{x}^*) \sim GSN_1^2(u_*, \tilde{\sigma}, \lambda_{0*}, 1, \lambda_*, 0,1),$$

where

$$u_* = \boldsymbol{u}'(\mathbf{V}_{11}+\tau^2\mathbf{I}_n)^{-1}\boldsymbol{y}, \tilde{\sigma} = \mathrm{R}-\boldsymbol{u}'(\mathbf{V}_{11}+\tau^2\mathbf{I}_n)^{-1}\boldsymbol{u}, \lambda_{0*} = \lambda_{00} + \lambda_*'\tilde{\sigma}^{-\frac{1}{2}}u_*,$$

$$\lambda_{00} = \lambda_0^+ + \left(\tilde{\lambda}_1' - \mathrm{c}\mathrm{R}^{-1}\boldsymbol{u}'\right)(\mathbf{V}_{11}+\tau^2\mathbf{I}_n - \boldsymbol{u}\mathrm{R}^{-1}\boldsymbol{u}')^{-1}\boldsymbol{y},$$

$$\lambda_*' = \left(\mathrm{c} - \tilde{\boldsymbol{\lambda}}_1'(\mathbf{V}_{11}+\tau^2\mathbf{I}_n)^{-1}\boldsymbol{u}\right)(\mathrm{R}-\boldsymbol{u}'(\mathbf{V}_{11}+\tau^2\mathbf{I}_n)^{-1}\boldsymbol{u})^{-\frac{1}{2}},$$

and $\lambda_0^+ = \lambda_0/\sqrt{1+Q}$.

(ii). The expectation and the variance of f^* given $\boldsymbol{Y}=\boldsymbol{y}, \boldsymbol{X}^*=\boldsymbol{x}^*$ respectively are

$$E(f^*|\boldsymbol{Y}=\boldsymbol{y}, \boldsymbol{X}^*=\boldsymbol{x}^*) = u_* + \frac{\varphi(\delta_0)}{\Phi(\delta_0)}\left(\tilde{\sigma}^{\frac{1}{2}}\lambda_*/\sqrt{1+\lambda_*^2}\right),$$

and

$$Var(f^*|\boldsymbol{Y}=\boldsymbol{y}, \boldsymbol{X}^*=\boldsymbol{x}^*) = \tilde{\sigma}(1-h),$$

where

$$h=\left(\left(\frac{\varphi(\delta_0)}{\Phi(\delta_0)\sqrt{1+\lambda_*^2}}\right)^2+\frac{\delta_0\varphi(\delta_0)}{\Phi(\delta_0)(1+\lambda_*^2)}\right)\lambda_*^2.$$

Proof of (i): To find the conditional distribution of f^* given $Y=y, X^*=x^*$, consider the following partitions

$$\mathbf{V}+\tau^2\tilde{\mathbf{I}}_{n+1}=\begin{pmatrix}\mathbf{V}_{11}+\tau^2\mathbf{I}_n & \boldsymbol{u}\\ \boldsymbol{u}' & \mathrm{R}\end{pmatrix}, \lambda^{+\prime}=(\tilde{\lambda}_1' \quad \mathrm{c})',$$

where

$$\lambda^{+\prime}=\tilde{\lambda}^{+\prime}(\mathbf{V}+\tau^2\tilde{\mathbf{I}}_{n+1})^{\frac{1}{2}}.$$

Also consider the following notation

$$\mathbf{V}_y=\mathbf{V}_{11}+\tau^2\mathbf{I}_n, \tilde{\sigma}=\mathrm{R}-\boldsymbol{u}'(\mathbf{V}_{11}+\tau^2\mathbf{I}_n)^{-1}\boldsymbol{u} \text{ and } u_*=\boldsymbol{u}'(\mathbf{V}_{11}+\tau^2\mathbf{I}_n)^{-1}\boldsymbol{y}.$$

Then

$$(\boldsymbol{y}' \quad f^*)(\mathbf{V}+\tau^2\tilde{\mathbf{I}}_{n+1})^{-1}\begin{pmatrix}\boldsymbol{y}\\ f^*\end{pmatrix}=\boldsymbol{y}'\mathbf{V}_y^{-1}\boldsymbol{y}+(f^*-u_*)'\tilde{\sigma}^{-1}(f^*-u_*),$$

$$=\boldsymbol{y}'\mathbf{V}_y^{-1}\boldsymbol{y}+(f^*-u_*)^2\tilde{\sigma}^{-1},$$

and

$$\tilde{\lambda}^{+\prime}(\mathbf{V}+\tau^2\tilde{\mathbf{I}}_{n+1})^{-\frac{1}{2}}\begin{pmatrix}\boldsymbol{y}\\ f^*\end{pmatrix}=\tilde{\lambda}^{+\prime}(\mathbf{V}+\tau^2\tilde{\mathbf{I}}_{n+1})^{\frac{1}{2}}(\mathbf{V}+\tau^2\tilde{\mathbf{I}}_{n+1})^{-1}\begin{pmatrix}\boldsymbol{y}\\ f^*\end{pmatrix},$$

$$=\lambda^{+\prime}(\mathbf{V}+\tau^2\tilde{\mathbf{I}}_{n+1})^{-1}\begin{pmatrix}\boldsymbol{y}\\ f^*\end{pmatrix},$$

$$= (\tilde{\boldsymbol{\lambda}}_1' \quad \mathrm{c})(\mathrm{V} + \tau^2 \tilde{\mathbf{I}}_{n+1})^{-1} \binom{\boldsymbol{y}}{f^*},$$

$$= \left(\tilde{\boldsymbol{\lambda}}_1' - \mathrm{c}\mathrm{R}^{-1}\boldsymbol{u}'\right)(\mathbf{V}_{11} + \tau^2 \mathbf{I}_n - \boldsymbol{u}\mathrm{R}^{-1}\boldsymbol{u}')^{-1}\boldsymbol{y} +$$

$$\left(\mathrm{c} - \tilde{\boldsymbol{\lambda}}_1'(\mathbf{V}_{11} + \tau^2 \mathbf{I}_n)^{-1}\boldsymbol{u}\right)(\mathrm{R} - \boldsymbol{u}'(\mathbf{V}_{11} + \tau^2 \mathbf{I}_n)^{-1}\boldsymbol{u})^{-1} f^*.$$

