University of Brasília
Institute of Exact Sciences
Department of Statistics

Dissertation

Multivariate asymmetric distributions on the unit hypercube: Properties and applications

by

Leonardo Santos da Cruz

Brasília, 09 September of 2024

Multivariate asymmetric distributions on the unit hypercube: Properties and applications

by

Leonardo Santos da Cruz

Dissertation submitted to the Departament of Statistics at the University of Brasília, as part of the requirements required to obtain the Master Degree in Statistics.

Advisor: Roberto Vila Gabriel
Co-Advisor: Helton Saulo Bezerra dos Santos

Brasília, 09 September 2024

Dissertation submitted to the Graduate Program in Statistics of the Departament of Statistics at the University of Brasília, as part of the requirements required to obtain the Master Degree in Statistics.

Approved by:

Prof. Dr. ROBERTO VILA GABRIEL
Advisor, EST/UnB

Prof. Dra. CAROLINA MARCHANT
Universidad Católica del Maule Talca/Chile

Prof. Dr. EDUARDO YOSHIO NAKANO
EST/UnB

Acknowledgments

I would like to express my deepest gratitude to God for the opportunity to complete this work. I sincerely thank my advisors, Prof. Dr. Roberto Vila and Prof. Helton Saulo, for providing me with valuable learning experiences, as well as their constant support and patience throughout the entire development of this research. To my family, I express my gratitude for their understanding and patience in the face of my absence and reduced contact. To my friends and colleagues in the postgraduate program, especially Melquisadec, Thiago, João Victor, Moisés, Rebeca, Estevão, and Mari, I am grateful for your invaluable friendship and companionship. I would also like to thank the Departments of Statistics and Mathematics at the University of Brasília for the opportunity to be part of the postgraduate program and for the solid foundation I received. My sincere thanks to all the professors in the Department of Statistics, with special acknowledgment to Dr. José A. Fiorucci, Prof. Raul Y. Matsushita, Prof. Gustavo L. Gilardoni, and Prof. Guilherme S. Rodrigues, for their excellent teaching and patience in addressing my questions. I also extend my gratitude to the administrative staff for their support in creating an efficient work environment, with special mention to the administrative team — Karen Luiza, Tathyanna Martins, Lucas, and Edenilson — for their constant support, availability, and patience.

My appreciation also goes to Professors Dra. Carolina Marchant and Dr. Eduardo Yoshio Nakano for their valuable insights, recommendations, and for kindly agreeing to participate in the evaluation committee of this work. I am equally grateful to Prof. Dr. Felipe Quintino and my colleague João Victor for their significant contributions to this research. I cannot forget to express my gratitude to my undergraduate professors, Dr. Igor Lima, Dr. Paulo Henrique da Costa, Dr. Raimundo Bastos, Dr. Luiz Amorim, Mr. Rubens Peres Foster, and Dra. Cristina Acciari, for their support during my undergraduate studies and for the encouragement that motivated me to join the postgraduate program.

This study was financed in part by the Coordenação de Aperfeiçoamento de Pessoal de Nível Superior - Brasil (CAPES) - Finance Code 001.

Resumo Expandido

DISTRIBUIÇÕES ASSIMÉTRICAS MULTIVARIADAS SOBRE O HIPERCUBO UNITÁRIO: PROPRIEDADE E APLICAÇÕES.

A simetria é crucial na modelagem de dados, pois muitas técnicas estatísticas, como testes de hipóteses e intervalos de confiança, assumem a normalidade ou simetria dos dados. Distribuições simétricas simplificam a análise e a interpretação. Em contextos multivariados, a simetria é avaliada através de momentos de ordem superior, matrizes de covariância e funções de densidade de probabilidade, por exemplo.

No contexto multivariado, onde a simetria está presente, algumas distribuições podem ser utilizadas para modelar os dados, como, por exemplo, distribuições esféricas e distribuições elípticas, ambas multivariadas. Porém, quando os dados apresentam algum grau de assimetria, que pode ser observado através de uma representação gráfica, por exemplo, funções classificadas de acordo com essas denominações podem não modelar de forma otimizada o conjunto de dados em estudo. Portanto, alguns erros podem ocorrer na análise decorrente do ajuste dessas distribuições.

Em geral, a assimetria está frequentemente presente em contextos multivariados, mas modelar dados multivariados que apresentam assimetria não é uma tarefa trivial. Para lidar com essas características, métodos específicos foram desenvolvidos. Alguns destes métodos baseiam-se em abordagens originalmente criadas para dados simétricos, generalizando assim modelos propostos anteriormente. Esses modelos são conhecidos por incorporarem certo grau de assimetria em modelos simétricos, o que facilita a descrição e ajuste de dados que possuem essa característica.

O primeiro capítulo do presente trabalho apresentará ferramentas básicas e consecutivamente mais sofisticadas relacionadas à modelagem de dados multivariados, com e sem simetria.

Serão definidas distribuições mais complexas, como a distribuição elíptica, e serão apresentados exemplos e resultados importantes relacionados a essas distribuições. Por fim, são discutidas distribuições que generalizam distribuições elípticas, incorporando a propriedade de modelar dados assimétricos. Estas distribuições assimétricas serão definidas e alguns exemplos serão apresentados.

No capítulo seguinte, uma nova família de distribuições assimétricas é apresentada. Inicialmente é apresentado o modelo do qual deriva esta família de distribuições. Este modelo consiste em uma relação condicional entre variáveis aleatórias, onde são incorporados parâmetros de locação, escala, assimetria e um parâmetro adicional que pode ser utilizado para refinar o ajuste do modelo ao conjunto de dados. Nessa fase do trabalho são discutidos aspectos importantes, como a definição da função densidade de probabilidade que pode ser derivada do modelo apresentado. Também são discutidas as possíveis configurações dessas funções, destacando como elas podem, dependendo dos intervalos, assumir a forma de uma função de densidade de probabilidade já conhecida, destacando o caráter generalista do modelo. Serão exploradas outras propriedades, incluindo os critérios de escolha das funções a utilizar no modelo, alguns casos especiais da função densidade de probabilidade, a sua representação gráfica, a não identificabilidade do modelo, em determinadas condições; os quantis marginais, entre outras características relevantes. Além disso, são apresentadas justificativas matemáticas para alguns fatos discutidos ao longo do texto. Por fim, é apresentada a função de máxima verossimilhança, com a caracterização explícita desta função e suas respectivas derivadas parciais, destacando a impossibilidade de descrever explicitamente os estimadores dos parâmetros em termos de expressões analíticas. Como consequência direta, as estimativas dos parâmetros precisarão ser obtidas utilizando métodos computacionais, que serão discutidos e detalhados no capítulo seguinte.

Na parte final do presente trabalho, são empregados estudos de simulação, bem como a aplicação de duas famílias de distribuições apresentadas no capítulo 3 em dados reais. O estudo de simulação foi realizado com versões da função de densidade de probabilidade representando a distribuição dos dados do modelo. A estimativa de máxima verossimilhança foi utilizada em conjunto com o algoritmo de Monte Carlo. As análises utilizadas para avaliar as estimativas dos parâmetros foram o viés relativo e o erro quadrático médio. Para melhor ilustrar os resultados, são apresentados gráficos que mostram o comportamento dessas duas métricas para cada um dos parâmetros. Além disso, diversas funções foram empregadas para realizar o estudo de simulação. Uma pequena seleção representativa dessas funções é apresentada no corpo principal do texto, enquanto as demais podem ser encontradas no apêndice deste trabalho.

A aplicação aos dados reais foi realizada com um conjunto de dados reais do software R. A estatística descritiva dos dados foi apresentada e comentada. Duas funções de densidade derivadas do modelo foram então ajustadas e o ajuste é avaliado usando algumas métricas, que são brevemente apresentadas e discutidas. Após a discussão dos dados, indica-se qual distribuição melhor se ajusta ao conjunto de dados com base nos critérios considerados e nas funções G escolhidas para o modelo. Por fim, são apresentadas conclusões quanto à aplicação dos dados e à estimação dos parâmetros dentro de uma perspectiva geral do trabalho desenvolvido.

Palavras-chave: Distribuição Multivariada G-elíptica-assimétrica; Distribuição Multivariada G-elíptica-assimétrica normal; Distribuição Multivariada G-elíptica-assimétrica t-Student.

Abstract

In this work, a family of multivariate asymmetric distributions over the unitary hypercube defined in terms of well-known symmetric elliptical distributions is proposed. Here we seek to study fundamental properties, such as the characterization of the density function for some types of distributions, as well as other properties, such as loss of identifiability, quantiles, conditional and marginal distributions, and moments. Furthermore, simulation studies were carried out to verify the asymptotic behavior of the estimated parameter values as the sample size increased. Finally, the developed model was used on real data where, using convenient metrics, the degree of quality of the model's adjustment to real data was verified.

Keywords: Multivariate extended G-skew-elliptical distribution; Multivariate extended G-skew-normal; Multivariate extended G-skew Student-t.

Contents

List of Tables

List of Figures

List of Symbols and Notations

$\mathbf{\Lambda}$	Orthogonal matrix
$\mathcal{O}$	Orthogonal group
ELL_n	Multivariate elliptical symmetric distribution
EUGSE_n	Multivariate extended unit-G-skew-elliptical distribution
$\boldsymbol{\lambda}$	Skewness parameter vector
$\boldsymbol{\mu}$	Constant vector
$\mathbf{\Sigma}$	Covariance Matrix
τ	Extension parameter (real number)
$g^{(n)}$	Density generator
PDF	Probability density function
CDF	Cumulative distribution function
EST_1	Univariate extended skew-Student-t
$\mathrm{S}_{\mathrm{ESN}_1}$	Univariate extended skew-Normal Standart

Chapter 1

Introduction

The formal properties of the normal distribution are fundamental to statistical theory. However, real-world data often exhibit characteristics such as skewness, multimodality, or censorship (Hill and Dixon, 1982). When extending the normal distribution to address these non-normal features, it becomes crucial to preserve its essential properties. One approach to modeling skewness is through the multivariate skew-normal distribution introduced by Azzalini and Dalla Valle, 1996, striking a balance between mathematical tractability and shape flexibility. Nevertheless, the skewness and kurtosis coefficients of the skew-normal distribution have limitations Azzalini, 1985, and it does not accommodate multimodality.

To address these challenges, Genton and Loperfido, 2002 introduced the generalized skew-elliptical (GSE) distributions. These GSE distributions serve as generalizations not only for skew-elliptical distributions (Azzalini and Capitanio, 1999; Branco and Dey, 2001) but also for other skewed extensions of normal distributions, including multivariate skew-t (Branco and Dey, 2001) and multivariate skew-Cauchy (Barry and Robert, 2000). Azzalini and Capitanio, 1999 proposed skew-elliptical densities as alternatives to skew-normal distributions, and Branco and Dey, 2001 provided a comprehensive discussion of such densities.

In general, skew-elliptical distributions, distinguished by their elliptical structure, define probability density functions within ellipsoids in a p-dimensional space. The introduction of asymmetry in skew-elliptical distributions is achieved through an asymmetry parameter.

In addition to introducing a multidimensional parameter that incorporates information about asymmetry, it is possible to treat this concept in a multidimensional way, also considering linear transformations represented by orthogonal matrices. These concepts can be synthesized through distributions that assimilate this information mathematically, known as spherical distributions. Generalizations of these distributions, which present broader characteristics and allow more flexible modeling of the data, are elliptical distributions. Finally, the association of a parameter that captures asymmetry with the concepts of spherical and elliptical distributions makes it possible to define asymmetric distributions. This work will seek to present a new family of asymmetric distributions that have a high degree of generality where, depending on the configuration of parameters belonging to their density function, it is possible to obtain already known asymmetric models and their respective properties.

In Chapter 3 , we will introduce a new family of multivariate asymmetric distributions over the unit hypercube, defined in terms of well-known elliptically symmetric distributions. To facilitate understanding, in Chapter 2, we will first present fundamental concepts before delving into the model that yields a family of skew-elliptical distributions and their respective properties.

Furthermore, the present work sought to computationally implement the functions derived from the developed model so that it was possible to establish a simulation study related to the evaluation of the parameters of its convergence to the real models as the sample size and the number of simulations increased (Monte Carlo method). Therefore, we sought to carry out the above study in principle for the unit-G-skew-normal extended density function and subsequently for the unit-G-skew-t-student extended density function. Once the simulation phase and its respective analyzes were completed, we sought to analyze the application of density functions to real data and the results arising from this application in terms of distribution function adjustments. In this sense, the data were used (enter the data) and the results are presented in Chapter 4.

Chapter 2

Preliminary Concepts

Symmetry is crucial in data modeling, as many statistical techniques, like hypothesis testing and confidence intervals, assume data normality or symmetry. Symmetric distributions simplify analysis and interpretation. In multivariate contexts, symmetry is assessed through higher-order moments, covariance matrices, and probability density functions.

In the multivariate context, where symmetry is present, some distributions can be used to model the data, such as, for example, spherical distributions and elliptical distributions, both multivariate. However, when the data presents some degree of asymmetry, which can be observed through a graphical representation, for example, functions classified according to these denominations may not optimally model the data set under study. Therefore, some mistakes may occur in an analysis arising from the adjustment of these distributions.

In general, asymmetry is often present in multivariate contexts. Modeling multivariate data that presents asymmetry is not a trivial task. To deal with these characteristics, specific methods have been developed. Some of these methods are based on approaches originally created for symmetric data, thus generalizing previously proposed models. These models are known for incorporating a certain degree of asymmetry into symmetric models, which facilitates the description and adjustment of data that have this characteristic.

This chapter will present both basic and more sophisticated tools related to modeling multivariate data, with and without symmetry. More complex distributions will be defined, such as the elliptical distribution, and examples and important results related to these distributions will be presented. Finally, distributions that generalize elliptical distributions will be discussed, incorporating the ability to model asymmetric data. These asymmetric distributions will be defined, and some examples will be presented.

2.1 Spherically and elliptically symmetric distribution

Definition 2.1.1. *An $n \times 1$ random vector $\boldsymbol{X} = (X_1, \ldots, X_n)^\top$ is said to have a spherically symmetric distribution (or simply spherical distribution) if for every $\boldsymbol{\Lambda} \in \mathcal{O}(n)$,*

$$\boldsymbol{\Lambda X} \stackrel{d}{=} \boldsymbol{X}. \tag{2.1.1}$$

Here $\stackrel{d}{=}$ means that the two sides have the same distribution (Section 1.2 of Fang, Kotz, and Ng, 1990), and $\mathcal{O}(n)$ denotes the set of $n \times n$ orthogonal matrices. The set $\mathcal{O}(n)$ is a group, called the orthogonal group, with the group operation being the ordinary matrix multiplication. In general, a spherical distribution is a probability distribution that is symmetrically distributed in all directions from a central point in a multidimensional space. Another way to define this family of distributions is as follows: A random variate $\boldsymbol{W}$ is spherically distributed if its distribution is invariant under rotations of $\mathbb{R}^n$, which is equivalent to having the stochastic representation

$$\boldsymbol{W} = R\boldsymbol{U},$$

where R is a non-negative random variable, $\boldsymbol{U}$ is uniform on sphere $\mathbb{S}_{n-1}$, and R and $\boldsymbol{U}$ are independent. Random variable R is called the generating variate, with the generating distribution F, and vector random variable $\boldsymbol{U}$ is the uniform base of the spherical distribution.

The characteristic function can be written in the form

$$\Psi_{\boldsymbol{W}}(\boldsymbol{\lambda}) = g(\boldsymbol{\lambda}^\top \boldsymbol{\lambda}),$$

where g is a scalar function called characteristic generator.

Theorem 2.1.1. *An n-dimensional random vector $\boldsymbol{X}$ has a spherical distribution if and only if its characteristic function $\Psi(\boldsymbol{t})$ satisfies one of the following equivalent conditions:*

1) $\Psi(\boldsymbol{\Lambda}^T \boldsymbol{t}) = \Psi(\boldsymbol{t})$, for any $\boldsymbol{\Lambda} \in \mathcal{O}(n)$ and $\boldsymbol{t} \in \mathbb{R}^n$;

2) There exists a function $\phi(\cdot)$ of a scalar variable such that $\Psi(\boldsymbol{t}) = \phi(\boldsymbol{t}^\top \boldsymbol{t})$.

Proof. Note that for any square matrix $\boldsymbol{A}$, the characteristic function of $\boldsymbol{AX}$ equals $\Psi(\boldsymbol{A}^\top t)$, that is,

$$\mathbb{E}[\exp(i\boldsymbol{t}^\top \boldsymbol{AX})] = \mathbb{E}[\exp(i(\boldsymbol{A}^\top \boldsymbol{t})^\top \boldsymbol{X})] = \Psi(\boldsymbol{A}^\top \boldsymbol{t}).$$

Thus Item 1) is equivalent to Definition (2.1.1). Now Item 2) implies 1), because

$$\Psi(\boldsymbol{\Lambda}^\top \boldsymbol{t}) = \phi((\boldsymbol{\Lambda}^\top \boldsymbol{t})^\top (\boldsymbol{\Lambda}^\top \boldsymbol{t})) = \phi(\boldsymbol{t}^\top \boldsymbol{\Lambda \Lambda t}) = \phi(\boldsymbol{t}^\top \boldsymbol{t}) = \Psi(\boldsymbol{t}).$$

Conversely, Item 1) implies that $\Psi(\boldsymbol{t})$ is an invariant function with respect to the group $\mathcal{O}(n)$ which has the maximal invariant $\phi(\boldsymbol{t}^\top\boldsymbol{t})$. Thus $\phi(\boldsymbol{t})$ must be a function of $\boldsymbol{t}^\top\boldsymbol{t}$, implying the statement in Item 2). $\square$

From now on, we will use the notation $\boldsymbol{X} \sim S_n(\phi)$ to indicate that $\boldsymbol{X}$ has a characteristic function of the form $\phi(\boldsymbol{t}^\top\boldsymbol{t})$, where $\phi(\cdot)$ is a function of a scalar variable known as the characteristic generator of the spherical distribution.

Example 2.1.1. *Let $\boldsymbol{U}^{(n)} = (U_1, \dots, U_n)^\top$ be a random vector uniformly distributed on the unit sphere $S_1^{n-1} \equiv \{\boldsymbol{u} \in \mathbb{R}^n : \|\boldsymbol{u}\| = 1\}$ in $\mathbb{R}^n$, where $\|\cdot\|$ is the Euclidean norm. For simplicity, we denote this fact by $\boldsymbol{U}^{(n)} \sim \mathcal{U}(S_1^{n-1})$. It follows that for any orthogonal matrix $\boldsymbol{\Lambda} \in \mathcal{O}(n)$, the vectors $\boldsymbol{\Lambda}\boldsymbol{U}^{(n)}$ and $\boldsymbol{U}^{(n)}$ are equal in distribution.*

Proof. It is well-known that $\boldsymbol{U}^{(n)} \sim \mathcal{U}(S_1^{n-1})$ has density given by $f_{\boldsymbol{U}^{(n)}}(\boldsymbol{u}) = 1/\mathrm{Vol}(S_1^{n-1})$, $\forall \boldsymbol{u} \in S_1^{n-1}$. By denoting $\boldsymbol{Y} \equiv \boldsymbol{\Lambda}\boldsymbol{U}^{(n)}$, we want to show that $\boldsymbol{Y} \overset{d}{=} \boldsymbol{U}^{(n)}$, for any orthogonal matrix $\boldsymbol{\Lambda}$. Indeed, if $\boldsymbol{y} = (y_1, \dots, y_n)^\top$ and $\boldsymbol{u} = (u_1, \dots, u_n)^\top$ are the corresponding values of $\boldsymbol{Y} = (Y_1, \dots, Y_n)^\top$ and $\boldsymbol{U}^{(n)} = (U_1, \dots, U_n)^\top$, respectively, we can write

$$\begin{cases} y_1 = a_{11}u_1 + \cdots + a_{1n}u_n \\ y_2 = a_{21}u_1 + \cdots + a_{2n}u_n \\ \vdots \qquad\qquad\quad \vdots \\ y_n = a_{n1}u_1 + \cdots + a_{nn}u_n. \end{cases}$$

Moreover, note that

$$\|\boldsymbol{y}\|^2 = \boldsymbol{y}^\top\boldsymbol{y} = (\boldsymbol{\Lambda}\boldsymbol{u})^\top\boldsymbol{\Lambda}\boldsymbol{u} = \boldsymbol{u}^\top(\boldsymbol{\Lambda}^\top\boldsymbol{\Lambda})\boldsymbol{u} = \boldsymbol{u}^\top\boldsymbol{u} = \|\boldsymbol{u}\|^2 = 1,$$

where we have used the fact that $\boldsymbol{\Lambda}^\top = \boldsymbol{\Lambda}^{-1}$. In simple terms, $\boldsymbol{y} \in S_1^{n-1}$.