To simplify the calculation, let

$$\lambda_{00} = \lambda_0^+ + \left(\tilde{\boldsymbol{\lambda}}_1' - \mathrm{c}\mathrm{R}^{-1}\boldsymbol{u}'\right)(\mathbf{V}_{11} + \tau^2 \mathbf{I}_n - \boldsymbol{u}\mathrm{R}^{-1}\boldsymbol{u}')^{-1}\boldsymbol{y},$$

$$\lambda_*' = \left(\mathrm{c} - \tilde{\boldsymbol{\lambda}}_1'(\mathbf{V}_{11} + \tau^2 \mathbf{I}_n)^{-1}\boldsymbol{u}\right)(\mathrm{R} - \boldsymbol{u}'(\mathbf{V}_{11} + \tau^2 \mathbf{I}_n)^{-1}\boldsymbol{u})^{-\frac{1}{2}}$$

$$\tilde{\sigma} = \mathrm{R} - \boldsymbol{u}'(\mathbf{V}_{11} + \tau^2 \mathbf{I}_n)^{-1}\boldsymbol{u}, \text{ and } u_* = \boldsymbol{u}'(\mathbf{V}_{11} + \tau^2 \mathbf{I}_n)^{-1}\boldsymbol{y}.$$

More simplification leads to

$$g_{f^*|y,x^*} \propto exp\left(-\frac{1}{2}\tilde{\sigma}^{-1}(f^* - u_*)^2\right)\Phi\left(\lambda_{00} + \lambda_*'\tilde{\sigma}^{-\frac{1}{2}}f^*\right),$$

$$\propto exp\left(-\frac{1}{2}\tilde{\sigma}^{-1}(f^* - u_*)^2\right)\Phi\left(\lambda_{00} + \lambda_*'\tilde{\sigma}^{-\frac{1}{2}}u_* + \lambda_*'\tilde{\sigma}^{-\frac{1}{2}}(f^* - u_*)\right).$$

Let $\lambda_{0*} = \lambda_{00} + \lambda_*'\tilde{\sigma}^{-\frac{1}{2}}u_*$, then

$$g_{f^*|y,x^*} \propto exp\left(-\frac{1}{2}\tilde{\sigma}^{-1}(f^* - u_*)^2\right)\Phi\left(\lambda_{0*} + \lambda_*'\tilde{\sigma}^{-\frac{1}{2}}(f^* - u_*)\right).$$

Thus

$$(f^*|\boldsymbol{y}, \boldsymbol{x}^*) \sim GSN_1^2(u_*, \tilde{\sigma}, \lambda_{0*}, 1, \lambda_*, 0,1).$$

Proof of (ii). To find the expectation of f^*given $\boldsymbol{Y} = \boldsymbol{y}, X^* = \boldsymbol{x}^*$, we use Theorem (3.3.3).

Hence

$$E(f^*|\boldsymbol{Y} = \boldsymbol{y}, X^* = \boldsymbol{x}^*) = u_* + \frac{\varphi(\delta_0)}{\Phi(\delta_0)}\left(\tilde{\sigma}^{\frac{1}{2}}\lambda_*/\sqrt{1+\lambda_*^2}\right).$$

If we use the second part of Theorem (3.3.3), then

$$Var(f^*|\boldsymbol{Y} = \boldsymbol{y}, X^* = \boldsymbol{x}^*) = \tilde{\sigma}(1-h).$$

Remark: if $\boldsymbol{\lambda} = \mathbf{0}$ and $\boldsymbol{\gamma} = \mathbf{0}$ i.e. $(\boldsymbol{f}', f^*)' \sim GSN_{n+1}^2(\boldsymbol{u}, \boldsymbol{\Sigma}, \lambda_0, \sigma^2, \mathbf{0}, \mathbf{0}, \boldsymbol{\Sigma})$, then $(\boldsymbol{f}', f^*)' \sim N_{n+1}(\boldsymbol{u}, \boldsymbol{\Sigma})$. Using the last substitution, then

$u_* = \boldsymbol{u}'(\mathbf{V}_{11} + \tau^2\mathbf{I}_n)^{-1}\boldsymbol{y},\ \lambda_{00} = \lambda_0^+,\ \lambda_0^+ = \lambda_0, \lambda_*' = 0$ and $\tilde{\sigma} = \mathrm{R} - \boldsymbol{u}'(\mathbf{V}_{11} + \tau^2\mathbf{I}_n)^{-1}\boldsymbol{u}.$

Hence

$$(f^*|\boldsymbol{y}, X^* = \boldsymbol{x}^*) \sim N_1(u_*, \tilde{\sigma}).$$

So

$$(f^*|\boldsymbol{y}, \boldsymbol{x}^*) \sim GSN_1(\boldsymbol{u}'(\mathbf{V}_{11} + \tau^2\mathbf{I}_n)^{-1}\boldsymbol{y}, \mathrm{R} - \boldsymbol{u}'(\mathbf{V}_{11} + \tau^2\mathbf{I}_n)^{-1}\boldsymbol{u}),$$

i.e., the Gaussian process is a special case of the generalized skew Gaussian process.

5.3 Parameter Estimation

The parameters τ^2 , λ_0 , $\boldsymbol{\lambda}_1$ can be estimated using maximum likelihood method, i.e., by maximizing the function $L(\lambda_0^*, \tau^2, \lambda)$, where L denotes the likelihood of $\boldsymbol{Y} \sim GSN_{n+1}^2(\mathbf{0}, \boldsymbol{\Sigma} + \tau^2\mathbf{I}_n, \lambda_0^*, 1, \tilde{\boldsymbol{\lambda}}_1, \mathbf{0}, \mathbf{I}_n)$. To illustrate this method, we consider the model

$$\boldsymbol{Y} = \boldsymbol{X} + \boldsymbol{\epsilon},$$

where, $\mathbf{X} \sim GSN_n^2(\mathbf{0}, \boldsymbol{\Sigma}, \lambda_0, 1, \lambda_1 \mathbf{1}_n, \mathbf{0}, \mathbf{I}_n)$ and $\boldsymbol{\epsilon} \sim N_n(\mathbf{0}, \tau^2\mathbf{I}_n)$, where $\tau > 0$. Using Theorem (4.2.1), then $\boldsymbol{Y} \sim GSN_{n+1}^2(\mathbf{0}, \boldsymbol{\Sigma} + \tau^2\mathbf{I}_n, \lambda_0^*, 1, \tilde{\boldsymbol{\lambda}}_1, \mathbf{0}, \mathbf{I}_n)$, where

$$\lambda_0^* = \lambda_0 / \sqrt{1 + \lambda_1 \mathbf{1}_n' \boldsymbol{\Sigma}^{-\frac{1}{2}} \left(\boldsymbol{\Sigma}^{-1} + \frac{1}{\tau^2}\mathbf{I}_n\right)^{-1} \boldsymbol{\Sigma}^{-\frac{1}{2}} \lambda_1 \mathbf{1}_n}$$

and

$$\tilde{\boldsymbol{\lambda}}_1' = \lambda_1 \mathbf{1}_n' \boldsymbol{\Sigma}^{\frac{1}{2}} (\boldsymbol{\Sigma} + \tau^2 \mathbf{I}_n)^{-\frac{1}{2}} / \sqrt{1 + \lambda_1 \mathbf{1}_n' \boldsymbol{\Sigma}^{-\frac{1}{2}} \left(\boldsymbol{\Sigma}^{-1} + \frac{1}{\tau^2}\mathbf{I}_n\right)^{-1} \boldsymbol{\Sigma}^{-\frac{1}{2}} \lambda_1 \mathbf{1}_n}.$$

To estimate the parameter λ_0^*, τ^2 and $\tilde{\boldsymbol{\lambda}}_1$, we may maximize the function L with respect to the parameters λ_0^*, τ^2 and $\tilde{\boldsymbol{\lambda}}_1$, where

$$L(\boldsymbol{y}, \lambda_0^*, \tau^2, \tilde{\boldsymbol{\lambda}}_1) = \frac{exp\left(-\frac{1}{2}\boldsymbol{y}'(\boldsymbol{\Sigma} + \tau^2\mathbf{I}_n)^{-1}\boldsymbol{y}\right)}{(2\pi)^{\frac{n}{2}} \Phi(\delta_0) |\boldsymbol{\Sigma} + \tau^2 \mathbf{I}_n|^{\frac{1}{2}}} \Phi\left(\lambda_0^* + \tilde{\boldsymbol{\lambda}}_1' (\boldsymbol{\Sigma} + \tau^2\mathbf{I}_n)^{-\frac{1}{2}} \boldsymbol{y}\right).$$

If the covariance function depends on some parameters, we need to add these parameters to $L(\boldsymbol{y}, \lambda_0^*, \tau^2, \tilde{\boldsymbol{\lambda}}_1)$. Since no closed forms are available for the maximum likelihood estimators, then intensive statistical computations are required to obtain the values of these estimators. In this thesis, we will not be interested in this computation problem. So, we leave it to our future works that concerning the applications of our theory to real data.

5.4 Gaussian Process with Generalized Skew Gaussian Error

In this section, we give another generalization to the GPR by keeping the input function varies over the class of all sample path functions of a GP, while we assume that the error process $\epsilon(x)$ follows a GSGP. In this case, the data are assumed to have a GSG distribution. Under this setup, we wish to predict $f(x)$ at x^*.

Theorem 5.4.1. Consider the model

$$Y(x) = f(x) + \epsilon(x), \tag{5.2}$$

where for all n and $\boldsymbol{x}_1, \ldots, \boldsymbol{x}_n \in D$, $\boldsymbol{f} = (f(\boldsymbol{x}_1), \ldots, f(\boldsymbol{x}_n))' \sim N_n(\mathbf{0}, \boldsymbol{\Sigma})$ and

$\boldsymbol{\epsilon} = \left(\epsilon(\boldsymbol{x}_1), \ldots, \epsilon(\boldsymbol{x}_n)\right)' \sim GSN_n^2(\mathbf{0}, \tau^2\mathbf{I}_n, \lambda_0, 1, \boldsymbol{\lambda}_1, \mathbf{0}, \mathbf{I}_n), \tau > 0$, then

$$\boldsymbol{Y} = \left(Y(\boldsymbol{x}_1), \ldots, Y(\boldsymbol{x}_n)\right)' \sim GSN_n^2\left(\mathbf{0}, \boldsymbol{\Sigma}^{-\frac{1}{2}}(\boldsymbol{\Sigma} + \tau^2\mathbf{I}_n)\boldsymbol{\Sigma}^{-\frac{1}{2}}, \dot{\lambda}_0, 1, \ddot{\boldsymbol{\lambda}}_1, \mathbf{0}, \mathbf{I}_n\right),$$

where

$$\dot{\lambda}_0 = \lambda_0 / \sqrt{1 + \frac{1}{\tau^2}\boldsymbol{\lambda}_1'\boldsymbol{\Sigma}^{\frac{1}{2}}(\tau^{-2}\boldsymbol{\Sigma} + \mathbf{I}_n)^{-1}\boldsymbol{\Sigma}^{\frac{1}{2}}\boldsymbol{\lambda}_1}$$

and

$$\ddot{\lambda}_1' = \frac{1}{\tau}\lambda_1'\Sigma^{\frac{1}{2}}(\tau^{-2}\Sigma + \mathbf{I}_n)^{-\frac{1}{2}} \Big/ \sqrt{1 + \frac{1}{\tau^2}\lambda_1'\Sigma^{\frac{1}{2}}(\tau^{-2}\Sigma + \mathbf{I}_n)^{-1}\Sigma^{\frac{1}{2}}\lambda_1}.$$

Proof. Multiply the model (5.2) by $\boldsymbol{\Sigma}^{-\frac{1}{2}}$, then

$$\boldsymbol{\Sigma}^{-\frac{1}{2}}Y = \boldsymbol{\Sigma}^{-\frac{1}{2}}f + \boldsymbol{\Sigma}^{-\frac{1}{2}}\epsilon.$$

Let $\dot{Y} = \boldsymbol{\Sigma}^{-\frac{1}{2}}Y$, $\dot{f}(x) = \boldsymbol{\Sigma}^{-\frac{1}{2}}f$ and $\dot{\epsilon} = \boldsymbol{\Sigma}^{-\frac{1}{2}}\epsilon$, then we have

$$\dot{Y} = \dot{f} + \dot{\epsilon}.$$

It is easy to show that $\dot{f} \sim N_n(\mathbf{0}, \mathbf{I}_n)$ and $\dot{\epsilon} \sim GSN_n^2(\mathbf{0}, \boldsymbol{\Sigma}^{-1}\tau^2, \lambda_0, 1, \boldsymbol{\lambda}_1, \mathbf{0}, \mathbf{I}_n)$. So using Theorem (4.2.1), then

$$\dot{Y} \sim GSN_n^2\left(\mathbf{0}, \Sigma^{-1}\tau^2 + \mathbf{I}_n, \dot{\lambda}_0, 1, \ddot{\lambda}_1, \mathbf{0}, \mathbf{I}_n\right),$$

where

$$\dot{\lambda}_0 = \lambda_0 \Big/ \sqrt{1 + \frac{1}{\tau^2}\lambda_1'\Sigma^{\frac{1}{2}}(\tau^{-2}\Sigma + \mathbf{I}_n)^{-1}\Sigma^{\frac{1}{2}}\lambda_1}$$

and

$$\ddot{\lambda}_1' = \frac{1}{\tau}\lambda_1'\Sigma^{\frac{1}{2}}(\tau^{-2}\Sigma + \mathbf{I}_n)^{-\frac{1}{2}} \Big/ \sqrt{1 + \frac{1}{\tau^2}\lambda_1'\Sigma^{\frac{1}{2}}(\tau^{-2}\Sigma + \mathbf{I}_n)^{-1}\Sigma^{\frac{1}{2}}\lambda_1}.$$