If J_* is the inverse of the Jacobian matrix of $\boldsymbol{y} = (y_1, \dots, y_n)^\top$, then, we have

$$\det(J_*) = \begin{vmatrix} a_{11} & \dots & a_{1n} \\ \vdots & \ddots & \vdots \\ a_{n1} & \dots & a_{nn} \end{vmatrix} = |\boldsymbol{\Lambda}| = \pm 1,$$

where in the last equality above we have used the fact that every orthogonal matrix has a determinant equal to ± 1.

Jacobian method gives

$$f_{\boldsymbol{Y}}(\boldsymbol{y}) = f_{\boldsymbol{U}^{(n)}}(\boldsymbol{u})\,\frac{1}{|\det(J_*)|} = f_{\boldsymbol{U}^{(n)}}(\boldsymbol{u}) = \frac{1}{\mathrm{Vol}(S^{n-1})}, \quad \forall \boldsymbol{y} \in S_1^{n-1}.$$

Therefore, we have proven that

$$f_{\boldsymbol{Y}}(\boldsymbol{y}) = f_{\boldsymbol{U}^{(n)}}(\boldsymbol{y}), \quad \forall \boldsymbol{y} \in S_1^{n-1}.$$

The required result then follows readily. □

Following the same reasoning as the proof in Example 2.1.1, the following example can be verified.

Example 2.1.2. *Let $\boldsymbol{X}$ denote a random vector distributed uniformly inside the unit sphere in $\mathbb{R}^n$. The random vector $\boldsymbol{X}$ has a spherical distribution.*

Example 2.1.3. *Let $\boldsymbol{X} = (X_1, \ldots, X_n)^\top$ be a random vector distributed according to $N_n(\mathbf{0}, \boldsymbol{I})$. Since the characteristic function of X_1, is $\exp(-t_1^2/2)$, then the characteristic function of $\boldsymbol{X}$, denoted by $\Psi(\boldsymbol{t})$, is*

$$\Psi(\boldsymbol{t}) = \exp\left(-\frac{1}{2}\,\boldsymbol{t}^\top \boldsymbol{t}\right) = \phi(\boldsymbol{t}^\top \boldsymbol{t}), \quad \boldsymbol{t} = (t_1, \ldots, t_n)^\top \in \mathbb{R}^n,$$

where we have defined $\phi(u) = \exp(-u/2)$, $u \in \mathbb{R}$. Hence, from Item (2) of Theorem 2.1.1, $\boldsymbol{X}$ has a spherical distribution $S_n(\phi)$ with characteristic generator $\phi(u) = \exp(-u/2)$.

Definition 2.1.2. *An $n \times 1$ random vector $\boldsymbol{X}$ is said to have an elliptically symmetric distribution (or simply elliptical distribution) with parameters $\boldsymbol{\mu}_{n\times 1}$ and $\boldsymbol{\Sigma}_{n\times n}$ if*

$$\boldsymbol{X} \overset{d}{=} \boldsymbol{\mu} + \boldsymbol{A}^\top \boldsymbol{Y}, \quad \boldsymbol{Y} \sim S_k(\phi), \tag{2.1.2}$$

where $\boldsymbol{A}$ is a $k \times n$ matrix such that $\boldsymbol{A}^\top \boldsymbol{A} = \boldsymbol{\Sigma}$ with $\mathrm{rank}(\boldsymbol{\Sigma}) = k$. We shall write $\boldsymbol{X} \sim EC_n(\boldsymbol{\mu}, \boldsymbol{\Sigma}, \phi)$.

Example 2.1.4. *(The multinormal distribution). If a random vector $\boldsymbol{X}$ has the following decomposition:*

$$\boldsymbol{X} \overset{d}{=} \boldsymbol{\mu} + \boldsymbol{A}^\top \boldsymbol{Y},$$

where $\boldsymbol{\mu} \in \mathbb{R}^n$, $\boldsymbol{A}$ is a $m \times n$ matrix and $\boldsymbol{Y} \sim N_m(\mathbf{0}, \boldsymbol{I})$, then we say that $\boldsymbol{X}$ follows a multinormal distribution $N_m(\boldsymbol{\mu}, \boldsymbol{\Sigma})$ with $\boldsymbol{\Sigma} = \boldsymbol{A}^\top \boldsymbol{A}$. From Example 2.1.3, $\boldsymbol{Y} \sim S_k(\phi)$ with

$\phi(u) = \exp(-u/2)$, *and we have that* $\boldsymbol{X} \sim EC_n(\boldsymbol{\mu}, \boldsymbol{\Sigma}, \phi)$. *Equivalently,*

$$\boldsymbol{X} \stackrel{d}{=} \boldsymbol{\mu} + r\boldsymbol{A}^\top \boldsymbol{U}^{(k)},$$

where $r \stackrel{d}{=} ||y|| \sim \chi^2_n$ *(chi-squared distribution with* n *degrees of freedom) and* $\boldsymbol{U}^{(k)}$ *is distributed uniformly on the unit sphere surface in* $\mathbb{R}^k$.

Example 2.1.5. *Let* $\boldsymbol{Z} \sim N_n(\boldsymbol{0}, \boldsymbol{I})$ *and* $S \sim \chi^2_m$ *be independent. Let*

$$\boldsymbol{Y} = m^{\frac{1}{2}} \frac{\boldsymbol{Z}}{S}. \tag{2.1.3}$$

We say that $\boldsymbol{Y}$ *has a multivariate t-distribution with* m *degrees of freedom and write* $\boldsymbol{Y} \sim Mt_n(m, \boldsymbol{0}, \boldsymbol{I})$. *Evidently, we can write* (2.1.3) *as follows:*

$$\boldsymbol{Y} \stackrel{d}{=} m^{\frac{1}{2}} \frac{R\boldsymbol{U}^{(n)}}{S} = R^* \boldsymbol{U}^{(n)},$$

where $R \sim \chi_n$, S *and* $\boldsymbol{U}^{(n)}$ *are independent, and* $R^* = m^{1/2}R/S$ *(*R^*/n *has an F-distribution with* n *and* m *degrees of freedom). Thus,* $\boldsymbol{Y}$ *has a spherical distribution. Let*

$$\boldsymbol{X} = \boldsymbol{\mu} + \boldsymbol{A}^\top \boldsymbol{Y},$$

where $\boldsymbol{A}$ *is a* $n \times n$ *matrix and* $\boldsymbol{\mu} \in \mathbb{R}^n$.

We say that $\boldsymbol{X}$ *has a multivariate t-distribution with parameters* $\boldsymbol{\mu}, \boldsymbol{\Sigma} = \boldsymbol{A}^\top\boldsymbol{A}$, *and* m *degrees of freedom, and write* $\boldsymbol{X} \sim Mt_n(m, \boldsymbol{\mu}, \boldsymbol{\Sigma})$. *Clearly,* $Mt_n(m, \boldsymbol{\mu}, \boldsymbol{\Sigma}) = EC_n(\boldsymbol{\mu}, \boldsymbol{\Sigma}, \phi)$ *with a special* ϕ.

The following result shows the structure of any elliptically distributed random vector. The proof is adapted from (Armerin, 2017) and can be found in detail in that same reference.

Theorem 2.1.2. *Let* $\boldsymbol{\mu} \in \mathbb{R}^n$ *and let* $\boldsymbol{\Sigma}$ *be an* $n \times n$ *symmetric and positive semidefinite matrix. For an n-dimensional random vector* $\boldsymbol{X}$ *the following statements are equivalent.*

1. $\boldsymbol{X} \sim EC_n(\boldsymbol{\mu}, \boldsymbol{\Sigma}, \Psi)$.

2. *We have*

$$\boldsymbol{h}^\top \boldsymbol{X} \stackrel{d}{=} \boldsymbol{h}^\top \boldsymbol{\mu} + \sqrt{\boldsymbol{h}^\top \boldsymbol{\Sigma} \boldsymbol{h}} Z, \quad \forall \boldsymbol{h} \in \mathbb{R}^n,$$

where Z is a symmetric random variable with

$$\mathbb{E}[\exp(itZ)] = \Psi(t^2).$$

Proof. $[1 \implies 2]$ If $\boldsymbol{X} \sim EC_n(\boldsymbol{\mu}, \boldsymbol{\Sigma}, \Psi)$, then for every $\boldsymbol{h} \in \mathbb{R}^n$ and some matrix $\boldsymbol{A}$ such that $\boldsymbol{A}\boldsymbol{A}^\top = \boldsymbol{\Sigma}$, we have

$$\begin{aligned} \boldsymbol{h}^\top \boldsymbol{X} \stackrel{d}{=} \boldsymbol{h}^\top \boldsymbol{\mu} + \boldsymbol{h}^\top \boldsymbol{A}\boldsymbol{Y} = \boldsymbol{h}^\top \boldsymbol{\mu} + (\boldsymbol{A}^\top \boldsymbol{h})^\top \boldsymbol{Y} &\stackrel{d}{=} \boldsymbol{h}^\top \boldsymbol{\mu} + ||\boldsymbol{A}^\top \boldsymbol{h}||\tilde{Y} \\ &\stackrel{d}{=} \boldsymbol{h}^\top \boldsymbol{\mu} + \sqrt{\boldsymbol{h}^\top \boldsymbol{A}\boldsymbol{A}^\top \boldsymbol{h}}\tilde{Y} \\ &\stackrel{d}{=} \boldsymbol{h}^\top \boldsymbol{\mu} + \sqrt{\boldsymbol{h}^\top \boldsymbol{\Sigma} \boldsymbol{h}}\tilde{Y}. \end{aligned}$$

Since $\boldsymbol{Y}$ exhibits a spherical distribution, it follows that $\tilde{Y}$ is a symmetric random variable with a characteristic function:

$$\mathbb{E}[\exp(it\tilde{Y})] = \Psi(t^2).$$

$[2 \implies 1]$ If $\boldsymbol{X}$ has the property that

$$\boldsymbol{h}^\top \boldsymbol{X} \stackrel{d}{=} \boldsymbol{h}^\top \boldsymbol{\mu} + \sqrt{\boldsymbol{h}^\top \boldsymbol{\Sigma} \boldsymbol{h}} Z, \quad \forall \boldsymbol{h} \in \mathbb{R}^n,$$

where $\mathbb{E}[\exp(itZ)] = \Psi(t^2)$, then

$$\mathbb{E}[\exp(i\boldsymbol{h}^\top \boldsymbol{X})] = \exp(i\boldsymbol{h}^\top \boldsymbol{\mu})\mathbb{E}[\exp(i\sqrt{\boldsymbol{h}^\top \boldsymbol{\Sigma} \boldsymbol{h}})Z] = \exp(i\boldsymbol{h}^\top \boldsymbol{\mu})\Psi(\boldsymbol{h}^\top \boldsymbol{\Sigma} \boldsymbol{h}),$$

That is, $\boldsymbol{X} \sim EC_n(\boldsymbol{\mu}, \boldsymbol{\Sigma}, \Psi)$. □

2.2 Properties of elliptically symmetric distribution

2.2.1 Moments

The mean vector of a n-dimensional elliptical random vector $\boldsymbol{x}$, where $\boldsymbol{\Lambda}\boldsymbol{\Lambda}^T = \boldsymbol{\Sigma}$ corresponds to

$$\mathbb{E}(\boldsymbol{X}) = \mathbb{E}\left[\boldsymbol{\mu} + \boldsymbol{\mathcal{P}}\boldsymbol{\Lambda}\boldsymbol{U}^{(k)}\right] = \boldsymbol{\mu} + \boldsymbol{\Lambda}\mathbb{E}(\boldsymbol{\mathcal{P}})\mathbb{E}\left[\boldsymbol{U}^{(k)}\right],$$

where $\boldsymbol{\mathcal{P}}$ represents a vector that adjusts the magnitude of $\boldsymbol{\mathcal{U}}^{(k)}$. Here $\boldsymbol{\mathcal{P}}$ and $\boldsymbol{U}^{(k)}$ are supposed to be independent. Here we assume that $\mathbb{E}(\boldsymbol{\mathcal{P}})$ is finite. Since $\mathbb{E}\left[\boldsymbol{U}^{(k)}\right] = 0$, we obtain $\mathbb{E}(\boldsymbol{X}) =$

$\boldsymbol{\mu}$.

The covariance matrix of $\boldsymbol{X}$ is

$$\mathrm{Var}(\boldsymbol{X}) = \mathbb{E}\left[\left(\boldsymbol{\mathcal{P}}\boldsymbol{\Lambda}\boldsymbol{U}^{(k)}\right)\left(\mathcal{P}\boldsymbol{\Lambda}\boldsymbol{U}^{(k)}\right)^{\top}\right] = \mathbb{E}\left(\boldsymbol{\mathcal{P}}^2\right)\boldsymbol{\Lambda}\mathbb{E}\left[\boldsymbol{U}^{(k)}(\boldsymbol{U}^{(k)})^{\top}\right]\boldsymbol{\Lambda}^{\top},$$

provided $\mathbb{E}\left(\boldsymbol{\mathcal{P}}^2\right)$ is finite. Since $\sqrt{\chi_k^2}\boldsymbol{U}^{(k)} \sim N_k\left(0, \boldsymbol{I}_k\right)$ and

$$\boldsymbol{I}_k = \mathbb{E}\left[\left(\sqrt{\chi_k^2}\boldsymbol{U}^{(k)}\right)\left(\sqrt{\chi_k^2}\boldsymbol{U}^{(k)}\right)^{\top}\right] = \mathbb{E}\left(\chi_k^2\right)\mathbb{E}\left[\boldsymbol{U}^{(k)}(\boldsymbol{U}^{(k)})^{\top}\right] = k\mathbb{E}\left[\boldsymbol{U}^{(k)}(\boldsymbol{U}^{(k)})^{\top}\right],$$

therefore $\mathbb{E}\left[\boldsymbol{U}^{(k)}(\boldsymbol{U}^{(k)})^{\top}\right] = \boldsymbol{I}_k/k$ and thus

$$\mathrm{Var}(\boldsymbol{X}) = \frac{\mathbb{E}\left(\boldsymbol{\mathcal{P}}^2\right)}{k}\boldsymbol{\Sigma}.$$

Note that k refers to the number of components in $\boldsymbol{U}^{(k)}$, not necessarily the rank of $\boldsymbol{\Sigma}$ or the dimension of $\boldsymbol{X}$. Additionally, the dispersion matrix usually differs from the covariance matrix. The normal distribution is an exception: in this case, $\mathbb{E}\left(\boldsymbol{\mathcal{P}}^2\right) = \mathbb{E}\left(\chi_k^2\right) = k$, which implies that $\mathrm{Var}(\boldsymbol{X}) = \boldsymbol{\Sigma}$. However, by scaling $\boldsymbol{\mathcal{P}}$ with $\sqrt{k/\mathbb{E}\left(\boldsymbol{\mathcal{P}}^2\right)}$, we can always obtain a representation where $\mathrm{Var}(\boldsymbol{X}) = \boldsymbol{\Sigma}$ (see Bingham and Kiesel, 2002).

Consider a spherical random vector underlying a location-scale family with the stochastic representation:

$$\boldsymbol{X} \overset{\mathrm{d}}{=} \boldsymbol{\mathcal{P}}^{(n)}\boldsymbol{U}^{(n)}, \quad \forall n \in \mathbb{N},$$

where $\boldsymbol{U}^{(n)}$ is uniformly distributed on the unit sphere $\boldsymbol{\mathcal{S}}^{n-1}$, and $\boldsymbol{\mathcal{P}}^{(n)}$ is a scaling factor such that $\boldsymbol{X}$ always has the characteristic function $t \mapsto \phi\left(\boldsymbol{t}^{\top}\boldsymbol{t}\right)$. This implies that the characteristic generator ϕ is independent of n. Consequently, the characteristic function of the marginal cumulative distribution function (c.d.f.) of any component of $\boldsymbol{X}$ is given by $s \mapsto \phi\left(s^2\right)$, for $s \in \mathbb{R}$, irrespective of n. Therefore, the marginal distribution functions and their moments do not depend on the dimension n. As a result, if the second moment of $\boldsymbol{\mathcal{P}}^{(n)}$ is finite, it must be proportional to n.

Example 2.2.1. *(The 2nd moment of $\boldsymbol{\mathcal{P}}^{(n)}$ for the normal distribution) Since the generating variate of $\boldsymbol{X} \sim N_n\left(\boldsymbol{0}, \boldsymbol{I}\right)$ corresponds to $\sqrt{\chi_n^2}$, we obtain*

$$\mathbb{E}\left[\left(\boldsymbol{\mathcal{P}}^{(n)}\right)^2\right] = \mathbb{E}\left(\chi_n^2\right) = n.$$

The following theorem proves to be highly valuable for determining the asymptotic covariances of covariance matrix estimators for (generalized) elliptical distributions.

Theorem 2.2.1. *(Dickey and Chen, 1985). Let $\boldsymbol{X} = (X_1, \dots, X_n)^\top$ be a spherically distributed random vector with stochastic representation $\mathcal{P}\boldsymbol{U}^{(n)}$. Its mixed moment of order $(m_1, \dots, m_n)$ corresponds to*

$$\mathbb{E}\left(\prod_{i=1}^{n} X_i^{m_i}\right) = \frac{\mathbb{E}\left(\mathcal{P}^m\right)}{\left(\frac{n}{2}\right)^{(m/2)}} \prod_{i=1}^{n} \frac{m_i!}{2^{m_i}\left(\frac{m_i}{2}\right)!}$$

where $m = \sum_{i=1}^{n} m_i$ and every $m_1, \dots, m_n$ is supposed to be an even nonnegative integer. Here $(\cdot)^{(k)}$ is the 'rising factorial', i.e. $(x)^{(k)} = x(x+1)\cdots(x+k-1)$ for $k \in \mathbb{N}$ and $(x)^{(0)} = 1$. If at least one of the m_i 's is odd then the mixed moment vanishes. The proof can be seen in (Fang, Kotz, and Ng, 1990).

Theorem 2.2.2. *Denote the family of all possible characteristic generators for an $n \times 1$ random vector by $\Phi_n = \{\phi(\cdot) : \phi(t_1^2 + \dots + t_n^2)\}$ is an n-dimensional characteristic function. Note that $\Phi_i \subset \Phi_{i+1}$, where $1 \le i \le n$. A scalar function $\phi(\cdot)$ can determine an elliptically symmetric distributions $EC_n(\boldsymbol{\mu}, \boldsymbol{\Sigma}, \phi)$ for every $\boldsymbol{\mu} \in \mathbb{R}^n$ and $\boldsymbol{\Sigma} \ge 0$ with $rank(\boldsymbol{\Sigma}) = k$ if and only if $\phi \in \Phi_k$. The proof can be seen in (Fang, Kotz, and Ng, 1990).*

Corollary 2.2.3. *The following statements are equivalent*

1. $\boldsymbol{X} \sim EC_n(\boldsymbol{\mu}, \boldsymbol{\Sigma}, \phi)$ *with* $rank(\boldsymbol{\Sigma}) = k$,
2. $\boldsymbol{X} \overset{d}{=} \boldsymbol{\mu} + r\boldsymbol{A}^\top \boldsymbol{U}^{(k)}$,

where $r \ge 0$ is independent of $\boldsymbol{U}^{(k)}$, and $\boldsymbol{A}$ is a $k \times n$ matrix such that $\boldsymbol{A}^\top \boldsymbol{A} = \boldsymbol{\Sigma}$.

Corollary 2.2.4. *Assume that $\boldsymbol{X} \sim EC_n(\boldsymbol{\mu}, \boldsymbol{\Sigma}, \phi)$ with $rank(\boldsymbol{\Sigma}) = k$, then*

$$r^2 \overset{d}{=} (\boldsymbol{X} - \boldsymbol{\mu})^\top \boldsymbol{\Sigma}^{-1} (\boldsymbol{X} - \boldsymbol{\mu}),$$

where $\boldsymbol{\Sigma}^{-1}$ is the generalized inverse of $\boldsymbol{\Sigma}$.

Theorem 2.2.5. *Assume that $\boldsymbol{X}$ is nondegenerate.*

1. *If $\boldsymbol{X} \sim EC_n(\boldsymbol{\mu}, \boldsymbol{\Sigma}, \phi)$ and $\boldsymbol{X} \sim EC_n(\boldsymbol{\mu}^*, \boldsymbol{\Sigma}^*, \phi^*)$, then there exists a constant $c > 0$, such that*

$$\boldsymbol{\mu}^* = \boldsymbol{\mu}, \quad \boldsymbol{\Sigma}^* = c\boldsymbol{\Sigma}, \quad \phi^*(\cdot) = \phi(c^{-1}).$$

2. *If* $\boldsymbol{X} \stackrel{d}{=} \boldsymbol{\mu} + r\boldsymbol{A}^\top \boldsymbol{U}^{(k)} \stackrel{d}{=} \boldsymbol{\mu}^* + r^* \boldsymbol{A}^{*\top} \boldsymbol{U}^{(l^*)}$ *, where* $l \geq l^*$*, then there exists a constant* $c > 0$ *such that*

$$\boldsymbol{\mu}^* = \boldsymbol{\mu}, \quad \boldsymbol{A}^{*\top} \boldsymbol{A}^* = c\boldsymbol{A}^\top \boldsymbol{A}, \quad r^* \stackrel{d}{=} c^{-1/2} rb,$$

where $b \geq 0$ *is independent of* r *and* $b^2 \sim Beta\,(l^*/2, (l-l^*)/2)$ *if* $l > l^*$ *and* $b \equiv 1$ *if* $l = l^*$.

This theorem shows that $\boldsymbol{\Sigma}, \phi, r, \boldsymbol{A}$ are not unique unless we impose the condition that $|\boldsymbol{\Sigma}| = 1$ or that $|\boldsymbol{A}^\top \boldsymbol{A}| = 1$. The next theorem points out that any linear combination of elliptically distributed variates is still elliptical.

Theorem 2.2.6. *Assume that* $\boldsymbol{X} \sim EC_n(\boldsymbol{\mu}, \boldsymbol{\Sigma}, \phi)$ *with* $rank(\boldsymbol{\Sigma}) = k$*,* $\boldsymbol{B}$ *is an* $n \times m$ *matrix and* $\boldsymbol{v}$ *is an* $m \times 1$ *vector, then*

$$\boldsymbol{v} + \boldsymbol{B}^\top \boldsymbol{X} \sim EC_m(v + \boldsymbol{B}^\top \boldsymbol{\mu}, \boldsymbol{B}^\top \boldsymbol{\Sigma} \boldsymbol{B}, \phi).$$

Proof. The proof of theorem follows directly from relation

$$\boldsymbol{v} + \boldsymbol{B}^\top \boldsymbol{X} \stackrel{d}{=} (\boldsymbol{v} + \boldsymbol{B}^\top \boldsymbol{\mu}) + r(\boldsymbol{A}\boldsymbol{B})^\top \boldsymbol{U}^{(k)}.$$

□

Theorem 2.2.7. *Assume that* $\boldsymbol{X} \sim EC_n(\boldsymbol{\mu}, \boldsymbol{\Sigma}, \phi)$ *and* $\mathbb{E}\,(r^2) < \infty$*. Then*

$$\mathbb{E}(\boldsymbol{X}) = \boldsymbol{\mu}, \quad \mathrm{Cov}(\boldsymbol{X}) = \frac{\mathbb{E}\,(r^2)}{\mathrm{rank}(\boldsymbol{\Sigma})} \boldsymbol{\Sigma} = -2\phi'(0)\boldsymbol{\Sigma},$$
$$\Lambda_2(\boldsymbol{X}) = \mathbb{E}\left(\boldsymbol{X}\boldsymbol{X}^\top\right) = \boldsymbol{\mu}\boldsymbol{\mu}^\top - 2\phi'(0)\boldsymbol{\Sigma},$$

where $\phi'(0)$ *is the derivative of* ϕ *at the origin and*

$$\Gamma_1(\boldsymbol{x}) = i^{-1} \left.\frac{\partial \phi(\boldsymbol{t})}{\partial \mathbf{t}}\right|_{\boldsymbol{t}=0},$$
$$\Gamma_2(\boldsymbol{x}) = i^{-2} \left.\frac{\partial^2 \phi(\boldsymbol{t})}{\partial \boldsymbol{t} \partial \boldsymbol{t}^\top}\right|_{\boldsymbol{t}=0}.$$

Proof. Denoting $k = \mathrm{rank}(\boldsymbol{\Sigma})$, we have

$$\boldsymbol{X} \stackrel{d}{=} \boldsymbol{\mu} + r\boldsymbol{A}^\top \boldsymbol{U}^{(k)}.$$

As $\mathbb{E}(U^{(n)}) = 0$ and $\mathrm{Cov}(U^{(n)}) = I_n/n$ (Fang, Kotz, and Ng, 1990), it follows that

$$\mathbb{E}(\boldsymbol{X}) = \boldsymbol{\mu} + \mathbb{E}(r)\boldsymbol{A}^\top \mathbb{E}\left(\mathbf{U}^{(k)}\right) = \boldsymbol{\mu},$$

and

$$\begin{aligned}\mathrm{Cov}(\boldsymbol{X}) = \mathrm{Cov}\left(r\boldsymbol{A}^\top \mathbf{U}^{(k)}\right) &= \mathbb{E}\left(r^2\right)\boldsymbol{A}^\top \mathrm{Cov}\left(\boldsymbol{U}^{(k)}\right)\mathbf{A} \\ &= \mathbb{E}\left(r^2\right)\frac{1}{k}\boldsymbol{A}^\top \boldsymbol{I}_k \boldsymbol{A} = \frac{1}{k}\mathbb{E}\left(r^2\right)\boldsymbol{\Sigma}.\end{aligned}$$

□

In general, a given variable $\boldsymbol{X} \sim EC_n(\boldsymbol{\mu}, \boldsymbol{\Sigma}, \phi)$ does not necessarily possess a density (cf. Section 2.3, Fang, Kotz, and Ng, 1990). We shall now consider two cases:

1. $\boldsymbol{X}$ has a probability density function;

2. $\boldsymbol{\Sigma} > \mathbf{0}$ and $\mathbb{P}(\boldsymbol{X} = \boldsymbol{\mu}) = 0$.

A necessary condition that $\boldsymbol{X} \sim EC(\boldsymbol{\mu}, \boldsymbol{\Sigma}, \phi)$ possesses a probability density function is that $\mathrm{rank}(\boldsymbol{\Sigma}) = n$. In this case, the stochastic representation becomes

$$\boldsymbol{X} \stackrel{d}{=} \boldsymbol{\mu} + \boldsymbol{A}^\top \boldsymbol{Y},$$

where $\boldsymbol{A}$ is a nonsingular matrix with $\boldsymbol{A}^\top\boldsymbol{A} = \boldsymbol{\Sigma}$ and $\boldsymbol{Y} \sim S_n(\phi)$ (cf. 2.10, Fang, Kotz, and Ng, 1990).

The probability density function of $\boldsymbol{Y}$ is of the form $g\left(\boldsymbol{y}^\top\boldsymbol{y}\right)$, where $g(\cdot)$ is the probability density function generator. Since $\boldsymbol{X} = \boldsymbol{\mu} + \mathbf{A}^\top\boldsymbol{Y}$, the probability density function of $\boldsymbol{X}$ is of the form

$$|\boldsymbol{\Sigma}|^{-1/2} g\left((\boldsymbol{x} - \boldsymbol{\mu})^\top \boldsymbol{\Sigma}^{-1}(\boldsymbol{x} - \boldsymbol{\mu})\right).$$

In this case we shall sometimes use the notation $EC_n(\boldsymbol{\mu}, \boldsymbol{\Sigma}, g)$ instead of $EC_n(\boldsymbol{\mu}, \boldsymbol{\Sigma}, \phi)$. If $\boldsymbol{X}$ does not possess a probability density function, $\mathbb{P}(r = 0) = 0$ (cf. 2.33, Fang, Kotz, and Ng, 1990) and $|\boldsymbol{\Sigma}| > 0$, carrying out the same transformation $\boldsymbol{X} = \boldsymbol{\mu} + \boldsymbol{A}^\top\boldsymbol{Y}$, we obtain that $\mathbb{P}(\boldsymbol{Y} = \mathbf{0}) = 0$ and $\boldsymbol{Y}$ has all the marginal probability density functions and so does $\boldsymbol{X}$. In this case, the marginal density of $\boldsymbol{X}_{(k)} = (X_1, \ldots, X_k)^\top - \mu_{(k)}$, $1 \leqslant k < n$, where

$\boldsymbol{\mu}_{(k)} = (\mu_1, \ldots, \mu_k)^\top$, is given by

$$\frac{\Gamma(n/2)\left|\boldsymbol{\Sigma}_k\right|^{1/2}}{\Gamma((n-k)/2)\pi^{k/2}} \int_{\left(\boldsymbol{x}_{(k)}^\top \boldsymbol{\Sigma}_k^{-1} \boldsymbol{x}_{(k)}\right)^{1/2}}^{\infty} r^{-(n-2)} \left(r^2 - \boldsymbol{x}_{(k)}^\top \boldsymbol{\Sigma}_{\boldsymbol{k}}^{-1} \boldsymbol{x}_{(k)}\right)^{(n-k)/2-1} \mathrm{d}F(r), \tag{2.2.1}$$

where $\boldsymbol{\Sigma}_k$ is the first principal minor of $\boldsymbol{\Sigma}$ of dimension k (cf. 2.44, Fang, Kotz, and Ng, 1990).

Any function $g(\cdot)$ satisfying

$$\int_0^\infty y^{n/2-1} g(y) \mathrm{d}y < \infty,$$

defines a probability density functions (2.2.1) of an elliptically symmetric distribution with a normalizing constant C_n, where

$$C_n = \frac{\Gamma(n/2)}{2\pi^{n/2} \int_0^\infty r^{n-1} g\left(r^2\right) \mathrm{d}r}. \tag{2.2.2}$$

Example 2.2.2. *For $r, s > 0, 2N + n > 2$, let*

$$g(t) = t^{N-1} \exp\left(-r t^s\right).$$

From Equation (2.2.2)*,*

$$C_n = \frac{s \pi^{-n/2} r^{(2N+n-2)/(2s)} \Gamma\left(\frac{n}{2}\right)}{\Gamma\left(\frac{2N+n-2}{2s}\right)}.$$

The multivariate normal distribution is the special case $n = 1, s = 1, r = 1/2$. The case $s = 1$ was introduced and studied by (Kotz, 1975).

2.3 Multivariate skew-elliptical distributions

According (Branco and Dey, 2001) the multivariate skew-elliptical distribution is defined as follows:

Definition 2.3.1. *Consider $\boldsymbol{X} = (X_1, \ldots, X_k)^\top$ a random vector. Let $\boldsymbol{X}^* = \left(X_0, \boldsymbol{X}^\top\right)^\top$ be a $(k+1)$-dimensional random vector, such that $\boldsymbol{X}^* \sim EC_{k+1}\left(\boldsymbol{\mu}^*, \boldsymbol{\Sigma}; \phi\right)$, where $\boldsymbol{\mu}^* = (0, \boldsymbol{\mu})$, $\boldsymbol{\mu} = (\mu_1, \ldots, \mu_k)^\top$, ϕ is the characteristic function, and the scale parameter matrix $\boldsymbol{\Sigma}$ has the form*

$$\boldsymbol{\Sigma} = \begin{pmatrix} 1 & \boldsymbol{\delta}^\top \\ \boldsymbol{\delta} & \boldsymbol{\Omega} \end{pmatrix},$$

with $\boldsymbol{\delta} = (\delta_1, \ldots, \delta_k)^\top$. *Here* $\boldsymbol{\Omega}$ *is the scale matrix associated to the vector* $\boldsymbol{X}$. *We say that the random vector* $\boldsymbol{Y} = \boldsymbol{X}|X_0 > 0$ *has a skew-elliptical distribution and denote for* $\boldsymbol{Y} \sim SE_k(\boldsymbol{\mu}, \boldsymbol{\Omega}, \boldsymbol{\delta}; \phi)$*, where* $\boldsymbol{\delta}$ *is the skewness parameter.*

If the probability density function of the random vector $\boldsymbol{X}^*$ exists and $P(\boldsymbol{X}^* = 0) = 0$, then the p.d.f. of $\boldsymbol{Y}$ will be of the form

$$f_{\boldsymbol{Y}}(\boldsymbol{y}) = 2 f_{g^{(k)}}(\boldsymbol{y}) F_{g_{q(y)}}\left(\lambda^\top(\boldsymbol{y} - \boldsymbol{\mu})\right), \tag{2.1}$$

where $f_{g^{(k)}}(\cdot)$ is the p.d.f. of $EC_k\left(\boldsymbol{\mu}, \boldsymbol{\Omega}; g^{(k)}\right)$ and $F_{g_{q(z)}}$ is the c.d.f. of $El_1\left(0, 1; g_{q(\mathbf{z})}\right)$, with

$$\lambda^\top = \frac{\boldsymbol{\delta}^\top \boldsymbol{\Omega}^{-1}}{\sqrt{1 - \boldsymbol{\delta}^\top \boldsymbol{\Omega}^{-1} \boldsymbol{\delta}}}, \tag{2.2}$$

$$g^{(k)}(u) = \frac{2\pi^{k/2}}{\Gamma(k/2)} \int_0^\infty g^{(k+1)}(r^2 + u) r^{k-1} \mathrm{d}r, \quad u \geqslant 0, \tag{2.3}$$

$$g_{q(\boldsymbol{y})}(u) = \frac{g^{(k+1)}(u + q(\boldsymbol{y}))}{g^{(k)}(q(\boldsymbol{y}))}, \tag{2.4}$$

and $q(\boldsymbol{y}) = (\boldsymbol{y} - \boldsymbol{\mu})^T \boldsymbol{\Omega}^{-1} (\boldsymbol{y} - \boldsymbol{\mu})$. In this case, we denote $\boldsymbol{Y} \sim SE_k(\boldsymbol{\mu}, \boldsymbol{\Omega}, \boldsymbol{\delta}; g^{(k+1)})$, where $g^{(k+1)}$ is the density generator function. The notation $\mathbf{X} \sim El_k(\boldsymbol{\mu}, \boldsymbol{\Sigma}; \phi)$ to indicate that $\mathbf{X}$ is a k-dimensional random vector, elliptically distributed with location vector $\boldsymbol{\mu} \in \mathbb{R}^k$ and a $k \times k$ (positive definite) dispersion matrix $\boldsymbol{\Sigma}$ and characteristic function (c.f.) ϕ.

2.3.1 Examples of Skew-Elliptical Distributions

A mixed distribution can be created by combining two or more probability distributions. This involves extracting random variables from multiple populations to compose a new composite distribution. The original distributions (which are created to create the new distribution) can be univariate or multivariate, but the mixed distribution must maintain consistent dimensionality across all components. Furthermore, the constituent distributions must be of the same type, that is, all discrete or all continuous. Several examples of distorted elliptical distributions will be presented below.

Skew-Scale Mixture of Normal Distribution

In (Liu and Dey, 2004) the Skew-Scale Mixture of Normal Distribution density function is presented. It has the following configuration.

$$\begin{aligned} f_{\boldsymbol{Y}}(\boldsymbol{y}) =& 2|\boldsymbol{\Sigma}|^{-1/2}\int_{-\infty}^{\boldsymbol{\alpha}^\top(\boldsymbol{y}-\boldsymbol{\mu})}\int_0^\infty [2\pi K(\eta)]^{-(k+1)/2}\exp\left(-\frac{r^2+(\boldsymbol{y}-\boldsymbol{\mu})^\top\boldsymbol{\Sigma}^{-1}(\boldsymbol{y}-\boldsymbol{\mu})}{2K(\eta)}\right)\mathrm{d}H(\eta)\mathrm{d}r \\ =& 2\int_0^\infty |\boldsymbol{\Sigma}|^{-1/2}[2\pi K(\eta)]^{-k/2}\exp\left(-\frac{(\boldsymbol{y}-\boldsymbol{\mu})^\top\boldsymbol{\Sigma}^{-1}(\boldsymbol{y}-\boldsymbol{\mu})}{2K(\eta)}\right)\Phi\left(\frac{\boldsymbol{\alpha}^\top(\boldsymbol{y}-\boldsymbol{\mu})}{K(\eta)^{1/2}}\right)\mathrm{d}H(\eta). \end{aligned}$$

Thus the probability density function reduces to

$$f_{\boldsymbol{Y}}(\boldsymbol{y}) = 2\int_0^\infty \phi_k(\boldsymbol{y};\boldsymbol{\mu},K(\eta)\boldsymbol{\Sigma})\Phi\left(\frac{\boldsymbol{\alpha}^\top(\boldsymbol{y}-\boldsymbol{\mu})}{K(\eta)^{1/2}}\right)\mathrm{d}H(\eta).$$

One particular case of this distribution is the skew-normal distribution, for which H is degenerate, with $K(\eta) = 1$. In this case the corresponding probability density function is given by

$$2\phi_k(\boldsymbol{y};\boldsymbol{\mu},\boldsymbol{\Sigma})\Phi(\boldsymbol{\alpha}^\top(\boldsymbol{y}-\boldsymbol{\mu})).$$

Thus, if $g^{(k+1)}(\cdot)$ serves as the density generating function for a scale mixture of normal distributions, it follows that $f_{\boldsymbol{Y}}(\cdot)$ is once again a scale mixture of skew-normal distributions. The subsequent examples illustrate particular instances of the skew-scale mixture of normal distributions.