To complete the proof, we will find the distribution of $\boldsymbol{\Sigma}^{\frac{1}{2}}\dot{Y}$. Since

$$M_{\Sigma^{\frac{1}{2}}\dot{Y}}(t) = M_{\dot{Y}}\left(\boldsymbol{\Sigma}^{\frac{1}{2}}\boldsymbol{t}\right),$$

$$= \frac{1}{\Phi(\delta_0)} exp\left(\frac{1}{2}\boldsymbol{t}'\boldsymbol{\Sigma}^{\frac{1}{2}}(\tau^2\Sigma^{-1} + \mathrm{I}_n)^{-1}\boldsymbol{\Sigma}^{\frac{1}{2}}\boldsymbol{t}\right)\Phi\left(\dot{\lambda}_0 + \ddot{\boldsymbol{\lambda}}_1'(\tau^2\boldsymbol{\Sigma}^{-1} + \tau^2\mathbf{I}_n)^{\frac{1}{2}}\boldsymbol{\Sigma}^{\frac{1}{2}}\boldsymbol{t}\right).$$

So the result follows.

Let $\boldsymbol{x}^*$ be a fixed input vector in the domain of $\boldsymbol{x}$. We are interested in predicting the value of $f(\boldsymbol{x})$ at $\boldsymbol{x}^*$. Let $f^* = f(\boldsymbol{x}^*)$ and $S = \{(\boldsymbol{x}_1, y_1), \dots, (\boldsymbol{x}_n, y_n)\}$ be observed data according to the model (5.2). Since $f(\boldsymbol{x})$ follows a GGP, then

$$(\boldsymbol{f}' \quad \dot{f})' \sim \mathrm{N}_{n+1}(\mathbf{0}, \boldsymbol{V}).$$

To obtain the predictive distribution of f^* given $\boldsymbol{x}^*$ and $\boldsymbol{y}$, we need the following lemma.

Lemma 5.4.1. Consider the model $y_i = f(\boldsymbol{x}_i) + \epsilon(\boldsymbol{x}_i), \boldsymbol{x}_i \in D, i = 1, \dots, n$, where $f(\boldsymbol{x})$ and $\epsilon(\boldsymbol{x})$ are the same in (5.2). Moreover, assume that $\epsilon(\boldsymbol{x})$ satisfies the weak third order stationary conditions. If $\boldsymbol{x}^* \in D$ and

$$\mathbf{V} = \begin{pmatrix} \mathbf{V}_{11} & \mathbf{V}_{12} \\ \mathbf{V}_{12}' & \mathbf{V}_{22} \end{pmatrix},$$

then

$$(\boldsymbol{y}' \quad f^*)' \sim GSN_{n+1}^2(\mathbf{0}, \mathbf{V} + \tau^2\tilde{\mathbf{I}}_{n+1}, \lambda_{01}, 1, \dot{\boldsymbol{\lambda}}_{11}, \mathbf{0}, \mathbf{I}_{n+1}).$$

Proof. Using the mgf, then

$$M_{\boldsymbol{y},f^*}(\boldsymbol{t}, s) = E(\exp(\boldsymbol{t}'\boldsymbol{y} + sf^*))$$

$$= M_{\boldsymbol{f},f^*}(\boldsymbol{t}, s)M_{\epsilon}(\boldsymbol{t}),$$

$$= \frac{1}{\Phi(\delta_0)} exp\left(\frac{1}{2}(\boldsymbol{t}' \quad s)\mathbf{V}\binom{\boldsymbol{t}}{s}\right)\Phi\left(\frac{\lambda_0 + \lambda_1 \mathbf{1}_n'(\tau^2\mathbf{I}_n)^{\frac{1}{2}}\boldsymbol{t}}{\sqrt{1+\lambda_1^2(n+1)}}\right) exp\left(\frac{1}{2}\boldsymbol{t}'\tau^2\mathbf{I}_n\boldsymbol{t}\right),$$

$$= \frac{1}{\Phi(\delta_0)} exp\left(\frac{1}{2}(\boldsymbol{t}' \quad s)(\mathbf{V} + \tau^2\tilde{\mathbf{I}}_{n+1})\binom{\boldsymbol{t}}{s}\right)\Phi\left(\frac{\lambda_0 + \lambda_1 \mathbf{1}_n'(\tau^2\mathbf{I}_n)^{\frac{1}{2}}\boldsymbol{t}}{\sqrt{1+\lambda_1^2(n+1)}}\right),$$

$$= \frac{1}{\Phi(\delta_0)} exp\left(\frac{1}{2}(\boldsymbol{t}' \quad s)(\mathbf{V} + \tau^2\tilde{\mathbf{I}}_{n+1})\binom{\boldsymbol{t}}{s}\right)\Phi\left(\frac{\lambda_0 + \tau(\lambda_1 \mathbf{1}_n' \quad 0)\binom{\boldsymbol{t}}{s}}{\sqrt{1+\lambda_1^2(n+1)}}\right).$$

Let

$$\lambda_{11}' = \tau(\lambda_1 \mathbf{1}_n' \quad 0)(\mathbf{V} + \tau^2\tilde{\mathbf{I}}_{n+1})^{-\frac{1}{2}},$$

then

$$M_{y,f^*}(\boldsymbol{t}, s) = \frac{exp\left(\frac{1}{2}(\boldsymbol{t}' \quad s)(\mathbf{V} + \tau^2\tilde{\mathbf{I}}_{n+1})\binom{\boldsymbol{t}}{s}\right)}{\Phi(\delta_0)}\Phi\left(\frac{\lambda_0 + \lambda_{11}'(\mathbf{V} + \tau^2\tilde{\mathbf{I}}_{n+1})^{\frac{1}{2}}\binom{\boldsymbol{t}}{s}}{\sqrt{1+\lambda_1^2(n+1)}}\right).$$