Skew-Finite Mixture of Normal

If the density generator function is

$$g^{(k+1)}(u) = \sum_{i=1}^n p_i\,[2\pi K\,(\eta_i)]^{-(k+1)/2}\exp(-u/2K\,(\eta_i)), \quad u \geq 0,$$

with $0 \leq p_i \leq 1$ and $\sum_{i=1}^n p_i = 1$, then the distribution H is a discrete measure on $\{\eta_1,\ldots,\eta_n\}$ with probabilities $p_1,\ldots,p_n$, respectively (Liu and Dey, 2004). The probability density function of the skew-finite mixture of normal is given by

$$f_{\boldsymbol{Y}}(\boldsymbol{y}) = 2\sum_{i=1}^n p_i\phi_k\left(\boldsymbol{y};\boldsymbol{\mu},K\,(\eta_i)\,\boldsymbol{\Sigma}\right)\Phi\left(\frac{\boldsymbol{\alpha}^\top(\boldsymbol{y}-\boldsymbol{\mu})}{K\,(\eta_i)^{1/2}}\right),$$

which is again a finite mixture of skew-normal distributions. In this case, for simplicity, often take $K(\eta_i) = 1, i = 1, \ldots, n$.

Skew-Logistic Distribution

The density generator function is

$$g^{(k+1)}(u) = \frac{\exp(-u)}{1 + \exp(-u)}, \quad u > 0.$$

(Choy, 1995) highlighted that the logistic distribution represents a specific instance of the scale mixture of normal distributions, when $K(\eta) = 4\eta^2$ and η follows an asymptotic Kolmogorov distribution with a probability density function.

$$f(\eta) = 8 \sum_{k=1}^{\infty} (-1)^{k+1} k^2 \eta \exp(-2k^2\eta^2).$$

2.4 Parameter estimation

2.4.1 The Maximum Likelihood Method

There are several ways to estimate distribution parameters. These forms depend on some circumstances, such as the approach you want to apply. In the case of frequentist approaches, one of the most traditional ways to estimate parameters is through the optimization of the maximum likelihood function. This section will seek to detail this method, which will be used in future computational analyses. The examples provided, along with additional intriguing cases, can be found in (Bolfarine and Sandoval, 2010).

Definition 2.4.1. *Let $X_1, \ldots, X_n$ be a random sample of size n of the random variable X with probability density function $f(x|\theta)$, with $\theta \in \Theta$, where Θ is the parameter space. The likelihood function for θ corresponding to the observed random sample is given by*

$$L(\theta; \boldsymbol{x}) = \prod_{i=1}^{n} f(x_i|\theta). \tag{2.4.1}$$

Definition 2.4.2. *The maximum likelihood estimator of θ (case exists) is the value $\widehat{\theta} \in \Theta$ that maximizes the likelihood function $L(\theta; \boldsymbol{x})$.*

The natural logarithm of the likelihood function of θ is denoted by

$$l(\theta; \boldsymbol{x}) \equiv \log L(\theta; \boldsymbol{x}). \tag{2.4.2}$$

It is not difficult to verify that the value of θ that maximizes the likelihood function $L(\theta; \boldsymbol{x})$ also maximizes $l(\theta; \boldsymbol{x})$ given by (2.4.2). Furthermore, in the uniparametric case where Θ is an interval of the line and $l(\theta; \boldsymbol{x})$ is derivable, the maximum likelihood estimator can be found as the root of the likelihood equation

$$l'(\theta; \boldsymbol{x}) \equiv \frac{\partial l(\theta; \boldsymbol{x})}{\partial \theta} = 0. \tag{2.4.3}$$

In certain straightforward cases, the solution to the likelihood equation can be determined explicitly. However, in more complex scenarios, the solution to equation (2.4.3) typically requires numerical methods. To verify that the solution to equation (2.4.3) corresponds to a maximum, it is essential to confirm whether

$$l''(\widehat{\theta}; \boldsymbol{x}) \equiv \left. \frac{\partial^2 \log L(\theta; \boldsymbol{x})}{\partial \theta^2} \right|_{\theta=\widehat{\theta}} < 0. \tag{2.4.4}$$

In cases where Θ is discrete or where the maximum of $l(\theta; \boldsymbol{x})$ occurs at the boundary of Θ, the maximum likelihood estimator cannot be derived from the solution of (2.4.3). In these instances, the maximum is determined by directly examining the likelihood function (2.4.1).

Example 2.4.1. *Let $X_1, \ldots, X_n$ be a random sample from the distribution of the random variable $X \sim N(\mu, 1)$. In this case, the likelihood function is given by*

$$L(\mu; x) = \left(\frac{1}{\sqrt{2\pi}}\right)^n \exp\left[-\frac{1}{2}\sum_{i=1}^{n}(x_i - \mu)^2\right],$$

with $\Theta = \{\mu : -\infty < \mu < \infty\}$. As

$$l(\mu; x) = -n \log \sqrt{2\pi} - \frac{1}{2}\sum_{i=1}^{n}(x_i - \mu)^2,$$

it follows from (2.4.3) *that the likelihood equation is given by*

$$\sum_{i=1}^{n}(x_i - \widehat{\mu}) = 0.$$

Then the maximum likelihood estimator of μ is given by

$$\widehat{\mu} = \frac{1}{n}\sum_{i=1}^{n} X_i = \overline{X}.$$

It is not difficult to verify in this case that (2.4.4) *is satisfied.*

In certain instances, particularly when the likelihood is linked to more intricate models, the likelihood function may not have a clear analytical solution. In these situations, maximum likelihood estimators are typically obtained through numerical methods. Let $U(\theta)$ represent the score function, defined as

$$U(\theta) \equiv \frac{\partial \log L(\theta; x)}{\partial \theta}$$

we have that, for the maximum likelihood estimator $\widehat{\theta}$,

$$U(\widehat{\theta}) = 0,$$

so that, by expanding $U(\widehat{\theta})$ in Taylor series around a point θ_0, we obtain

$$0 = U(\widehat{\theta}) \cong U\left(\theta_0\right) + (\widehat{\theta} - \theta_0)U'\left(\theta_0\right).$$

That is, we arrive at the equation

$$\widehat{\theta} \cong \theta_0 - \frac{U\left(\theta_0\right)}{U'\left(\theta_0\right)}. \tag{2.4.5}$$

From the equation (2.4.5), we obtain the iterative procedure (Newton-Raphson)

$$\theta_{j+1} = \theta_j - \frac{U\left(\theta_j\right)}{U'\left(\theta_j\right)}, \tag{2.4.6}$$

which begins with the value θ_0, followed by a new value θ_1 obtained from (2.4.6), and continues iteratively until the process stabilizes. This occurs when, for a given small ϵ, the condition $|\theta_{j+1} - \theta_j| < \epsilon$ is satisfied. In this scenario, the point $\widehat{\theta}$ at which the process stabilizes is considered the maximum likelihood estimator of θ. In some cases, substituting $U'\left(\theta_j\right)$ in (2.4.5) with $\mathbb{E}\left[U'\left(\theta_j\right)\right]$, which represents the Fisher information at θ_j corresponding to the observed sample, multiplied by -1, can significantly simplify the procedure. This approach is known as the score method.

2.4.2 Multiparametric case

In the previous sections, we explored how to derive maximum likelihood estimators and examined their properties when the likelihood function depends on a single parameter. In this section, we will address scenarios where $\boldsymbol{\theta} = (\theta_1, \ldots, \theta_r)^\top$, $r \in \mathbb{N}$, meaning that the likelihood function involves two or more parameters. The parameter space will be denoted by Θ. When the regularity conditions are met, the maximum likelihood estimators of $\theta_1, \ldots, \theta_r$ can be obtained by solving the following equations.

$$\frac{\partial \log L(\boldsymbol{\theta}; \boldsymbol{x})}{\partial \theta_i} = 0, \quad i = 1, \ldots, r.$$

In cases where the support of the distribution of X depends on $\boldsymbol{\theta}$ or the maximum occurs at the boundary of Θ, the maximum likelihood estimator is generally obtained by inspecting the graph of the likelihood function, as in the case uniparametric. In cases where the likelihood function depends on two parameters, θ_1 and θ_2, by using the equation

$$\frac{\partial \log L\left(\theta_1, \theta_2; \boldsymbol{x}\right)}{\partial \theta_1} = 0,$$

we obtain a solution for θ_1 as a function of θ_2, which we can denote by $\widehat{\theta}_1\left(\theta_2\right)$. Substituting the solution for θ_1 into the joint likelihood, we now have a function of just θ_2, that is,

$$g\left(\theta_2; \boldsymbol{x}\right) = l(\widehat{\theta}_1\left(\theta_2\right), \theta_2; \boldsymbol{x}),$$

referred to as the profiled likelihood of θ_2, which can be utilized to obtain the maximum likelihood estimator of θ_2. The optimization of $g\left(\theta_2; \boldsymbol{x}\right)$ can then be performed in the standard manner, that is, by differentiation, when feasible.

Example 2.4.2. *Let $X_1, \ldots, X_n$ be a random sample of the random variable $X \sim N\left(\mu, \sigma^2\right)$, where μ and σ^2 are unknown. We then have that $\boldsymbol{\theta} = (\mu, \sigma^2)$, with*

$$L(\boldsymbol{\theta}; \boldsymbol{x}) = \left(\frac{1}{2\pi\sigma^2}\right)^{n/2} \exp\left[-\sum_{i=1}^{n} \frac{\left(x_i - \mu\right)^2}{2\sigma^2}\right],$$

so that

$$l\left(\boldsymbol{\theta}; \boldsymbol{x}\right) = -\frac{n}{2}\log(2\pi\sigma^2) - \sum_{i=1}^{n} \frac{\left(x_i - \mu\right)^2}{2\sigma^2}.$$

Consequently,

$$\frac{\partial l\left(\boldsymbol{\theta};\boldsymbol{x}\right)}{\partial\mu}=2\sum_{i=1}^{n}\frac{\left(x_i-\widehat{\mu}\right)}{2\sigma^2}=0,$$

which leads to the estimator $\widehat{\mu}=\overline{X}$. Therefore, the logarithm of the profiled likelihood of σ^2 is given by

$$g\left(\sigma^2;\boldsymbol{x}\right)=-\frac{n}{2}\log(2\pi\sigma^2)-\frac{1}{2\sigma^2}\sum_{i=1}^{n}\left(x_i-\overline{x}\right)^2.$$

Then the maximum likelihood estimator of σ^2 is obtained as a solution to the equation

$$\frac{\partial g\left(\sigma^2;\boldsymbol{x}\right)}{\partial\sigma^2}=-\frac{n}{2\hat{\sigma}^2}+\sum_{i=1}^{n}\frac{\left(x_i-\overline{x}\right)^2}{2\hat{\sigma}^4}=0,$$

which leads to the estimator

$$\widehat{\sigma}^2=\frac{1}{n}\sum_{i=1}^{n}\left(X_i-\overline{X}\right)^2.$$

Therefore, the maximum likelihood estimators of μ and σ^2 are given, respectively, by

$$\widehat{\mu}=\overline{X}=\frac{1}{n}\sum_{i=1}^{n}X_i \quad \textit{and} \quad \widehat{\sigma}^2=\frac{1}{n}\sum_{i=1}^{n}\left(X_i-\overline{X}\right)^2.$$

In the multiparametric scenario, properties such as invariance, sufficient statistics, and others remain applicable. The same holds true for the situation involving multiple independent samples, as demonstrated in the following example.

Example 2.4.3. *Let $X_1,\ldots,X_n$ and $Y_1,\ldots,Y_m$ be random samples of $X\sim N\left(\mu_X,\sigma^2\right)$ and $Y\sim N\left(\mu_Y,\sigma^2\right)$, respectively. In this case, $\boldsymbol{\theta}=(\mu_X,\mu_Y,\sigma^2)$. Therefore, the likelihood corresponding to the observed sample is given by*

$$L(\boldsymbol{\theta};x,y)=\left(\frac{1}{\sqrt{2\pi}\sigma}\right)^n\left(\frac{1}{\sqrt{2\pi}\sigma}\right)^m\exp\left[-\frac{1}{2\sigma^2}\sum_{i=1}^{n}\left(x_i-\mu_X\right)^2-\frac{1}{2\sigma^2}\sum_{i=1}^{m}\left(y_i-\mu_Y\right)^2\right].$$

Hence,

$$l(\boldsymbol{\theta};x,y)=-\frac{(n+m)}{2}\log(2\pi)-\frac{(m+n)}{2}\log(\sigma^2)-\sum_{i=1}^{n}\frac{\left(x_i-\mu_X\right)^2}{2\sigma^2}-\sum_{i=1}^{m}\frac{\left(y_i-\mu_Y\right)^2}{2\sigma^2}.$$

By deriving $l(\boldsymbol{\theta}; x, y)$ with respect to μ_X, μ_Y and σ^2, we arrive at the equations

$$\frac{\partial l(\boldsymbol{\theta}; x, y)}{\partial \mu_X} = \sum_{i=1}^{n} (x_i - \widehat{\mu}_X) = 0,$$

$$\frac{\partial l(\boldsymbol{\theta}; x, y)}{\partial \mu_Y} = \sum_{j=1}^{m} (y_i - \widehat{\mu}_Y) = 0,$$

and

$$\frac{\partial l(\boldsymbol{\theta}; x, y)}{\partial \sigma^2} = -\frac{(m+n)}{2}\frac{1}{\widehat{\sigma}^2} + \frac{1}{2\widehat{\sigma}^4}\left[\sum_{i=1}^{n} (x_i - \widehat{\mu}_X)^2 + \sum_{j=1}^{m} (y_j - \widehat{\mu}_Y)^2\right] = 0,$$

whose solution presents the estimators

$$\widehat{\mu}_X = \overline{X}, \quad \widehat{\mu}_Y = \overline{Y},$$

and

$$\widehat{\sigma}^2 = \frac{\sum_{i=1}^{n} \left(X_i - \bar{X}\right)^2 + \sum_{j=1}^{m} \left(Y_j - \bar{Y}\right)^2}{m+n}.$$

2.4.3 Multivariate case

Let $\mathbf{X} = (\boldsymbol{X}_1, \ldots, \boldsymbol{X}_n)$ be a multivariate random sample (independently and identically distributed) of size n, where $\boldsymbol{X}_i \in \mathbb{R}^d$ is a random vector of dimension d with observed value $\mathbf{x}_i$, $i = 1, \ldots, n$. Suppose $\mathbf{X}$ has a joint density $f(\mathbf{x}; \boldsymbol{\theta})$, where $\boldsymbol{\theta}$ is the vector of unknown parameters.

The likelihood function is defined as:

$$L(\boldsymbol{\theta}; \mathbf{x}) = \prod_{i=1}^{n} f(\mathbf{x}_i; \boldsymbol{\theta}).$$

The corresponding log-likelihood function is:

$$\ell(\boldsymbol{\theta}; \mathbf{x}) = \log L(\boldsymbol{\theta}; \mathbf{x}) = \sum_{i=1}^{n} \log f(\mathbf{x}_i; \boldsymbol{\theta}).$$

This is the general format for an arbitrary multivariate distribution, where the joint density $f(\mathbf{x}; \boldsymbol{\theta})$ is replaced by the specific density of the distribution in question.

Example 2.4.4. *For a bivariate normal distribution with mean vector $\boldsymbol{\mu} = (\mu_1, \mu_2)^\top$ and covariance matrix $\boldsymbol{\Sigma}$, the joint density is given put:*

$$f(\mathbf{x}; \boldsymbol{\mu}, \boldsymbol{\Sigma}) = \frac{1}{2\pi|\boldsymbol{\Sigma}|^{1/2}} \exp\left(-\frac{1}{2}(\mathbf{x} - \boldsymbol{\mu})^\top \boldsymbol{\Sigma}^{-1}(\mathbf{x} - \boldsymbol{\mu})\right),$$

where $\mathbf{x} = (x_1, x_2)^\top$ is the vector of observations, $\boldsymbol{\mu} = (\mu_1, \mu_2)^\top$ is the vector of means and $\boldsymbol{\Sigma}$ is the covariance matrix, defined as:

$$\boldsymbol{\Sigma} = \begin{pmatrix} \sigma_1^2 & \sigma_{12} \\ \sigma_{12} & \sigma_2^2 \end{pmatrix}.$$

In the above, σ_1^2 and σ_2^2 are the variances and σ_{12} is the covariance between X_1 and X_2.

The log-likelihood function for $(\boldsymbol{\mu}, \boldsymbol{\Sigma})^\top$ is given by

$$\ell(\boldsymbol{\mu}, \boldsymbol{\Sigma}; \mathbf{x}) = -\frac{n}{2}\log(2\pi) - \frac{n}{2}\log|\boldsymbol{\Sigma}| - \frac{1}{2}\sum_{i=1}^{n}(\mathbf{x}_i - \boldsymbol{\mu})^\top \boldsymbol{\Sigma}^{-1}(\mathbf{x}_i - \boldsymbol{\mu}),$$

where the term $-\frac{n}{2}\log(2\pi)$ is constant, $|\boldsymbol{\Sigma}|$ is the determinant of the covariance matrix and $(\mathbf{x}_i - \boldsymbol{\mu})^\top \boldsymbol{\Sigma}^{-1}(\mathbf{x}_i - \boldsymbol{\mu})$ is the quadratic term that measures the Mahalanobis distance between $\mathbf{x}_i$ and $\boldsymbol{\mu}$.

To obtain parameter estimates from the maximum likelihood function in the multivariate case, we follow the standard procedure of maximizing the log-likelihood function as presented in Section 2.4.1. This process involves finding the parameters that maximize this function, deriving it with respect to the parameters, and solving the resulting system of equations.

Example 2.4.5. *To illustrate this process, consider the specific case of multivariate normal distribution with mean vector $\boldsymbol{\mu}$ and covariance matrix $\boldsymbol{\Sigma}$. The log-likelihood function for $(\boldsymbol{\mu}, \boldsymbol{\Sigma})^\top$ is written as*

$$\ell(\boldsymbol{\mu}, \boldsymbol{\Sigma}; \mathbf{x}) = -\frac{nd}{2}\log(2\pi) - \frac{n}{2}\log|\boldsymbol{\Sigma}| - \frac{1}{2}\sum_{i=1}^{n}(\mathbf{x}_i - \boldsymbol{\mu})^\top \boldsymbol{\Sigma}^{-1}(\mathbf{x}_i - \boldsymbol{\mu}).$$

Average estimate $\boldsymbol{\mu}$. Deriving the log-likelihood function with respect to $\boldsymbol{\mu}$:

$$\frac{\partial \ell(\boldsymbol{\mu}, \boldsymbol{\Sigma}; \mathbf{x})}{\partial \boldsymbol{\mu}} = \boldsymbol{\Sigma}^{-1}\sum_{i=1}^{n}(\mathbf{x}_i - \boldsymbol{\mu}).$$

Equating the derivative to zero for maximization:

$$\hat{\boldsymbol{\mu}} = \frac{1}{n}\sum_{i=1}^{n}\mathbf{x}_i.$$

The maximum likelihood estimate for the mean $\boldsymbol{\mu}$ is the sample mean $\hat{\boldsymbol{\mu}}$. Deriving the log-likelihood function with respect to $\boldsymbol{\Sigma}$:

$$\frac{\partial \ell(\boldsymbol{\mu}, \boldsymbol{\Sigma}; \mathbf{x})}{\partial \boldsymbol{\Sigma}} = -\frac{n}{2}\boldsymbol{\Sigma}^{-1} + \frac{1}{2}\boldsymbol{\Sigma}^{-1}\sum_{i=1}^{n}(\mathbf{x}_i - \boldsymbol{\mu})(\mathbf{x}_i - \boldsymbol{\mu})^{\top}\boldsymbol{\Sigma}^{-1}.$$

Equating the derivative to zero:

$$\sum_{i=1}^{n}(\mathbf{x}_i - \boldsymbol{\mu})(\mathbf{x}_i - \boldsymbol{\mu})^{\top} = n\boldsymbol{\Sigma}.$$

Solving for $\boldsymbol{\Sigma}$:

$$\hat{\boldsymbol{\Sigma}} = \frac{1}{n}\sum_{i=1}^{n}(\mathbf{x}_i - \boldsymbol{\mu})(\mathbf{x}_i - \boldsymbol{\mu})^{\top}.$$

The maximum likelihood estimate for the covariance matrix $\boldsymbol{\Sigma}$ is the sample covariance matrix $\hat{\boldsymbol{\Sigma}}$.