Now

$$\lambda_{11}'\lambda_{11} = \tau^2(\lambda_1 \mathbf{1}_n' \quad 0)(\mathbf{V} + \tau^2\tilde{\mathbf{I}}_{n+1})^{-1}\binom{\lambda_1 \mathbf{1}_n}{0},$$

$$= \tau^2(\lambda_1 \mathbf{1}_n' \quad 0)\begin{pmatrix} \mathbf{V}_{11} + \tau^2\mathbf{I}_n & \mathbf{V}_{12} \\ \mathbf{V}_{12}' & \mathbf{V}_{22} \end{pmatrix}^{-1}\binom{\lambda_1 \mathbf{1}_n}{0}.$$

To simplify the calculations, let

$$\boldsymbol{a} = -(\mathbf{V}_{11} + \tau^2\mathbf{I}_n)^{-1}\mathbf{V}_{12}(\mathbf{V}_{22} - \mathbf{V}_{12}'(\mathbf{V}_{11} + \tau^2\mathbf{I}_n)^{-1}\mathbf{V}_{12})^{-1}$$

and

$$\mathbf{L} = (\mathbf{V}_{22} - \mathbf{V}_{12}'(\mathbf{V}_{11} + \tau^2\mathbf{I}_n)^{-1}\mathbf{V}_{12})^{-1},$$

then

$$\boldsymbol{\lambda}_{11}'\boldsymbol{\lambda}_{11} = \tau^2(\lambda_1\mathbf{1}_n' \quad 0)\begin{pmatrix} (\mathbf{V}_{11} + \tau^2\mathbf{I}_n - \mathbf{V}_{12}\mathbf{V}_{22}^{-1}\mathbf{V}_{12}')^{-1} & \boldsymbol{a} \\ \boldsymbol{a}' & \mathbf{L} \end{pmatrix}\begin{pmatrix} \lambda_1\mathbf{1}_n \\ 0 \end{pmatrix},$$

$$= \tau^2\lambda_1^2\mathbf{1}_n'(\mathbf{V}_{11} + \tau^2\mathbf{I}_n - \mathbf{V}_{12}\mathbf{V}_{22}^{-1}\mathbf{V}_{12}')^{-1}\mathbf{1}_n.$$

Let

$$\mathbf{Q}_1 = \boldsymbol{V}_{11} - \boldsymbol{V}_{12}\boldsymbol{V}_{22}^{-1}\boldsymbol{V}_{12}'.$$

Then

$$\boldsymbol{\lambda}_{11}'\boldsymbol{\lambda}_{11} = \tau^2\lambda_1^2\mathbf{1}_n'(\mathbf{Q}_1 + \tau^2\mathbf{I}_n)^{-1}\mathbf{1}_n,$$

$$= \tau^2\lambda_1^2\mathbf{1}_n'(\tau^{-2}\mathbf{I}_n - \tau^{-2}[\tau^{-2}\mathbf{I}_n + \mathbf{Q}_1^{-1}]^{-1}\tau^{-2})\mathbf{1}_n,$$

$$= \lambda_1^2(n) - \tau^{-2}\lambda_1^2\mathbf{1}_n'[\tau^{-2}\mathbf{I}_n + \mathbf{Q}_1^{-1}]^{-1}\mathbf{1}_n.$$

Let

$$Q_2 = \lambda_1^2 + \tau^{-2}\lambda_1^2\mathbf{1}_n'[\tau^{-2}\mathbf{I}_n + \mathbf{Q}_1^{-1}]^{-1}\mathbf{1}_n,$$

then

$$\lambda_1^2(n+1) = \boldsymbol{\lambda}_{11}'\boldsymbol{\lambda}_{11} + Q_2.$$

To do more simplifications, let $\lambda_{01} = \frac{\lambda_0}{\sqrt{1+Q_2}}$, $\dot{\boldsymbol{\lambda}}_{11}' \frac{\lambda_{11}'}{\sqrt{1+Q_2}}$ and $\dot{\boldsymbol{t}}' = (\boldsymbol{t}' \quad s)$. Then

$$M_{y,f^*}(\boldsymbol{t},s)=\frac{exp\left(\frac{1}{2}\dot{\boldsymbol{t}}'(\mathbf{V}+\tau^2\tilde{\mathbf{I}}_{n+1})\dot{\boldsymbol{t}}\right)}{\Phi(\delta_0)}\Phi\left(\frac{\lambda_{01}+\dot{\boldsymbol{\lambda}}'_{11}(\mathbf{V}+\tau^2\tilde{\mathbf{I}}_{n+1})^{\frac{1}{2}}\dot{\boldsymbol{t}}}{\sqrt{1+\dot{\boldsymbol{\lambda}}'_{11}\dot{\boldsymbol{\lambda}}_{11}}}\right).$$

which implies

$$(y' \quad f^*)' \sim GSN^2_{n+1}\left(\mathbf{0},\mathbf{V}+\tau^2\tilde{\mathbf{I}}_{n+1},\lambda_{01},1,\dot{\boldsymbol{\lambda}}_{11},\mathbf{0},\mathbf{I}_{n+1}\right).$$

Theorem 5.4.2. Assume that, the result in Theorem (5.4.1) is satisfied, then under the following partitions

$$\mathbf{V}+\tau^2\tilde{\mathbf{I}}_{n+1}=\begin{pmatrix}\mathbf{V}_{11}+\tau^2\mathbf{I}_n & \boldsymbol{u}\\ \boldsymbol{u}' & \mathrm{R}\end{pmatrix} \text{ and } \dot{\boldsymbol{\lambda}}_{11}=(\boldsymbol{\lambda}'_y \quad l).$$

The conditional distribution of f^* given y and x^* is

$$(f^*|y,x^*)\sim GSN^2_{n+1}(\dot{\mu},\dot{\sigma},\xi_0,1,\xi,0,1),$$

where

$\dot{\mu}=\boldsymbol{u}'(\mathbf{V}_{11}+\tau^2\mathbf{I}_n)^{-1}\boldsymbol{y}, \dot{\sigma}=\mathrm{R}-\boldsymbol{u}'(\mathbf{V}_{11}+\tau^2\mathbf{I}_n)^{-1}\boldsymbol{u}, \xi_0=\xi_{00}+\xi\dot{\sigma}^{-\frac{1}{2}}\dot{\mu},$

$\xi_{00}=\lambda_{01}+\left(\boldsymbol{\lambda}'_y-l\mathrm{R}^{-1}\boldsymbol{u}'\right)(V_{11}+\tau^2\mathbf{I}_n-\boldsymbol{u}\mathrm{R}^{-1}\boldsymbol{u}')^{-1}\boldsymbol{y},$

and $\xi=\left(l-\boldsymbol{\lambda}'_y(\mathbf{V}_{11}+\tau^2\mathbf{I}_n)^{-1}\boldsymbol{u}\right)(\mathrm{R}-\boldsymbol{u}'(\mathbf{V}_{11}+\tau^2\mathbf{I}_n)^{-1}\boldsymbol{u})^{-\frac{1}{2}}.$

Proof. To proof this theorem, we follow the same approach in the proof of Theorem (5.3.1).