For multivariate normal distributions, maximum likelihood parameter estimates are

$$\hat{\boldsymbol{\mu}} = \frac{1}{n}\sum_{i=1}^{n}\mathbf{x}_i, \quad \hat{\boldsymbol{\Sigma}} = \frac{1}{n}\sum_{i=1}^{n}(\mathbf{x}_i - \hat{\boldsymbol{\mu}})(\mathbf{x}_i - \hat{\boldsymbol{\mu}})^{\top}.$$

These are the maximum likelihood estimates (MLEs) for a multivariate normal distribution. For other types of distributions, the process will be similar, but with different forms for $f(\boldsymbol{x}; \boldsymbol{\theta})$.

It is important to note that it is not always possible to estimate parameters analytically. It is often necessary to use computational resources to make this estimate, especially when applying the maximum likelihood function. In this context, the optimization problem can be converted into other well-known questions, such as the search for roots of polynomials, among others.

Chapter 3

The multivariate unit-asymmetric model

In this chapter, a new family of asymmetric distributions is presented. Initially, the model from which this family of distributions is derived is presented. This model consists of a conditional relationship between researched variables, where parameters of location, scale, asymmetry and an additional parameter that can be used to refine the model's adjustment to the data set are incorporated.

Important aspects will be discussed, such as the definition of probability density function that can be derived from the presented model. The possible configurations of these functions will also be discussed, highlighting how they can, depending on the intervals, take the form of an already known probability density function, highlighting the generalist nature of the model. Other properties will be explored, including criteria for choosing the functions to be used in the model, the special cases of probability density function, their graphical representation, the non-identifiability of the model, the marginal quantiles, among other relevant characteristics. In addition, mathematical justifications will be presented for some facts discussed throughout the text. For example, it will be demonstrated that the finiteness of moments is directly determined by the choice of functions initially defined in the model, and that a change in the configuration of the domain of these functions can result in fundamental changes in properties such as moments.

Finally, the maximum likelihood function will be presented, with the explicit characterization of this function and its respective partial derivatives, highlighting the impossibility of explicitly describing the parameter estimators in terms of the samples. As a direct consequence, parameter estimates will need to be obtained using computational methods, which will be discussed and detailed in the following chapter.

3.1 The multivariate unit-asymmetric model

Let $G_1, \ldots, G_n : (0,1) \to \mathbb{R}$, $n \in \mathbb{N}$, be monotonically and strictly increasing functions, and let $\boldsymbol{X} = (X_1, \ldots, X_n)^\top$ and Z denote a n-dimensional (absolutely) continuous random vector, with support $\mathbb{R}^n$, and a continuous random variable, respectively. Based on $G_1^{-1}, \ldots, G_n^{-1}$ (the inverse functions of $G_1, \ldots, G_n$), $\boldsymbol{X}$ and Z, we define a new random vector $\boldsymbol{Y} = (Y_1, \ldots, Y_n)^\top$, with support $(0,1)^n$ (the unit hypercube), as follows

$$\boldsymbol{Y} = \boldsymbol{T} \,|\, \boldsymbol{\lambda}^\top(\boldsymbol{X} - \boldsymbol{\mu}) + \tau > Z, \tag{3.1.1}$$

where $\boldsymbol{T} = (G_1^{-1}(X_1), \ldots, G_n^{-1}(X_n))^\top$, $\tau \in \mathbb{R}$ is the extension parameter, $\boldsymbol{\lambda} = (\lambda_1, \ldots, \lambda_n)^\top \in \mathbb{R}^n$ is the skewness parameter vector and $\boldsymbol{\mu} = (\mu_1, \ldots, \mu_n)^\top \in \mathbb{R}^n$ is a constant vector. That is, $\boldsymbol{Y}$ is the conditional random vector for $\boldsymbol{T}$ given $\boldsymbol{\lambda}^\top(\boldsymbol{X} - \boldsymbol{\mu}) + \tau > Z$. Let $f_{\boldsymbol{Y}}$ denote the joint probability density function (PDF) of $\boldsymbol{Y}$. Bayes' rule provides

$$\begin{aligned}
f_{\boldsymbol{Y}}(\boldsymbol{y}) &= \frac{\displaystyle\int_0^\infty f_{\boldsymbol{T}, \boldsymbol{\lambda}^\top(\boldsymbol{X}-\boldsymbol{\mu}) - Z + \tau}(\boldsymbol{y}, s)\,\mathrm{d}s}{\mathbb{P}(\boldsymbol{\lambda}^\top(\boldsymbol{X} - \boldsymbol{\mu}) + \tau > Z)}, \qquad \boldsymbol{y} = (y_1, \ldots, y_n)^\top \in (0,1)^n, \\
&= f_{\boldsymbol{T}}(\boldsymbol{y}) \, \frac{\displaystyle\int_0^\infty f_{\boldsymbol{\lambda}^\top(\boldsymbol{X}-\boldsymbol{\mu}) - Z + \tau \,|\, \boldsymbol{T} = \boldsymbol{y}}(s)\,\mathrm{d}s}{\mathbb{P}(Z - \boldsymbol{\lambda}^\top(\boldsymbol{X} - \boldsymbol{\mu}) < \tau)} \\
&= f_{\boldsymbol{T}}(\boldsymbol{y}) \, \frac{F_Z(\boldsymbol{\lambda}^\top(\boldsymbol{y}_G - \boldsymbol{\mu}) + \tau \,|\, \boldsymbol{X} = \boldsymbol{y}_G)}{F_{Z - \boldsymbol{\lambda}^\top(\boldsymbol{X} - \boldsymbol{\mu})}(\tau)}, \quad \boldsymbol{y}_G \equiv (G_1(y_1), \ldots, G_n(y_n))^\top \in \mathbb{R}^n.
\end{aligned} \tag{3.1.2}$$

Chain rule gives $f_{\boldsymbol{T}}(\boldsymbol{y}) = f_{\boldsymbol{X}}(\boldsymbol{y}_G) \prod_{i=1}^n G_i'(y_i)$. So, from (3.1.2) we have

$$f_{\boldsymbol{Y}}(\boldsymbol{y}) = f_{\boldsymbol{X}}(\boldsymbol{y}_G) \, \frac{F_Z(\boldsymbol{\lambda}^\top(\boldsymbol{y}_G - \boldsymbol{\mu}) + \tau \,|\, \boldsymbol{X} = \boldsymbol{y}_G)}{F_{Z - \boldsymbol{\lambda}^\top(\boldsymbol{X} - \boldsymbol{\mu})}(\tau)} \prod_{i=1}^n G_i'(y_i), \quad \boldsymbol{y} \in (0,1)^n, \tag{3.1.3}$$

where $\boldsymbol{y}_G$ is as given in (3.1.2).

Table 3.1 presents some examples of functions G_i's for use in (3.1.3).

Table 3.1: Some functions G_i's with its respective inverses and derivatives.

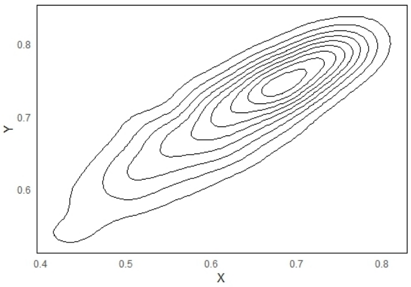
$G_i(x)$	$G_i^{-1}(x)$	$G_i'(x)$
$\tan((x-\frac{1}{2})\pi)$	$\frac{1}{2}+\frac{\arctan(x)}{\pi}$	$\frac{\pi}{\sin^2(\pi x)}$
$\log(\frac{x^3}{1-x^3})$	$\left[\frac{\exp(x)}{1+\exp(x)}\right]^{\frac{1}{3}}$	$\frac{3}{x(1-x^3)}$
$\log(\frac{x^5}{1-x^5})$	$\left[\frac{\exp(x)}{1+\exp(x)}\right]^{\frac{1}{5}}$	$\frac{5}{x(1-x^5)}$
$\log(-\log(1-x))$	$1-\exp(-\exp(x))$	$\frac{1}{(1-x)\log(\frac{1}{1-x})}$

Determining functions with the initially presented characteristics can follow a specific rule. One suggestion is to consider the inverse functions as cumulative distribution functions. It is important to note that the range of these functions is restricted to the interval $(0,1)$. By imposing constraints on G_i, we can achieve the desired domains. Consider the following examples presented in Table 3.2.

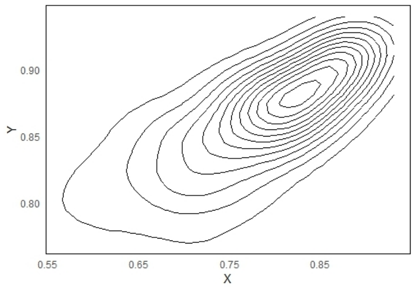
Table 3.2: Some functions G_i's and their inverses are obtained from cumulative distribution functions.

Distribution	$G_i(x)$	$G_i^{-1}(x)$	$G_i'(x)$
Hyperbolic secant distribution	$\frac{2}{\pi}\log(\tan\frac{\pi}{2}x))$	$\frac{2}{\pi}\arctan(\exp(\frac{\pi}{2}x))$	$\frac{\sec^2(\frac{\pi}{2}x)}{\tan(\frac{\pi}{2}x)}$
Gumbel ($\mu=\beta=1$)	$1-\log(-\log(x))$	$\exp(-\exp(-x+1))$	$\frac{-1}{x\log(x)}$
Gompertz ($\eta=\beta=1$)	$\log(\log(\frac{1}{-x+1})+1)$	$1-\exp(-\exp(x)+1)$	$\frac{(-x+1)^{-1}}{\log(\frac{1}{-x+1})+1}$
Logistic ($k=L=1, x_0=0$)	$\log(\frac{x}{1-x})$	$\frac{\exp(x)}{1+\exp(x)}$	$\frac{1}{x(1-x)}$

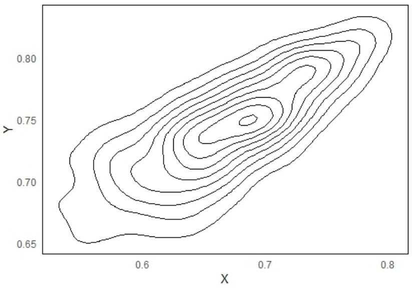
In what follows we present some two-dimensional illustrations of the outputs from the model with some variations in the parameters values and functions, where the distribution is bivariate normal. In the first column $G_i^{-1}(x)=(x-2)/(2x)+\sqrt{x^2+4}/(2x)$ and the second column $G_i^{-1}(x)=(1/2)+\arctan(x)/\pi$. Note that the graphs appear to exhibit an elliptical configuration with a certain degree of asymmetry. This data distribution model will be fundamental for the development of this work. In Chapter 2, we used the notation EC_n a for symmetric elliptic distributions. From now on, we will use the notation ELL_n for these distributions.



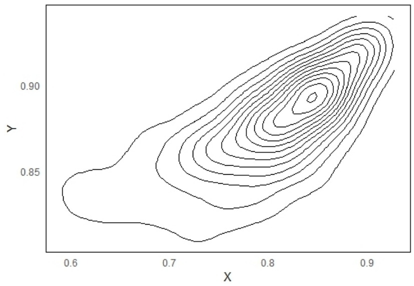
$\boldsymbol{\mu} = (1, 2)^T; \boldsymbol{\lambda} = (1, -0.7)^T; \tau = 0.5.$



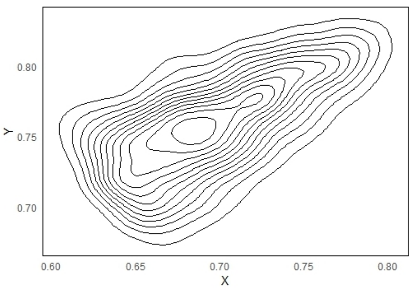
$\boldsymbol{\mu} = (0.8, 2)^T; \boldsymbol{\lambda} = (1.5, 1)^T; \tau = 0.7.$



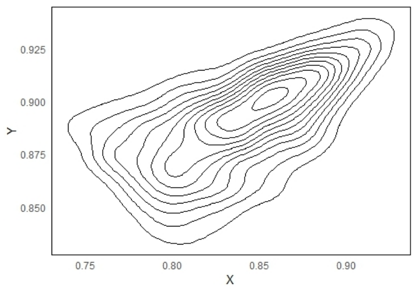
$\boldsymbol{\mu} = (1, 2)^T; \boldsymbol{\lambda} = (1, 2)^T; \tau = 1.$



$\boldsymbol{\mu} = (2.5, 1)^T; \boldsymbol{\lambda} = (1, 0.7)^T; \tau = 1.2.$



$\boldsymbol{\mu} = (1, 2)^T; \boldsymbol{\lambda} = (50, 20)^T; \tau = -10.$



$\boldsymbol{\mu} = (1.5, 1.8)^T; \boldsymbol{\lambda} = (40, 30)^T; \tau = -8.$

Figure 3.1: Scatterplots for data with elliptical distribution.

So far we have not established any probabilistic dependency relationship between Z and $\boldsymbol{X}$. From now on we assume that the $(n+1)$-dimensional vector $\boldsymbol{X}^*$, defined as $\boldsymbol{X}^* = (Z, \boldsymbol{X})^\top$, has a multivariate elliptical (symmetric) (ELL_{n+1}) distribution (Fang, Kotz, and Ng, 1990) with location vector $\boldsymbol{\mu}^* = (0, \boldsymbol{\mu})^\top$, for $\boldsymbol{\mu} = (\mu_1, \ldots, \mu_n)^\top \in \mathbb{R}^n$, positive definite $(n+1) \times (n+1)$

dispersion matrix

$$\boldsymbol{\Sigma}^* = \begin{pmatrix} 1 & \mathbf{0}^\top \\ \mathbf{0} & \boldsymbol{\Sigma} \end{pmatrix}, \quad \boldsymbol{\Sigma} = (\Sigma_{i,j})_{n\times n},\ \Sigma_{i,j} = \mathrm{Cov}(X_i, X_j),\ i,j = 1,\ldots,n,$$

and density function generator $g^{(n+1)}$. For simplicity we use the notation $\boldsymbol{X}^* \sim \mathrm{ELL}_{n+1}(\boldsymbol{\mu}^*, \boldsymbol{\Sigma}^*, g^{(n+1)})$. The PDF of $\boldsymbol{X}^* \sim \mathrm{ELL}_{n+1}(\boldsymbol{\mu}^*, \boldsymbol{\Sigma}^*, g^{(n+1)})$ at $\boldsymbol{x}^* = (x_1,\ldots,x_{n+1})^\top \in \mathbb{R}^{n+1}$ is given by

$$\begin{aligned} f_{\boldsymbol{X}^*}(\boldsymbol{x}^*) &= f_{\boldsymbol{X}^*}(\boldsymbol{x}^*; \boldsymbol{\mu}^*, \boldsymbol{\Sigma}^*, g^{(n+1)}) \\ &= \frac{1}{|\boldsymbol{\Sigma}^*|^{1/2} Z_{g^{(n+1)}}}\, g^{(n+1)}((\boldsymbol{x}^* - \boldsymbol{\mu}^*)^\top [\boldsymbol{\Sigma}^*]^{-1} (\boldsymbol{x}^* - \boldsymbol{\mu}^*)), \end{aligned} \tag{3.1.4}$$

where

$$Z_{g^{(n+1)}} = \frac{\pi^{(n+1)/2}}{\Gamma((n+1)/2)} \int_0^\infty u^{(n+1)/2-1} g^{(n+1)}(u)\mathrm{d}u$$

is a normalization constant.

Table 3.3 presents some examples of densities generators for use in (3.1.4).

Table 3.3: Normalization functions $(Z_{g^{(n)}})$ and density generators $(g^{(n)})$.

Multivariate distribution	$Z_{g^{(n)}}$	$g^{(n)}(x)$	Parameter
Extended unit-G-skew-Student-t	$\frac{\Gamma(\nu/2)(\nu\pi)^{n/2}}{\Gamma((\nu+n)/2)}$	$(1+\frac{x}{\nu})^{-(\nu+n)/2}$	$\nu > 0$
Extended unit-G-skew-Cauchy	$\frac{\pi^{(n+1)/2}}{\Gamma((n+1)/2)}$	$\frac{1}{(1+x)^{(n+1)/2}}$	$-$
Extended unit-G-skew-normal	$(2\pi)^{n/2}$	$\exp(-x/2)$	$-$

It is well-known that all elliptic distributions are invariant to linear transformations (see Fang, Kotz, and Ng, 1990), that is, if $\boldsymbol{S} \sim \mathrm{ELL}_n(\boldsymbol{\mu}, \boldsymbol{\Sigma}, g^{(n)})$ then $\boldsymbol{c} + \boldsymbol{A}\boldsymbol{S} \sim \mathrm{ELL}_n(\boldsymbol{c} + \boldsymbol{A}\boldsymbol{\mu}, \boldsymbol{A}\boldsymbol{\Sigma}\boldsymbol{A}^\top, g^{(n)})$, where $\boldsymbol{A}$ is a square matrix and $\boldsymbol{c} \in \mathbb{R}^n$ is a constant vector. In particular, this implies that a linear combination of the components of $\boldsymbol{X}$ is again elliptically distributed. More precisely, we have

$$Z - \boldsymbol{\lambda}^\top(\boldsymbol{X} - \boldsymbol{\mu}) \sim \mathrm{ELL}_1\left(0, 1 + \boldsymbol{\lambda}^\top \boldsymbol{\Sigma} \boldsymbol{\lambda}, g^{(1)}\right). \tag{3.1.5}$$

As a consequence of the last statement, we have that marginals of an elliptic distribution are elliptic. Hence,

$$\boldsymbol{X} \sim \mathrm{ELL}_n(\boldsymbol{\mu}, \boldsymbol{\Sigma}, g^{(n)}). \tag{3.1.6}$$

On the other hand, it is well-known that conditionals of an elliptic distribution are again elliptic (see Theorem 2.18 of Fang, Kotz, and Ng, 1990). This provides that

$$Z|\boldsymbol{X} = \boldsymbol{x} \sim \mathrm{ELL}_1(0, 1, g_{q(\boldsymbol{x})}), \tag{3.1.7}$$

where

$$q(\boldsymbol{x}) = (\boldsymbol{x} - \boldsymbol{\mu})^\top \boldsymbol{\Sigma}^{-1}(\boldsymbol{x} - \boldsymbol{\mu}) \quad \text{and} \quad g_{q(\boldsymbol{x})}(s) = \frac{g^{(2)}(s + q(\boldsymbol{x}))}{g^{(1)}(q(\boldsymbol{x}))}. \tag{3.1.8}$$

Let $F_{\mathrm{ELL}_1}(\cdot;\, 0, 1, g)$ be the CDF of $\mathrm{ELL}_1(0, 1, g)$ with density generator function g, where g can be $g_{q(\boldsymbol{x})}$ or $g^{(1)}$. So, from (3.1.5), (3.1.6) and (3.1.7), the identity in (3.1.3) can be written as

$$f_{\boldsymbol{Y}}(\boldsymbol{y}) = f_{\boldsymbol{X}}(\boldsymbol{y}_G)\, \frac{F_{\mathrm{ELL}_1}(\boldsymbol{\lambda}^\top(\boldsymbol{y}_G - \boldsymbol{\mu}) + \tau;\, 0, 1, g_{q(\boldsymbol{y}_G)})}{F_{\mathrm{ELL}_1}(\tau;\, 0, 1 + \boldsymbol{\lambda}^\top\boldsymbol{\Sigma}\boldsymbol{\lambda}, g^{(1)})} \prod_{i=1}^{n} G_i'(y_i), \quad \boldsymbol{y} \in (0,1)^n,$$

with $\boldsymbol{y}_G$ being as in (3.1.2) and $\boldsymbol{X} \sim \mathrm{ELL}_n(\boldsymbol{\mu}, \boldsymbol{\Sigma}, g^{(n)})$.

Note that $F_{\mathrm{ELL}_1}(\tau = 0;\, 0, 1 + \boldsymbol{\lambda}^\top\boldsymbol{\Sigma}\boldsymbol{\lambda}, g^{(1)}) = 1/2$ because $Z - \boldsymbol{\lambda}^\top(\boldsymbol{X} - \boldsymbol{\mu})$ is symmetric about 0.

Definition 3.1.1. *We say that a random vector $\boldsymbol{Y} = (Y_1, \ldots, Y_n)^\top$ has a multivariate extended unit-G-skew-elliptical (EUGSE$_n$) distribution if $\boldsymbol{Y}$ has the probability density function*

$$f_{\boldsymbol{Y}}(\boldsymbol{y}) = f_{\boldsymbol{X}}(\boldsymbol{y}_G)\, \frac{F_{\mathrm{ELL}_1}(\boldsymbol{\lambda}^\top(\boldsymbol{y}_G - \boldsymbol{\mu}) + \tau;\, 0, 1, g_{q(\boldsymbol{y}_G)})}{F_{\mathrm{ELL}_1}(\tau;\, 0, 1 + \boldsymbol{\lambda}^\top\boldsymbol{\Sigma}\boldsymbol{\lambda}, g^{(1)})} \prod_{i=1}^{n} G_i'(y_i), \quad \boldsymbol{y} \in (0,1)^n, \tag{3.1.9}$$

where $\boldsymbol{X} \sim$ ELL$_n(\boldsymbol{\mu}, \boldsymbol{\Sigma}, g^{(n)})$. For simplicity, we write $\boldsymbol{Y} \sim$ EUGSE$_n(\boldsymbol{\mu}, \boldsymbol{\Sigma}, \boldsymbol{\lambda}, \tau, g^{(n)})$ and we commonly say that $\boldsymbol{Y}$ is an EUGSE$_n$ random vector.

Explicit formulas for the PDF of $\boldsymbol{Y} \sim$ EUGSE$_n(\boldsymbol{\mu}, \boldsymbol{\Sigma}, \boldsymbol{\lambda}, \tau, g^{(n)})$ corresponding to multivariate extended unit-G-skew-Student-t, multivariate extended unit-G-skew-Cauchy and multivariate extended unit-G-skew-normal models (see Table 3.3), are provided in Subsection 3.2.1.

The EUGSE$_n$ distribution provides a very flexible class of statistical models. Depending on the choice of the functions $G_1, \ldots, G_n$ we have a family of multivariate extended unit distributions with presence of asymmetry. For $\boldsymbol{\lambda} = \boldsymbol{0}$, $\tau = 0$, $G_1(x) = G_2(x) = \log(-\log(1 - x))$,

$0 < x < 1$, and $n = 2$, we obtain the bivariate unit model studied in reference (Vila et al., 2023b). In general, for the EUGSE$_n$ model, it is not necessary to consider all G_i's equal as in (Vila et al., 2023b). For $g^{(n)}(x) = (1 + x/\nu)^{-(\nu+n)/2}$, $\nu > 0$, we get the multivariate extended unit-G-skew-Student-t, which reduces to the multivariate extended G-skew-Cauchy and multivariate extended G-skew-normal distributions by letting $\nu = 1$ and $\nu \to \infty$, respectively.

3.2 Some structural properties

3.2.1 Special cases

In this subsection we develop some examples of multivariate EUGSE$_n$ distributions as special cases.

Proposition 3.2.1 (Multivariate extended unit-G-skew-Student-t)**.** *Let $\boldsymbol{Y} \sim$ EUGSE$_n(\boldsymbol{\mu}, \boldsymbol{\Sigma}, \boldsymbol{\lambda}, \tau, g^{(n)})$, where $g^{(n)}(x) = (1 + x/\nu)^{-(\nu+n)/2}$, $x \in \mathbb{R}$, is the PDF generator of the multivariate Student-t distribution with $\nu > 0$ degrees of freedom. Then, the PDF of $\boldsymbol{Y}$ at $\boldsymbol{y} \in (0,1)^n$ is given by*

$$f_{\boldsymbol{Y}}(\boldsymbol{y}) = t_n(\boldsymbol{y}_G;\, \boldsymbol{\mu}, \boldsymbol{\Sigma}, \nu)\, \frac{F_{\nu+1}\left([\boldsymbol{\lambda}^\top(\boldsymbol{y}_G - \boldsymbol{\mu}) + \tau]\sqrt{\frac{\nu+1}{\nu+q(\boldsymbol{y}_G)}}\right)}{F_\nu\left(\frac{\tau}{\sqrt{1+\boldsymbol{\lambda}^\top\boldsymbol{\Sigma}\boldsymbol{\lambda}}}\right)} \prod_{i=1}^n G_i'(y_i), \tag{3.2.1}$$

where $\boldsymbol{y}_G$ and $q(\boldsymbol{y}_G)$ are as given in (3.1.2) and (3.1.8), respectively. Moreover, $t_n(\boldsymbol{y}_G;\, \boldsymbol{\mu}, \boldsymbol{\Sigma}, \nu) = g^{(n)}(q(\boldsymbol{y}_G))/(|\boldsymbol{\Sigma}|^{1/2} Z_{g^{(n)}})$, with $Z_{g^{(n)}}$ being as in Table 3.3, denotes the PDF of the usual n-dimensional Student-t distribution with location $\boldsymbol{\mu} \in \mathbb{R}^n$, positive definite $n \times n$ dispersion matrix $\boldsymbol{\Sigma}$, and degrees of freedom $\nu > 0$, and F_ν denotes the univariate standard Student-t CDF with degrees of freedom $\nu > 0$.

Proof. By using formula (3.1.9), it is enough to verify that

$$F_{\mathrm{ELL}_1}(\boldsymbol{\lambda}^\top(\boldsymbol{y}_G - \boldsymbol{\mu}) + \tau;\, 0, 1, g_{q(\boldsymbol{y}_G)}) = F_{\nu+1}\left([\boldsymbol{\lambda}^\top(\boldsymbol{y}_G - \boldsymbol{\mu}) + \tau]\sqrt{\frac{\nu+1}{\nu + q(\boldsymbol{y}_G)}}\right) \tag{3.2.2}$$

and

$$F_{\mathrm{ELL}_1}(\tau;\, 0, 1 + \boldsymbol{\lambda}^\top\boldsymbol{\Sigma}\boldsymbol{\lambda}, g^{(1)}) = F_\nu\left(\frac{\tau}{\sqrt{1 + \boldsymbol{\lambda}^\top\boldsymbol{\Sigma}\boldsymbol{\lambda}}}\right). \tag{3.2.3}$$

The identity (3.2.3) follows directly when standardizing the corresponding random variable of $F_{\mathrm{ELL}_1}(\cdot;\, 0, 1 + \boldsymbol{\lambda}^\top\boldsymbol{\Sigma}\boldsymbol{\lambda}, g^{(1)})$. Therefore, it remains to verify (3.2.2). Indeed, as

$F_{\mathrm{ELL}_1}(\cdot;\, 0, 1, g_{q(\boldsymbol{y}_G)})$ is the CDF of $\mathrm{ELL}_1(0, 1, g_{q(\boldsymbol{y}_G)})$ with density generator function $g_{q(\boldsymbol{y}_G)}$ as given in (3.1.8), we have

$$F_{\mathrm{ELL}_1}(\boldsymbol{\lambda}^\top(\boldsymbol{y}_G - \boldsymbol{\mu}) + \tau;\, 0, 1, g_{q(\boldsymbol{y}_G)}) = \frac{1}{Z_{g^{(2)}}/Z_{g^{(1)}}} \int_{-\infty}^{\boldsymbol{\lambda}^\top(\boldsymbol{y}_G-\boldsymbol{\mu})+\tau} \frac{g^{(2)}(s^2 + q(\boldsymbol{y}_G))}{g^{(1)}(q(\boldsymbol{y}_G))} \mathrm{d}s, \tag{2.2.4}$$

which, by simple algebraic manipulations, can be written as

$$= \frac{1}{Z_{g^{(2)}}/Z_{g^{(1)}}} \int_{-\infty}^{\boldsymbol{\lambda}^\top(\boldsymbol{y}_G-\boldsymbol{\mu})+\tau} \frac{(1 + \frac{s^2+q(\boldsymbol{y}_G)}{\nu})^{-(\nu+2)/2}}{(1 + \frac{q(\boldsymbol{y}_G)}{\nu})^{-(\nu+1)/2}} \mathrm{d}s$$

$$= \frac{1}{Z_{g^{(2)}}/Z_{g^{(1)}}} \int_{-\infty}^{\boldsymbol{\lambda}^\top(\boldsymbol{y}_G-\boldsymbol{\mu})+\tau} \frac{\left(1 + \frac{1}{\nu+1}\left[s\sqrt{\frac{\nu+1}{\nu+q(\boldsymbol{y}_G)}}\right]^2\right)^{-(\nu+2)/2}}{\sqrt{1 + \frac{q(\boldsymbol{y}_G)}{\nu}}} \mathrm{d}s.$$

By making the change of variable $t = s\sqrt{(\nu+1)/(\nu+q(\boldsymbol{y}_G))}$, the last integral is

$$= \frac{1}{Z_{g^{(2)}}/Z_{g^{(1)}}} \sqrt{\frac{\nu}{\nu+1}} \int_{-\infty}^{(\boldsymbol{\lambda}^\top(\boldsymbol{y}_G-\boldsymbol{\mu})+\tau)\sqrt{\frac{\nu+1}{\nu+q(\boldsymbol{y}_G)}}} \left(1 + \frac{t^2}{\nu+1}\right)^{-(\nu+2)/2} \mathrm{d}t. \tag{3.2.4}$$

A simple observation shows that

$$\frac{1}{Z_{g^{(2)}}/Z_{g^{(1)}}} \sqrt{\frac{\nu}{\nu+1}} = \left[\frac{((\nu+1)\pi)^{1/2}\Gamma((\nu+1)/2)}{\Gamma((\nu+2)/2)}\right]^{-1}.$$

So, the integral in (3.2.4) is written as

$$= \left[\frac{((\nu+1)\pi)^{1/2}\Gamma((\nu+1)/2)}{\Gamma((\nu+2)/2)}\right]^{-1} \int_{-\infty}^{(\boldsymbol{\lambda}^\top(\boldsymbol{y}_G-\boldsymbol{\mu})+\tau)\sqrt{\frac{\nu+1}{\nu+q(\boldsymbol{y}_G)}}} \left(1 + \frac{t^2}{\nu+1}\right)^{-(\nu+2)/2} \mathrm{d}t$$

$$= F_{\nu+1}\left([\boldsymbol{\lambda}^\top(\boldsymbol{y}_G - \boldsymbol{\mu}) + \tau]\sqrt{\frac{\nu+1}{\nu+q(\boldsymbol{y}_G)}}\right).$$

Then, the required formula in (3.2.2) follows. □

By letting $\nu = 1$ in Proposition 3.2.1, we have the following result.

Proposition 3.2.2 (Multivariate extended unit-G-skew-Cauchy)**.** *Let* $\boldsymbol{Y} \sim$ *EUGSE*$_n(\boldsymbol{\mu}, \boldsymbol{\Sigma}, \boldsymbol{\lambda}, \tau, g^{(n)})$*, where* $g^{(n)}(x) = 1/(1+x)^{(n+1)/2}$*,* $x \in \mathbb{R}$*, is the PDF gener-*

ator of the multivariate Cauchy distribution. Then, the PDF of $\boldsymbol{Y}$ at $\boldsymbol{y} \in (0,1)^n$ is given by

$$f_{\boldsymbol{Y}}(\boldsymbol{y}) = c_n(\boldsymbol{y}_G;\, \boldsymbol{\mu}, \boldsymbol{\Sigma})\, \frac{F_2\left([\boldsymbol{\lambda}^\top(\boldsymbol{y}_G - \boldsymbol{\mu}) + \tau]\sqrt{\frac{2}{1+q(\boldsymbol{y}_G)}}\right)}{F_1\left(\frac{\tau}{\sqrt{1+\boldsymbol{\lambda}^\top\boldsymbol{\Sigma}\boldsymbol{\lambda}}}\right)} \prod_{i=1}^n G_i'(y_i), \tag{3.2.5}$$

where $\boldsymbol{y}_G$ and $q(\boldsymbol{y}_G)$ are as given in (3.1.2) *and* (3.1.8)*, respectively. Moreover, $c_n(\boldsymbol{y}_G;\, \boldsymbol{\mu}, \boldsymbol{\Sigma}) = g^{(n)}(q(\boldsymbol{y}_G))/(|\boldsymbol{\Sigma}|^{1/2} Z_{g^{(n)}})$, with $Z_{g^{(n)}}$ being as in Table 3.3, denotes the PDF of the usual n-dimensional Cauchy distribution with location $\boldsymbol{\mu} \in \mathbb{R}^n$ and positive definite $n \times n$ dispersion matrix $\boldsymbol{\Sigma}$, and F_ν denotes the univariate standard Student-t CDF with degrees of freedom $\nu \in \{1,2\}$.*

By letting $\nu \to \infty$ in Proposition 3.2.1, the following result follows.

Proposition 3.2.3 (Multivariate extended unit-G-skew-normal)**.** *Let $\boldsymbol{Y} \sim$ EUGSE$_n(\boldsymbol{\mu}, \boldsymbol{\Sigma}, \boldsymbol{\lambda}, \tau, g^{(n)})$, where $g^{(n)}(x) = \exp(-x/2)$, $x \in \mathbb{R}$, is the PDF generator of the multivariate Gaussian distribution. Then, the PDF of $\boldsymbol{Y}$ at $\boldsymbol{y} \in (0,1)^n$ is given by*

$$f_{\boldsymbol{Y}}(\boldsymbol{y}) = \phi_n(\boldsymbol{y}_G;\, \boldsymbol{\mu}, \boldsymbol{\Sigma})\, \frac{\Phi\left(\boldsymbol{\lambda}^\top(\boldsymbol{y}_G - \boldsymbol{\mu}) + \tau\right)}{\Phi\left(\frac{\tau}{\sqrt{1+\boldsymbol{\lambda}^\top\boldsymbol{\Sigma}\boldsymbol{\lambda}}}\right)} \prod_{i=1}^n G_i'(y_i), \tag{3.2.6}$$

where $\boldsymbol{y}_G$ is as given in (3.1.2)*. Here, $\phi_n(\boldsymbol{y}_G;\, \boldsymbol{\mu}, \boldsymbol{\Sigma}, \nu) = g^{(n)}((\boldsymbol{y}_G - \boldsymbol{\mu})^\top \boldsymbol{\Sigma}^{-1}(\boldsymbol{y}_G - \boldsymbol{\mu}))/(|\boldsymbol{\Sigma}|^{1/2} Z_{g^{(n)}})$, with $Z_{g^{(n)}}$ being as in Table 3.3, denotes the PDF of the usual n-dimensional Gaussian distribution with location $\boldsymbol{\mu} \in \mathbb{R}^n$ and positive definite $n \times n$ dispersion matrix $\boldsymbol{\Sigma}$, and Φ denotes the univariate standard Gaussian CDF.*

Table 3.4 summarizes the results found in Propositions 3.2.1, 3.2.2 and 3.2.3.

Table 3.4: Probability density functions $f_{\boldsymbol{Y}}$ of the EUGSE$_n$ distributions of Table 3.3.

Multivariate distribution	$f_{\boldsymbol{Y}}(\boldsymbol{y})$
Extended unit-G-skew-Student-t	$t_n(\boldsymbol{y}_G;\, \boldsymbol{\mu}, \boldsymbol{\Sigma}, \nu)\, \frac{F_{\nu+1}\left([\boldsymbol{\lambda}^\top(\boldsymbol{y}_G - \boldsymbol{\mu}) + \tau]\sqrt{\frac{\nu+1}{\nu+q(\boldsymbol{y}_G)}}\right)}{F_\nu\left(\frac{\tau}{\sqrt{1+\boldsymbol{\lambda}^\top\boldsymbol{\Sigma}\boldsymbol{\lambda}}}\right)} \prod_{i=1}^n G_i'(y_i)$
Extended unit-G-skew-Cauchy	$c_n(\boldsymbol{y}_G;\, \boldsymbol{\mu}, \boldsymbol{\Sigma})\, \frac{F_2\left([\boldsymbol{\lambda}^\top(\boldsymbol{y}_G - \boldsymbol{\mu}) + \tau]\sqrt{\frac{2}{1+q(\boldsymbol{y}_G)}}\right)}{F_1\left(\frac{\tau}{\sqrt{1+\boldsymbol{\lambda}^\top\boldsymbol{\Sigma}\boldsymbol{\lambda}}}\right)} \prod_{i=1}^n G_i'(y_i)$
Extended unit-G-skew-normal	$\phi_n(\boldsymbol{y}_G;\, \boldsymbol{\mu}, \boldsymbol{\Sigma})\, \frac{\Phi\left(\boldsymbol{\lambda}^\top(\boldsymbol{y}_G - \boldsymbol{\mu}) + \tau\right)}{\Phi\left(\frac{\tau}{\sqrt{1+\boldsymbol{\lambda}^\top\boldsymbol{\Sigma}\boldsymbol{\lambda}}}\right)} \prod_{i=1}^n G_i'(y_i)$

Below, in Figure 3.2, are some graphs that represent the Extended unit-G-skew-normal distribution where the transformation $G^{-1}(x) = (1/2) + \arctan(x)/\pi$ where the parameters used were: $\boldsymbol{\mu} = (2,3)^{\top}$; $\boldsymbol{\lambda} = (0.5, 0.6)^{\top}$; $\sigma_1 = 1$; $\sigma_2 = 1$; $\rho = 0.5$; $\tau = 0$.