The expectation and the variance of f^* given y and x^*, respectively, are

$$E(f^*|Y=y,X^*=x^*)=\dot{\mu}+\frac{\varphi(\delta_0)}{\Phi(\delta_0)}\left(\frac{\dot{\sigma}^{\frac{1}{2}}\xi}{\sqrt{1+\xi^2}}\right),$$

$$Var(f^*|Y=y,X^*=x^*)=\dot{\sigma}\left(1-\dot{h}\right),$$

where

$$\dot{h}=\left(\left(\frac{\varphi(\delta_0)}{\Phi(\delta_0)\sqrt{1+\xi^2}}\right)^2+\frac{\delta_0\varphi(\delta_0)}{\Phi(\delta_0)(1+\xi^2)}\right)\xi^2.$$

5.5 Prediction at Random Input

In this section, we use the simple Monte-Carlo approach to approximate the prediction distribution of $f(x^*)$ at a random input x^*.

If $x^*\sim N_n(\mu_{x^*},\Sigma_{x^*})$, then the predictive distribution is obtained by integrating over the input distribution, i.e,

$$P(f^*|\mu_{x^*},\Sigma_{x^*},y)=\int P(f^*|x^*,y)\,P(x^*)dx^*,$$

where $P(f^*|x^*,y)$ is the pdf of

$$(f^*|y,x^*)\sim GSN_1^2(u_*,\tilde{\sigma},\lambda_{0*},1,\lambda_*,0,1),$$

Note that the mean and variance of $(f^*|y,x^*)$, respectively, are

$$u_*+\frac{\varphi(\delta_0)}{\Phi(\delta_0)}\left(\tilde{\sigma}^{\frac{1}{2}}\lambda_*/\sqrt{1+\lambda_*^2}\right),\qquad \tilde{\sigma}(1-h).$$

Since, $P(f^*|x^*, y)$ is complicated as a function of x^*, then the above integral is intractable. To approximate this integral, one may use the simple Monte-Carlo approach, i.e.

$$P(f^*|\mu_{x^*}, \Sigma_{x^*}, y) \approx \frac{1}{T}\sum_{t=1}^{T} P(f^*|x^{*t}, y),$$

where $x^{*1}, x^{*2}, \ldots, x^{*T}$ are independent observations from $P(x^*)$.

5.6 Simulation of GSGP

In this section, we present two algorithms to simulate random observations from a generalized skew-Gaussian distribution. Also these two algorithms can be employed to simulate a sample path of a GSGP.

Algorithm 1. This algorithm employs the definition of GSND to generate random observations from GSND. The steps of this algorithm are given as follows:

1. Generate an observation; say $(X_1, \ldots, X_n, X_{n+1})' = (X', X_{n+1})'$ from $N_{n+1}(\boldsymbol{\mu}, \boldsymbol{\Delta})$, where $\boldsymbol{\mu}$ and $\boldsymbol{\Delta}$ are the mean and the covariance matrix respectively.
2. If $\lambda_0 + \boldsymbol{\lambda}'\boldsymbol{X} \geq X_{n+1}$, then deliver $\boldsymbol{X}$ from $GSN_n^1(\lambda_0, \sigma^2, \boldsymbol{\lambda}, \boldsymbol{\gamma}, \boldsymbol{\Sigma})$.
3. If $\lambda_0 + \boldsymbol{\lambda}'\boldsymbol{X} < X_{n+1}$, go to step 1

Algorithm 2. In this algorithm, we apply the accept-reject method presented in Rubinstein (1981) to the pdf of $f(x)$. The accept-reject method assume that $f(x)$ can be written as

$f(x) = cg(x)h(x)$, where $c \geq 1, 0 < g(x) \leq 1$ and $h(x)$is a pdf. Hence, we need to rewrite $f(x)$ in the accept-reject form. Note that

$$f(x) = ch(x)g(x),$$

where

$$c = \left(\Phi(\delta_0)\right)^{-1}, h(\boldsymbol{x}) = \varphi_n(x; \mathbf{0}, \boldsymbol{\Sigma}) \text{ and } g(\boldsymbol{x}) = \Phi\left(\frac{\lambda_0 + (\boldsymbol{\lambda}' - \boldsymbol{\gamma}'\boldsymbol{\Sigma}^{-1})\boldsymbol{x}}{\sqrt{\sigma^2 - \boldsymbol{\gamma}'\boldsymbol{\Sigma}^{-1}\boldsymbol{\gamma}}}\right).$$

Note that, $c \geq 1$, since $0 < \Phi(\delta_0) \leq 1$, $h(\boldsymbol{x})$ is also pdf and $0 < g(\boldsymbol{x}) \leq 1$. So we give the following algorithm:

1. Generate U from $U(0,1)$.
2. Generate $\boldsymbol{Y}$ from the pdf $h(\boldsymbol{y})$.
3. If U $\leq g(\boldsymbol{Y})$, deliver $\boldsymbol{Y}$ as the variate generated from $f(\boldsymbol{x})$.
4. Go to step 1.

To simulate a sample path from a GSGP $X(\boldsymbol{t}), \boldsymbol{t} \in D$, we simulate a generalized skew-normal distribution on a grid of D, i.e, if $\{t_1, \dots, t_n\}$ represent a grid of D, then we simulate $(X(t_1), \dots, X(t_n))'$ from $GSN_n^1(\boldsymbol{\mu}, \lambda_0, \sigma^2, \boldsymbol{\lambda}, \boldsymbol{\gamma}, \boldsymbol{\Sigma})$, where $\boldsymbol{\mu}, \boldsymbol{\lambda}, \boldsymbol{\gamma}$ and $\boldsymbol{\Sigma}$ are given as in chapter five.