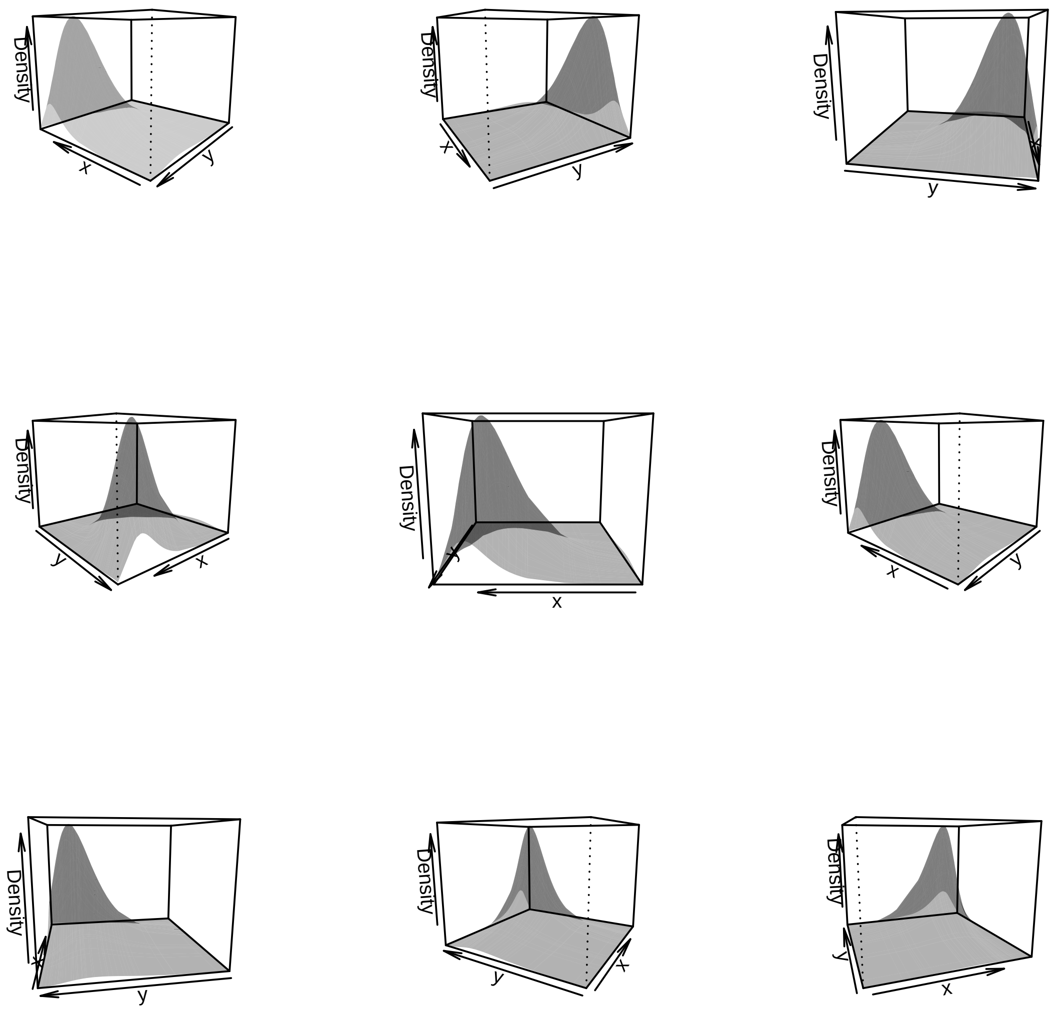


Figure 3.2: Extended unit-G-skew-normal density function with $G^{-1}(x) = (1/2) + \arctan(x)/\pi$.

Based on the presented graphs, a clear asymmetry in the data distribution can be observed. Additionally, it is worth noting that the graphical representation was constructed using 45 samples generated through computational simulations.

3.2.2 Stochastic representation

Let $\boldsymbol{X} = (X_1, \ldots, X_n)^\top$ be a n-dimensional random vector and Z be a real-valued random variable. Assume that the $(n+1)$-dimensional vector $(Z, \boldsymbol{X})^\top$ has a multivariate elliptical (symmetric) (ELL_{n+1}) distribution (Fang, Kotz, and Ng, 1990) with location vector $(0, \boldsymbol{\mu})^\top$, positive definite $(n+1)\times(n+1)$ dispersion matrix

$$\begin{pmatrix} 1 & 0 \\ \boldsymbol{0} & \boldsymbol{\Sigma} \end{pmatrix}, \quad \boldsymbol{\Sigma} = (\Sigma_{i,j})_{n\times n},\ \Sigma_{i,j} = \mathrm{Cov}(X_i, X_j),\ i,j = 1, \ldots, n,$$

and density generator $g^{(n+1)}$. For simplicity, we write

$$\begin{pmatrix} Z \\ \boldsymbol{X} \end{pmatrix} \sim \mathrm{ELL}_{n+1}\left(\begin{pmatrix} 0 \\ \boldsymbol{\mu} \end{pmatrix}, \begin{pmatrix} 1 & 0 \\ \boldsymbol{0} & \boldsymbol{\Sigma} \end{pmatrix}, g^{(n+1)}\right).$$

Well-known results by Fang, Kotz, and Ng (1990) on marginals and conditionals of multivariate elliptic distributions provide the following statements:

$$Z - \boldsymbol{\lambda}^\top(\boldsymbol{X} - \boldsymbol{\mu}) \sim \mathrm{ELL}_1\left(0, 1 + \boldsymbol{\lambda}^\top\boldsymbol{\Sigma}\boldsymbol{\lambda}, g^{(1)}\right), \tag{3.2.7}$$

$$\boldsymbol{X} \sim \mathrm{ELL}_n(\boldsymbol{\mu}, \boldsymbol{\Sigma}, g^{(n)}), \tag{3.2.8}$$

$$Z \sim \mathrm{ELL}_1(0, 1, g^{(1)}), \tag{3.2.9}$$

$$Z \,|\, \boldsymbol{X} = \boldsymbol{x} \sim \mathrm{ELL}_1(0, 1, g_{q(\boldsymbol{x})}), \quad \boldsymbol{x} = (x_1, \ldots, x_n)^\top \in \mathbb{R}^n, \tag{3.2.10}$$

where $q(\boldsymbol{x})$ and $g_{q(\boldsymbol{x})}$ are as in (3.1.8). From (3.1.6), $\boldsymbol{X}$ is multivariate elliptic, then its corresponding PDF is

$$f_{\boldsymbol{X}}(\boldsymbol{x}) = \frac{1}{|\boldsymbol{\Sigma}|^{1/2} Z_{g^{(n)}}}\, g^{(n)}(q(\boldsymbol{x})), \quad \boldsymbol{x} \in \mathbb{R}^n.$$

Setting $\boldsymbol{T} = (G_1^{-1}(X_1), \ldots, G_n^{-1}(X_n))^\top$, by chain rule it is clear that

$$f_{\boldsymbol{T}}(\boldsymbol{y}) = f_{\boldsymbol{X}}(\boldsymbol{y}_G) \prod_{i=1}^n G_i'(y_i).$$

Hence, from (3.2.7), (3.2.8) and (3.2.10), the PDF (3.1.9) of $\boldsymbol{Y} \sim \text{EUGSE}_n(\boldsymbol{\mu}, \boldsymbol{\Sigma}, \boldsymbol{\lambda}, \tau, g^{(n)})$ is written as

$$f_{\boldsymbol{Y}}(\boldsymbol{y}) = f_{\boldsymbol{T}}(\boldsymbol{y})\, \frac{F_Z(\boldsymbol{\lambda}^\top(\boldsymbol{y}_G - \boldsymbol{\mu}) + \tau \,|\, \boldsymbol{X} = \boldsymbol{y}_G)}{F_{Z-\boldsymbol{\lambda}^\top(\boldsymbol{X}-\boldsymbol{\mu})}(\tau)}, \quad \boldsymbol{y} = (y_1, \ldots, y_n)^\top \in (0,1)^n.$$

By using the above expression of $f_{\boldsymbol{Y}}(\boldsymbol{y})$ and then Bayes' rule, we get

$$\begin{aligned} f_{\boldsymbol{Y}}(\boldsymbol{y}) &= f_{\boldsymbol{T}}(\boldsymbol{y})\, \frac{\displaystyle\int_0^\infty f_{\boldsymbol{\lambda}^\top(\boldsymbol{X}-\boldsymbol{\mu})-Z+\tau \,|\, \boldsymbol{T}=\boldsymbol{y}}(s)\mathrm{d}s}{\mathbb{P}(Z - \boldsymbol{\lambda}^\top(\boldsymbol{X}-\boldsymbol{\mu}) < \tau)} \\ &= \frac{\displaystyle\int_0^\infty f_{\boldsymbol{T},\boldsymbol{\lambda}^\top(\boldsymbol{X}-\boldsymbol{\mu})-Z+\tau}(\boldsymbol{y}, s)\mathrm{d}s}{\mathbb{P}(\boldsymbol{\lambda}^\top(\boldsymbol{X}-\boldsymbol{\mu}) + \tau > Z)} = f_{\boldsymbol{T} \,|\, \boldsymbol{\lambda}^\top(\boldsymbol{X}-\boldsymbol{\mu})+\tau>Z}(\boldsymbol{y}). \end{aligned}$$

This shows that $\boldsymbol{Y} \sim \text{EUGSE}_n(\boldsymbol{\mu}, \boldsymbol{\Sigma}, \boldsymbol{\lambda}, \tau, g^{(n)})$ admits the stochastic representation:

$$\boldsymbol{Y} = \boldsymbol{T} \,|\, \boldsymbol{\lambda}^\top(\boldsymbol{X}-\boldsymbol{\mu}) + \tau > Z, \tag{3.2.11}$$

where $\boldsymbol{T} = (G_1^{-1}(X_1), \ldots, G_n^{-1}(X_n))^\top$, and $\boldsymbol{X}$ and Z are distributionally related by Items (3.2.7)-(3.2.10).

3.2.3 Reparameterization for to enforce identifiability

In general, identifiability is lost when a multivariate normal distribution is reduced by conditioning (Florens, Mouchart, and Rolin, 1990). This leads us to believe that for any choices of density generators $\left(g^{(n)}\right)$ the EUGSE_n model (3.1.3) loses identifiability. In this subsection we will prove the non-identifiability of the EUGSE_n model. To this end, by considering the notations $\boldsymbol{\lambda}_* \equiv \boldsymbol{\omega}\boldsymbol{\lambda}$,

$$\boldsymbol{\Sigma}_* \equiv \boldsymbol{\omega}^{-1}\boldsymbol{\Sigma}\boldsymbol{\omega}^{-1} = \begin{pmatrix} 1 & \frac{\sigma_{12}}{\sqrt{\sigma_{11}\sigma_{22}}} & \cdots & \frac{\sigma_{1n}}{\sqrt{\sigma_{11}\sigma_{nn}}} \\ \frac{\sigma_{21}}{\sqrt{\sigma_{22}\sigma_{11}}} & 1 & \cdots & \frac{\sigma_{2n}}{\sqrt{\sigma_{22}\sigma_{nn}}} \\ \vdots & \vdots & \ddots & \vdots \\ \frac{\sigma_{n1}}{\sqrt{\sigma_{nn}\sigma_{11}}} & \frac{\sigma_{n2}}{\sqrt{\sigma_{nn}\sigma_{22}}} & \cdots & 1 \end{pmatrix},$$

$$\boldsymbol{\omega} \equiv \sqrt{\operatorname{diag}(\boldsymbol{\Sigma})} = \begin{pmatrix} \sqrt{\sigma_{11}} & 0 & \cdots & 0 \\ 0 & \sqrt{\sigma_{22}} & \cdots & 0 \\ 0 & 0 & \cdots & \sqrt{\sigma_{nn}} \end{pmatrix},$$

It is natural to ask whether through reparameterization the model gains the property of identifiability. At least for the extended G-skew-normal distribution (see Table 3.4) the answer is positive. To verify this statement we consider the reparameterization $(\boldsymbol{\mu}, \boldsymbol{\Sigma}, \boldsymbol{\lambda}, \tau)^\top \longmapsto \boldsymbol{\psi} = (\boldsymbol{\mu}, \boldsymbol{\Sigma}_*, \boldsymbol{\delta}, \boldsymbol{\gamma})^\top$, where $\boldsymbol{\Sigma}_* = \boldsymbol{\omega}^{-1}\boldsymbol{\Sigma}\boldsymbol{\omega}^{-1}$ is the correlation matrix defined in above section, and

$$\delta \equiv \frac{\boldsymbol{\Sigma}_*\boldsymbol{\lambda}}{\sqrt{1+\lambda^\top\boldsymbol{\Sigma}_*\boldsymbol{\lambda}}}, \quad \gamma \equiv \frac{\tau}{\sqrt{1+\lambda^\top\boldsymbol{\Sigma}_*\boldsymbol{\lambda}}}. \tag{3.2.12}$$

In what remains of this subsection we will prove that the parametrization ψ is identifiable. Indeed, note that

$$\delta^\top = \frac{\boldsymbol{\lambda}^\top\boldsymbol{\Sigma}_*}{\sqrt{1+\boldsymbol{\lambda}^\top\boldsymbol{\Sigma}_*\boldsymbol{\lambda}}} \Longrightarrow \sqrt{1+\boldsymbol{\lambda}^\top\boldsymbol{\Sigma}_*\boldsymbol{\lambda}} = \frac{1}{\sqrt{1-\boldsymbol{\delta}^\top\boldsymbol{\Sigma}_*^{-1}\delta}}. \tag{3.2.13}$$

By using (3.2.13), we obtain

$$\boldsymbol{\lambda}^\top = \boldsymbol{\delta}^\top\boldsymbol{\Sigma}_*^{-1}\sqrt{1+\boldsymbol{\lambda}^\top\boldsymbol{\Sigma}_*\boldsymbol{\lambda}} = \frac{\boldsymbol{\delta}^\top\boldsymbol{\Sigma}_*^{-1}}{\sqrt{1-\boldsymbol{\delta}^\top\boldsymbol{\Sigma}_*^{-1}\boldsymbol{\delta}}}, \tag{3.2.14}$$

$$\tau = \gamma\sqrt{1+\boldsymbol{\lambda}^\top\boldsymbol{\Sigma}_*\boldsymbol{\lambda}} = \frac{\gamma}{\sqrt{1-\boldsymbol{\delta}^\top\boldsymbol{\Sigma}_*^{-1}\delta}}. \tag{3.2.15}$$

Hence, by (3.2.12), (3.2.14) and (3.2.15), the extended G-skew-normal PDF (see Table 3.4) can be written as a function of ψ as follows:

$$f_{\boldsymbol{Y}}(\boldsymbol{y};\boldsymbol{\psi}) = \phi_n\left(\boldsymbol{y}_G;\boldsymbol{\mu},\boldsymbol{\Sigma}_*\right)\frac{\Phi\left(\frac{\boldsymbol{\delta}^\top\boldsymbol{\Sigma}_*^{-1}(\boldsymbol{y}_G-\boldsymbol{\mu})+\gamma}{\sqrt{1-\boldsymbol{\delta}^\top\boldsymbol{\Sigma}_*^{-1}\boldsymbol{\delta}}}\right)}{\Phi(\gamma)}\prod_{i=1}^n G_i'\left(y_i\right) = f_{\mathrm{SN}}\left(\boldsymbol{y}_G;\boldsymbol{\psi}\right)\prod_{i=1}^n G_i'\left(y_i\right),$$

where $f_{\mathrm{SN}}(\cdot;\psi)$ is the skew-normal distribution defined as see (Castro, San Martín, and Arellano-Valle, 2013)

$$f_{\mathrm{SN}}(\boldsymbol{z};\boldsymbol{\psi}) \equiv \phi_n\left(\boldsymbol{z};\boldsymbol{\mu},\boldsymbol{\Sigma}_*\right)\frac{\Phi\left(\frac{\boldsymbol{\delta}^\top\boldsymbol{\Sigma}_*^{-1}(\boldsymbol{z}-\boldsymbol{\mu})+\gamma}{\sqrt{1-\boldsymbol{\delta}^\top\boldsymbol{\Sigma}_*^{-1}\boldsymbol{\delta}}}\right)}{\Phi(\gamma)}, \quad \boldsymbol{z} \in \mathbb{R}^n. \tag{3.2.16}$$

By using the r th cumulants of random vector corresponding to PDF $f_{\mathrm{SN}}(\cdot;\boldsymbol{\psi})$, in Section 2

(Castro, San Martín, and Arellano-Valle, 2013), it was proven that the skew-normal distribution (3.2.16) is identifiable. In other words, it was shown that

$$f_{\mathrm{SN}}(\boldsymbol{z};\boldsymbol{\psi}) = f_{\mathrm{SN}}\left(\boldsymbol{z};\boldsymbol{\psi}'\right), \forall \boldsymbol{z} \in \mathbb{R}^n \quad \Longrightarrow \quad \boldsymbol{\psi} = \boldsymbol{\psi}'.$$

As an immediate consequence of the above result, we obtain

$$f_{\boldsymbol{Y}}(\boldsymbol{y};\boldsymbol{\psi}) = f_{\mathrm{SN}}\left(\boldsymbol{y}_G;\boldsymbol{\psi}\right)\prod_{i=1}^{n} G_i'\left(y_i\right) = f_{\mathrm{SN}}\left(\boldsymbol{y}_G;\boldsymbol{\psi}'\right)\prod_{i=1}^{n} G_i'\left(y_i\right)$$

$$= f_{\boldsymbol{Y}}\left(\boldsymbol{y};\boldsymbol{\psi}'\right), \forall \boldsymbol{y} \in D^n \quad \Longrightarrow \quad \boldsymbol{\psi} = \boldsymbol{\psi}'.$$

This shows the identifiability of the extended G-skew-normal distribution model when considering reparameterization $\psi = (\boldsymbol{\mu}, \boldsymbol{\Sigma}_*, \boldsymbol{\delta}, \boldsymbol{\gamma})^\top$.

3.2.4 Marginal Quantiles

We can rewrite $Y_i = T_i|\lambda^T(X-\mu)+\tau > Z; i = 1, ..., n$. Where $T_i = G_i^{-1}(X_i), i = 1, ..., n$.

Let $p \in (0,1)$. The p-quantile for Y_i (which we call marginal quantile for $\boldsymbol{Y} = (Y_1, ..., Y_n)^\top$), denoted by $Q_{Y_i}(p)$, is a real number such that:

$$\mathbb{P}(Y_i \le Q_{Y_i}(p)) = p, \quad i = 1, \ldots, n.$$

We can define the (conditional) random variable $W_i = X_i|\lambda^T(X-\mu)+\tau > Z, i = 1, ..., n$. Since G_i is monotone, we can rewarite the above relation, hence:

$$\begin{aligned} p = \mathbb{P}(Y_i \le Q_{Y_i}(p)) &= \mathbb{P}(T_i \le Q_{Y_i}(p)|\lambda^T(X-\mu)+\tau > Z) \\ &= \mathbb{P}(X_i \le G_i(Q_{Y_i}(p))|\lambda^T(X-\mu)+\tau > Z) \\ &= \mathbb{P}(W_i \le G_i(Q_{Y_i}(p))). \end{aligned}$$

Equivalently,

$$Q_{Y_i}(p) = G_i^{-1}(Q_{W_i}(p)), \quad i = 1, \ldots, n.$$

In other words, the p-quantile for Y_i is determined by the p-quantile for W_i, and vice-versa. Therefore, is suficient find the distribuition of the W_i to determinate the p-quantile for Y_i. The distribuition of W_i, for the cases considered in this work, are known (see Subsection 3.2.5).

3.2.5 Conditional distributions

In the context of multivariate sample selection models (Heckman, 1976), the interest lies in finding the PDF of $Y_i \,|\, Y_j > \kappa$, $i \neq j \in \{1, \ldots, n\}$, given that $\boldsymbol{Y} = (Y_1, \ldots, Y_n)^\top \sim \text{EUGSE}_n(\boldsymbol{\mu}, \boldsymbol{\Sigma}, \boldsymbol{\lambda}, \tau, g^{(n)})$, with $\kappa \in (0,1)$. For this purpose, let $\boldsymbol{W} = (W_1, \ldots, W_n)^\top = \boldsymbol{X} \,|\, \boldsymbol{\lambda}^\top \boldsymbol{X} + \tau > Z$ be a multivariate extended skew-elliptical random vector.