Chapter Six

Conclusions and Future Work

In this thesis, we have extended the *skew-normal distribution* of Arnold and Beaver (2002). Also, we have studied its properties such as, moment generating function, closure under marginal and conditioning. The concept of skewness for the random process was entered and was used to generalize the Gaussian process to the generalized-skew Gaussian process. We have shown that the new distribution is amenable to extend the two prediction problems: the ordinary kriging and the Gaussian process for regression. Moreover, we gave two algorithms to simulate the generalized skew Gaussian distribution.

Several future works can be established concerning prediction and the GSN distributions. For example, prediction of a vector of values of $f(x)$, i.e., $(f(x_1^*),\dots,f(x_m^*))'$ and prediction of the excursion set of $Y(t)$ in a domain D', i.e., the set $\{t \in D': Y(t) \geq y\}$, where y is a given threshold. Also several geometric characteristics of the excursion set could be studied for prediction. Moreover, the GSGPR can be used as a family of processes, indexed by the skewness parameters, to study the robustness of the GPR anti departure from Gaussianity. The prediction problem when x^* is random has not been attacked in details in this thesis, since it needs further working on approximation theory specifically edgeworth and saddle point approximations. So we leave it to future research.

The main computational problem in GPR is the inversion of the matrix $(\mathbf{K}+\tau^2\mathbf{I}_n)$ which has the complexity $O(n^3)$. The idea of Nyström approximation could be used to reduce the rank of ***K*** (William and Seeger, 2000). Other methods to treat this problem are given in (Smola and Schölkopf, 2004; Fowlkes et al., 2001). We believe that similar future researches could be of central interest but under GSGPR.

References

1. Adler, R. J. and Taylor, J. (2007). *Random Fields and Geometry*, Springer, Boston.

2. Allard, D. & Naveau, P. (2007). A new spatial skew-normal random field model. *Commun. Statist. – Theory & Methods* **36**, 1821-1834.

3. Alodat, M. T. and Al-Rawwash. (2009). Skew Gaussian random field. *Journal of Computational and Applied Mathematics*, **232**, 2, p 496-504.

4. Alodat and Aludaat. (2006). The generalized hyperbolic process. *Parazilian Journal of Probability and Statistics*, **22**, 1, p 1-8.

5. Arlond, B. C & Beaver, R. J. (2002). Skewed multivariate models related to hidden truncation and/ or selective reporting. *Test. Soc. Estadist. Investigation Oper*. **11**, 7-54.

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

7. Azzalini, A. (1986). Further results on a class of distributions which includes the normal ones. *Statisicat*. **XLVI**, 199-208.

8. Azzalini, A. (2005). The skew-normal distribution and related multivariate families (with discussion). *Scand. J. Statist.* **32**, 159-188.

9. Azzalini, A. & Capitanio, A. (1999). Statistical application of the multivariate skew normal distributions. *J. R. Stat. Soc., ser. B* **61**, 579-602.

10. Azzalini, A. and Dalla Valle, A. (1996). The multivariate skew normal distribution. *Biometrika*, **83**, 715–726.

11. Bucciantі, A. and V. Pawlowsky-Glahn. (2005). New perspectives on water chemistry and compositional data analysis. *Mathematical Gology* 37(7), 707-731.

12. Cressie, N. A. C. (1993), *Statistics for Spatial Data*, New York: Wiley.

13. Dominguez-Molina, J. A., Gonzalez-Farias, G., Gupta, A. K. (2003). The multivariate close skew normal distribution. Technical Report No. 03-12, Department of mathematics and statistics, Bowling Green State University, Bowling Green, OH.

14. Fowlkes, C., Belongie, S., and Malik, J. (2001). Efficient Spatiotemporal Grouping using the Nystrom Method. *In* proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, CVPR 2001. P. 172.

15. Genton, M. G. (2004a). Skew-symmetric and generalized skew-elliptical distributions. In Genton, M. G., editor, *Skew-elliptical distributions and their applications: a journey beyond normality*, chapter 5, pages 81-100. Chapman & Hall/CRC.

16. Gupta, A. K., Gonzalez-Fraias, G., & Dominguez-Molina, J. A. (2003). A multivariate skew normal distribution. *J. Multivariate Anal.* **89**, 181-190.

17. Kim, H.-M., HA, E., & Mallick, B. K. (2004). Spatial prediction of rainfall using skew-normal process. In Genton, M. G., editor, *skew-elliptical distribution and their applications: a journey beyond normality*, chapter 16, pages 279-289. Chapman & Hall/CRC.

18. Kim, H.-M. and Mallick, B. K. (2004b). A Bayesian prediction using the skew Gaussian distribution. *Journal of Statistical Planning and Inference*, **120**, 85-101.

19. Machado, U. E. B. (2002). *Statistical analysis of non-Gaussian environmental loads and responses*. Doctoral Thesis, Lund University, Sweden.

20. Neal, R. M. (1995). *Bayesian learning for Neural Networks*. Ph.D. thesis, Dep. of Computer Science, University of Toronto.

21. O'Hagan, A. (1978). Curve fitting and optimal design for prediction. *J. R. Stat. Soc., ser. B*, **40**, 1-42.

22. Rasmussen, C. E. (1996). *Evaluation of Gaussian processes and other methods for nonlinear regression*. Ph.D. thesis, Dept. of Computer Science, University of Toronto.

23. Rubinstein, R. Y. (1981). *Simulation and the Montecarlo Method*. John Wiley and Sons, New York.

24. Schott, J. R. (1997). *Matrix analysis for statistics*. Wiley *Series In probability and statistics*.

25. Smola, A. J. and Schölkopf, B. (2004). A Tutorial on Support Vector Regression. *Statistics and Computing*, **14** (3), 199-228.

26. Vernic, R. (2006). Multivariate skew-normal distribution with applications in insurance. *Insurance Mathematics & Economics* **38,** 413-426.

27. Williams, C. K. I. and Seeger, M. (2000). Effect of the input density distribution on the kernel-based classifiers. In Proceedings of the seventeenth International Conference on Machine Learning (ICML 2000). Morgan Kaufmann.

28. Zhang, H. and El-Sharaawi. (2009). On spatial skew Gaussian process and applications. *Environmetic,* **10**, 982-1002.

Appendix A

1. An $n \times n$ matrix $\mathbf{P}$ is orthogonal matrix if $\mathbf{P}'\mathbf{P} = \mathbf{I}_n$.

2. The sets of vectors $\boldsymbol{X}_1, \boldsymbol{X}_2, \dots, \boldsymbol{X}_n$ are linearly independent if and only if $\sum_{i=1}^{n} \alpha_i \boldsymbol{X}_i = 0$, then $\alpha_i = 0$, for every $i = 1, 2, \dots, n$.

3. A matrix $\mathbf{A}_{n\times n}$ is said to be a positive semi definite matrix if $\boldsymbol{x}'\mathbf{A}\boldsymbol{x} \geq 0, \forall\, \boldsymbol{x} \in \mathbb{R}^n$ and is said to be a positive definite matrix if $\boldsymbol{x}'\mathbf{A}\boldsymbol{x} > 0, \forall\, \boldsymbol{x} \in \mathbb{R}^n, \boldsymbol{x} \neq \mathbf{0}$.

4. The vector of one's, i.e., $\mathbf{1}_n$ is an eigenvector for every circulant matrix of dimension n.

5. If $\mathbf{A}$ is $n \times n$ *symmetric matrix, then there exist an orthogonal matrix* $\mathbf{P}$, such that

$$\mathbf{P}'\mathbf{A}\mathbf{P} = \mathbf{P}^{-1}\mathbf{A}\mathbf{P} = diag(\lambda_1, \dots, \lambda_n),$$

where $\lambda_i^{\prime s}$ are the n eigen-values of $\mathbf{A}$, and the column of $\mathbf{P}$ is the eigenvector corresponding to $\lambda_1^{\prime s}$.

6. A matrix $\mathbf{A}$ is positive definite symmetric if and only if, all eigen-values are positive.

7. If $\mathbf{A}$ is a positive definite symmetric matrix, then there exist a matrix $\mathbf{B}$, called the square root of $\mathbf{A}$ such that $\mathbf{A} = \mathbf{B}'\mathbf{B}$. Here, we may use $\mathbf{A}^{\frac{1}{2}}$ instead of $\mathbf{B}$.

8. If $\mathbf{A}$, $\mathbf{B}$ and $\mathbf{A} + \mathbf{B}$ are all $n \times n$ nonsingular matrices, then

$$(\mathbf{A} + \mathbf{B})^{-1} = \mathbf{A}^{-1} - (\mathbf{A}^{-1} + \mathbf{B}^{-1})^{-1}\mathbf{A}^{-1}.$$

9. Suppose that $\mathbf{A}$ is $m \times n$ and the positive integers m_1, m_2, n_1, n_2 are such that $m = m_1 + m_2$ and $n = n_1 + n_2$. Then we can write $\mathbf{A}$ as a partitioned *matrix in*

the form

$$\mathbf{A} = \begin{array}{cc} m_1 & m_2 & \\ \end{array} \begin{pmatrix} \mathbf{A}_{11} & \mathbf{A}_{12} \\ \mathbf{A}_{21} & \mathbf{A}_{22} \end{pmatrix} \begin{array}{c} n_1 \\ n_2 \end{array},$$

where $\mathbf{A}_{11}$ is $m_1 \times n_1$, $\mathbf{A}_{12}$ is $m_2 \times n_1$, $\mathbf{A}_{21}$ is $m_1 \times n_2$ and $\mathbf{A}_{22}$ is $m_2 \times n_2$.

10. Let $\mathbf{A}$ $m \times m$ matrix, and suppose that we partition the matrix $\mathbf{A}$ as in (10), with $n_1 = m_1$, $n_2 = m_2$. Also suppose that $\mathbf{A}, \mathbf{A}_{11}$ and $\mathbf{A}_{22}$ are non-singular matrices, then

$$\mathbf{A}^{-1} = \begin{pmatrix} \dot{\mathbf{A}}_{11} & \dot{\mathbf{A}}_{12} \\ \dot{\mathbf{A}}_{21} & \dot{\mathbf{A}}_{22} \end{pmatrix}$$

where

$$\dot{\mathbf{A}}_{11} = (\mathbf{A}_{11} - \mathbf{A}_{12}\mathbf{A}_{22}^{-1}\mathbf{A}_{21})^{-1}, \dot{\mathbf{A}}_{12} = -\mathbf{A}_{11}^{-1}\mathbf{A}_{12}(\mathbf{A}_{22} - \mathbf{A}_{21}\mathbf{A}_{11}^{-1}\mathbf{A}_{12})^{-1},$$

$$\dot{\mathbf{A}}_{21} = -\mathbf{A}_{22}^{-1}\mathbf{A}_{21}(\mathbf{A}_{11} - \mathbf{A}_{12}\mathbf{A}_{22}^{-1}\mathbf{A}_{21})^{-1} \text{ and } \dot{\mathbf{A}}_{22} = (\mathbf{A}_{22} - \mathbf{A}_{21}\mathbf{A}_{11}^{-1}\mathbf{A}_{12})^{-1}$$

11. If $\mathbf{A}$ is a positive definite, then

$$\boldsymbol{x}'\mathbf{A}^{-1}\boldsymbol{x} = \boldsymbol{x}_1'\mathbf{A}_{11}^{-1}\boldsymbol{x}_1 + (\boldsymbol{x}_2 - \mathbf{A}_{21}\mathbf{A}_{11}^{-1}\boldsymbol{x}_1)'\mathbf{B}_{22}(\boldsymbol{x}_2 - \mathbf{A}_{21}\mathbf{A}_{11}^{-1}\boldsymbol{x}_1),$$

where $\mathbf{B}_{22} = (\mathbf{A}_{22} - \mathbf{A}_{21}\mathbf{A}_{11}^{-1}\mathbf{A}_{12})^{-1}$.

12. If $\boldsymbol{X} = (x_1, \dots, x_n)'$, then

$$\frac{\partial}{\partial X} = \left(\frac{\partial}{\partial x_i}\right)_{i=1}^{n} = \left(\frac{\partial}{\partial x_1}, \dots, \frac{\partial}{\partial x_n}\right)'.$$

13. If $f(\boldsymbol{x}) = \boldsymbol{x}'\boldsymbol{a}$, then

$$\frac{\partial f(x)}{\partial x} = a.$$

14. If $f(x) = x'\mathbf{A}x$ and $\mathbf{A}$ is a symmetric matrix, then

$$\frac{\partial f(x)}{\partial x} = 2Ax.$$

15. $\frac{\partial f(x)}{\partial x'} = \left(\frac{\partial f(x)}{\partial x}\right)'$