Analogously to the steps developed in (3.1.2), Bayes' rule provides

$$f_{Y_i \,|\, Y_j > \kappa}(y) = f_{Y_i}(y) \, \frac{\displaystyle\int_\kappa^\infty f_{Y_j \,|\, Y_i = y}(s) \mathrm{d}s}{\mathbb{P}(Y_j > \kappa)}, \quad y \in (0,1), \ \kappa \in (0,1). \tag{3.2.17}$$

If $Y_i = y$ then $W_i = G_i(y)$. So, the distribution of $Y_j \,|\, Y_i = y$ is the same as the distribution of $G_j^{-1}(W_j) \,|\, W_i = G_i(y)$. Consequently, the PDF of Y_j given $Y_i = y$ is given by

$$f_{Y_j \,|\, Y_i = y}(s) = f_{W_j \,|\, W_i = G_i(y)}(G_j(s)) \, G_j'(s). \tag{3.2.18}$$

Since $f_{Y_i}(y) = f_{W_i}(G_i(y)) G_i'(y)$ and $f_{Y_j}(s) = f_{W_j}(G_j(s)) G_j'(s)$, from (3.2.17) and (3.2.18) we get

$$f_{Y_i \,|\, Y_j > \kappa}(y) = f_{W_i}(G_i(y)) G_i'(y) \, \frac{\displaystyle\int_\kappa^\infty f_{W_j \,|\, W_i = G_i(y)}(G_j(s)) \, G_j'(s) \mathrm{d}s}{\displaystyle\int_\kappa^\infty f_{W_j}(G_j(s)) G_j'(s) \mathrm{d}s}.$$

Equivalently,

$$f_{Y_i \,|\, Y_j > \kappa}(y) = f_{W_i}(G_i(y)) G_i'(y) \, \frac{S_{W_j \,|\, W_i = G_i(y)}(G_j(\kappa))}{S_{W_j}(G_j(\kappa))}, \quad y \in (0,1), \ \kappa \in (0,1), \tag{3.2.19}$$

where S_X denotes the survival function (SF) of X. In other words, to determine the distribution of $Y_i \,|\, Y_j > \kappa$ it is sufficient to know the unconditional and conditional distributions of the multivariate extended skew-elliptical random vector $\boldsymbol{W}$.

In what remains of this subsection we present closed-forms for the PDF (3.2.19) of $Y_i \,|\, Y_j > \kappa$ by considering the Student-t, Cauchy and Gaussian generator densities.

Student-t density generator

Let $g^{(n)}(x) = (1 + x/\nu)^{-(\nu+n)/2}$, $x \in \mathbb{R}$ (see Table 3.3), be the Student-t density generator of the EUGSE$_n$ (multivariate extended unit-G-skew-Student-t) distribution.

Definition 3.2.1. *A random variable X follows a univariate extended skew-Student-t (EST$_1$) distribution, denoted by $X \sim$ EST$_1(\mu, \sigma^2, \lambda, \nu, \tau)$, if its PDF is given by (Arellano-Valle and Genton, 2010)*

$$f_{\mathrm{EST}_1}(x; \mu, \sigma^2, \lambda, \nu, \tau) = \frac{1}{\sigma}\, f_\nu(z)\, \frac{F_{\nu+1}\Big((\lambda z + \tau)\, \sqrt{\frac{\nu+1}{\nu+z^2}} \Big)}{F_\nu\Big(\frac{\tau}{\sqrt{1+\lambda^2}}\Big)}, \quad x \in \mathbb{R};\ \mu, \lambda, \tau \in \mathbb{R},\ \sigma, \nu > 0,$$

where $z = (x - \mu)/\sigma$, and f_ν and F_ν denote the PDF and CDF of the standard Student-t distribution with $\nu > 0$ degrees of freedom, respectively. Let $S_{\mathrm{ESN}_1}(x; \mu, \sigma^2, \lambda, \tau)$ be the SF corresponding to EST$_1$ PDF.

Let $\boldsymbol{W} = (W_1, \ldots, W_n)^\top = \boldsymbol{X} \,|\, \boldsymbol{\lambda}^\top \boldsymbol{X} + \tau > Z$. From Arellano-Valle and Genton, 2010, the unconditional and conditional distributions of $\boldsymbol{W}$ are respectively given by

$$W_i \sim \mathrm{EST}_1\left(\mu_i,\, \sigma_i^2,\, \frac{\lambda_i\sigma_i + \lambda_j\sigma_j\rho_{ij}}{\sigma_i\sqrt{1+\lambda_j^2\sigma_j^2(1-\rho_{ij}^2)}},\, \nu,\, \frac{\tau}{\sqrt{1+\lambda_j^2\sigma_j^2(1-\rho_{ij}^2)}}\right), \tag{3.2.20}$$

$$W_j \sim \mathrm{EST}_1\left(\mu_j,\, \sigma_j^2,\, \frac{\lambda_j\sigma_j + \lambda_i\sigma_i\rho_{ij}}{\sigma_j\sqrt{1+\lambda_i^2\sigma_i^2(1-\rho_{ij}^2)}},\, \nu,\, \frac{\tau}{\sqrt{1+\lambda_i^2\sigma_i^2(1-\rho_{ij}^2)}}\right), \tag{3.2.21}$$

and

$$W_j \,|\, W_i = y \sim \mathrm{EST}_1\left(\boldsymbol{\mu}_y,\, \boldsymbol{\sigma}^2_{y;\nu},\, \lambda_j\sigma_j\sqrt{1-\rho_{ij}^2},\, \nu+1,\, \boldsymbol{\tau}_{y;\nu}\right), \tag{3.2.22}$$

where we are adopting the following notation:

$$
\begin{aligned}
\boldsymbol{\mu}_y &= \mu_j + \sigma_j \rho_{ij} \left(\frac{y - \mu_i}{\sigma_i} \right); \\
\boldsymbol{\sigma}_{y;\nu}^2 &= \frac{\nu + \left(\frac{y - \mu_{1i}}{\sigma_i} \right)^2}{\nu + 1} \sigma_j^2 (1 - \rho_{ij}^2); \\
\boldsymbol{\tau}_{y;\nu} &= \left[(\lambda_i \sigma_i + \lambda_j \sigma_j \rho_{ij}) \left(\tfrac{y - \mu_i}{\sigma_i} \right) + \tau \right] \sqrt{\tfrac{\nu+1}{\nu + \left(\frac{y - \mu_i}{\sigma_i} \right)^2}}.
\end{aligned}
\tag{3.2.23}
$$

Hence, by combining (3.2.19) with (3.2.21), (3.2.22) and (3.2.23), we obtain

$$
\begin{aligned}
f_{Y_i \,|\, Y_j > \kappa}(y) = f_{\mathrm{EST}_1} &\left(G_i(y);\, \mu_i,\, \sigma_i^2,\, \frac{\lambda_i \sigma_i + \lambda_j \sigma_j \rho_{ij}}{\sigma_i \sqrt{1 + \lambda_j^2 \sigma_j^2 (1 - \rho_{ij}^2)}},\, \nu,\, \frac{\tau}{\sqrt{1 + \lambda_j^2 \sigma_j^2 (1 - \rho_{ij}^2)}} \right) G_i'(y) \\
&\times \frac{S_{\mathrm{EST}_1} \left(G_j(\kappa);\, \boldsymbol{\mu}_{G_i(y)},\, \boldsymbol{\sigma}^2_{G_i(y);\nu},\, \lambda_j \sigma_j \sqrt{1 - \rho_{ij}^2},\, \nu + 1,\, \boldsymbol{\tau}_{G_i(y);\nu} \right)}{S_{\mathrm{EST}_1} \left(G_j(\kappa);\, \mu_j,\, \sigma_j^2,\, \frac{\lambda_j \sigma_j + \lambda_i \sigma_i \rho_{ij}}{\sigma_j \sqrt{1 + \lambda_i^2 \sigma_i^2 (1 - \rho_{ij}^2)}},\, \nu,\, \frac{\tau}{\sqrt{1 + \lambda_i^2 \sigma_i^2 (1 - \rho_{ij}^2)}} \right)},
\end{aligned}
\tag{3.2.24}
$$

for $y \in (0, 1)$ and $\kappa \in (0, 1)$.

Example 3.2.1. *By taking $G_i(x) = \log(-\log(1 - x))$, $0 < x < 1$, $i = 1, \ldots, n$ (see Table 3.3), we get the multivariate asymmetric version of the unit-Student-t model addressed in (Vila et al., 2023b). So, from* (3.2.24) *and Table 3.1, we have*

$$
\begin{aligned}
f_{Y_i \,|\, Y_j > \kappa}(y) = f_{\mathrm{EST}_1} &\left(\log(-\log(1 - y));\, \mu_i,\, \sigma_i^2,\, \frac{\lambda_i \sigma_i + \lambda_j \sigma_j \rho_{ij}}{\sigma_i \sqrt{1 + \lambda_j^2 \sigma_j^2 (1 - \rho_{ij}^2)}},\, \nu,\, \frac{\tau}{\sqrt{1 + \lambda_j^2 \sigma_j^2 (1 - \rho_{ij}^2)}} \right) \frac{1}{(1 - y) \log(\frac{1}{1-y})} \\
&\times \frac{S_{\mathrm{EST}_1} \left(\log(-\log(1 - \kappa));\, \boldsymbol{\mu}_{\log(-\log(1-y))},\, \boldsymbol{\sigma}^2_{\log(-\log(1-y));\nu},\, \lambda_j \sigma_j \sqrt{1 - \rho_{ij}^2},\, \nu + 1,\, \boldsymbol{\tau}_{\log(-\log(1-y));\nu} \right)}{S_{\mathrm{EST}_1} \left(\log(-\log(1 - \kappa));\, \mu_j,\, \sigma_j^2,\, \frac{\lambda_j \sigma_j + \lambda_i \sigma_i \rho_{ij}}{\sigma_j \sqrt{1 + \lambda_i^2 \sigma_i^2 (1 - \rho_{ij}^2)}},\, \nu,\, \frac{\tau}{\sqrt{1 + \lambda_i^2 \sigma_i^2 (1 - \rho_{ij}^2)}} \right)},
\end{aligned}
$$

for $y \in (0, 1)$, $\kappa \in (0, 1)$, and $\boldsymbol{\mu}_y$, $\boldsymbol{\sigma}^2_{y;\nu}$ and $\boldsymbol{\tau}_{y;\nu}$ being as in (3.2.23).

Cauchy density generator

Let $g^{(n)}(x) = 1/(1+x)^{(n+1)/2}$, $x \in \mathbb{R}$ (see Table 3.3), be the Cauchy density generator of the EUGSE$_n$ (multivariate extended unit-G-skew-Cauchy) distribution.

By taking $\nu = 1$ in formula (3.2.24), we have

$$f_{Y_i\,|\,Y_j>\kappa}(y) = f_{\mathrm{EST}_1}\left(G_i(y);\, \mu_i,\, \sigma_i^2,\, \frac{\lambda_i\sigma_i + \lambda_j\sigma_j\rho_{ij}}{\sigma_i\sqrt{1+\lambda_j^2\sigma_j^2(1-\rho_{ij}^2)}},\, 1,\, \frac{\tau}{\sqrt{1+\lambda_j^2\sigma_j^2(1-\rho_{ij}^2)}}\right) G_i'(y)$$
$$\times \frac{S_{\mathrm{EST}_1}\left(G_j(\kappa);\, \boldsymbol{\mu}_{G_i(y)},\, \boldsymbol{\sigma}^2_{G_i(y);1},\, \lambda_j\sigma_j\sqrt{1-\rho_{ij}^2},\, 2,\, \boldsymbol{\tau}_{G_i(y);1}\right)}{S_{\mathrm{EST}_1}\left(G_j(\kappa);\, \mu_j,\, \sigma_j^2,\, \frac{\lambda_j\sigma_j + \lambda_i\sigma_i\rho_{ij}}{\sigma_j\sqrt{1+\lambda_i^2\sigma_i^2(1-\rho_{ij}^2)}},\, 1,\, \frac{\tau}{\sqrt{1+\lambda_i^2\sigma_i^2(1-\rho_{ij}^2)}}\right)}, \tag{3.2.25}$$

for $y \in (0,1)$ and $\kappa \in (0,1)$.

Example 3.2.2. *By taking $G_i(x) = (2x-1)/(x(1-x))$, $0 < x < 1$, $i = 1, \ldots, n$, from* (3.2.25) *and Table 3.1, we have*

$$f_{Y_i\,|\,Y_j>\kappa}(y) = f_{\mathrm{EST}_1}\left(\frac{2y-1}{y(1-y)};\, \mu_i,\, \sigma_i^2,\, \frac{\lambda_i\sigma_i + \lambda_j\sigma_j\rho_{ij}}{\sigma_i\sqrt{1+\lambda_j^2\sigma_j^2(1-\rho_{ij}^2)}},\, 1,\, \frac{\tau}{\sqrt{1+\lambda_j^2\sigma_j^2(1-\rho_{ij}^2)}}\right) \left[\frac{1}{(x-1)^2} + \frac{1}{x^2}\right]$$
$$\times \frac{S_{\mathrm{EST}_1}\left(\frac{2\kappa-1}{\kappa(1-\kappa)};\, \boldsymbol{\mu}_{\frac{2y-1}{y(1-y)}},\, \boldsymbol{\sigma}^2_{\frac{2y-1}{y(1-y)};1},\, \lambda_j\sigma_j\sqrt{1-\rho_{ij}^2},\, 2,\, \boldsymbol{\tau}_{\frac{2y-1}{y(1-y)};1}\right)}{S_{\mathrm{EST}_1}\left(\frac{2\kappa-1}{\kappa(1-\kappa)};\, \mu_j,\, \sigma_j^2,\, \frac{\lambda_j\sigma_j + \lambda_i\sigma_i\rho_{ij}}{\sigma_j\sqrt{1+\lambda_i^2\sigma_i^2(1-\rho_{ij}^2)}},\, 1,\, \frac{\tau}{\sqrt{1+\lambda_i^2\sigma_i^2(1-\rho_{ij}^2)}}\right)},$$

for $y \in (0,1)$, $\kappa \in (0,1)$, and $\boldsymbol{\mu}_y$, $\boldsymbol{\sigma}^2_{y;\nu}$ and $\boldsymbol{\tau}_{y;\nu}$ being as in (3.2.23).

Gaussian density generator

Let $g^{(n)}(x) = \exp(-x/2)$, $x \in \mathbb{R}$ (see Table 3.3), be the Gaussian density generator of the EUGSE$_n$ (multivariate extended unit-G-skew-normal) distribution.

Definition 3.2.2. *A random variable X follows a univariate extended skew-normal (ESN$_1$) distribution, denoted by $X \sim ESN_1(\mu, \sigma^2, \lambda, \tau)$, if its PDF is given by (Vernic, 2005; Arellano-*

Valle and Genton, 2010)

$$f_{\mathrm{ESN}_1}(x;\mu,\sigma^2,\lambda,\tau)=\frac{1}{\sigma}\,\phi(z)\,\frac{\Phi(\lambda z+\tau)}{\Phi\big(\frac{\tau}{\sqrt{1+\lambda^2}}\big)},\quad x\in\mathbb{R};\ \mu,\lambda,\tau\in\mathbb{R},\ \sigma>0,$$

where $z=(x-\mu)/\sigma$, and ϕ and Φ denote the PDF and CDF of the standard normal distribution, respectively. Let $S_{\mathrm{ESN}_1}(x;\mu,\sigma^2,\lambda,\tau)$ denote the SF corresponding to ESN$_1$ PDF.

Since

$$\lim_{\nu\to\infty}\boldsymbol{\sigma}^2_{y;\nu}=\sigma_j^2(1-\rho_{ij}^2),\quad \lim_{\nu\to\infty}\boldsymbol{\tau}_{y;\nu}=(\lambda_i\sigma_i+\lambda_j\sigma_j\rho_{ij})\left(\frac{y-\mu_i}{\sigma_i}\right)+\tau,$$

and $\lim_{\nu\to\infty}f_{\mathrm{EST}_1}(x;\mu,\sigma^2,\lambda,\nu,\tau)=f_{\mathrm{ESN}_1}(x;\mu,\sigma^2,\lambda,\tau)$, by letting $\nu\to\infty$ in (3.2.24), we obtain

$$f_{Y_i\,|\,Y_j>\kappa}(y)=f_{\mathrm{ESN}_1}\left(G_i(y);\,\mu_i,\,\sigma_i^2,\,\frac{\lambda_i\sigma_i+\lambda_j\sigma_j\rho_{ij}}{\sigma_i\sqrt{1+\lambda_j^2\sigma_j^2(1-\rho_{ij}^2)}},\,\nu,\,\frac{\tau}{\sqrt{1+\lambda_j^2\sigma_j^2(1-\rho_{ij}^2)}}\right)G_i'(y)$$
$$\times\frac{S_{\mathrm{ESN}_1}\left(G_j(\kappa);\,\mu_j+\sigma_j\rho_{ij}\left(\frac{G_i(y)-\mu_i}{\sigma_i}\right),\,\sigma_j^2(1-\rho_{ij}^2),\,\lambda_j\sigma_j\sqrt{1-\rho_{ij}^2},\,(\lambda_i\sigma_i+\lambda_j\sigma_j\rho_{ij})\left(\frac{G_i(y)-\mu_i}{\sigma_i}\right)+\tau\right)}{S_{\mathrm{ESN}_1}\left(G_j(\kappa);\,\mu_j,\,\sigma_j^2,\,\frac{\lambda_j\sigma_j+\lambda_i\sigma_i\rho_{ij}}{\sigma_j\sqrt{1+\lambda_i^2\sigma_i^2(1-\rho_{ij}^2)}},\,\nu,\,\frac{\tau}{\sqrt{1+\lambda_i^2\sigma_i^2(1-\rho_{ij}^2)}}\right)},\tag{3.2.26}$$

for $y\in(0,1)$ and $\kappa\in(0,1)$.

Example 3.2.3. *By taking $G_i(x)=\tan((x-1/2)\pi)$, $0<x<1$, $i=1,\ldots,n$, from* (3.2.26) *and Table 3.1, we have*

$$f_{Y_i\,|\,Y_j>\kappa}(y)=f_{\mathrm{ESN}_1}\left(\tan\left(\left(y-\frac{1}{2}\right)\pi\right);\,\mu_i,\,\sigma_i^2,\,\frac{\lambda_i\sigma_i+\lambda_j\sigma_j\rho_{ij}}{\sigma_i\sqrt{1+\lambda_j^2\sigma_j^2(1-\rho_{ij}^2)}},\,\nu,\,\frac{\tau}{\sqrt{1+\lambda_j^2\sigma_j^2(1-\rho_{ij}^2)}}\right)\frac{\pi}{\sin^2(\pi y)}$$
$$\times\frac{S_{\mathrm{ESN}_1}\left(\tan((\kappa-\frac{1}{2})\pi);\,\mu_j+\sigma_j\rho_{ij}\left(\frac{\tan((y-\frac{1}{2})\pi)-\mu_i}{\sigma_i}\right),\,\sigma_j^2(1-\rho_{ij}^2),\,\lambda_j\sigma_j\sqrt{1-\rho_{ij}^2},\,(\lambda_i\sigma_i+\lambda_j\sigma_j\rho_{ij})\left(\frac{\tan((y-\frac{1}{2})\pi)-\mu_i}{\sigma_i}\right)+\tau\right)}{S_{\mathrm{ESN}_1}\left(\tan((\kappa-\frac{1}{2})\pi);\,\mu_j,\,\sigma_j^2,\,\frac{\lambda_j\sigma_j+\lambda_i\sigma_i\rho_{ij}}{\sigma_j\sqrt{1+\lambda_i^2\sigma_i^2(1-\rho_{ij}^2)}},\,\nu,\,\frac{\tau}{\sqrt{1+\lambda_i^2\sigma_i^2(1-\rho_{ij}^2)}}\right)},$$

for $y\in(0,1)$ and $\kappa\in(0,1)$.

3.2.6 Expected value of a function of an EUGSE$_n$ random vector

Let $\boldsymbol{Y} = (Y_1, \ldots, Y_n)^\top \sim \text{EUGSE}_n(\boldsymbol{\mu}, \boldsymbol{\Sigma}, \boldsymbol{\lambda}, \tau, g^{(n)})$. If $\varphi : (0,1)^n \to \mathbb{R}$ is a real-valued measurable-analytic function, from stochastic representation in Subsection 3.2.2 it follows that

$$\varphi(\boldsymbol{Y}) \stackrel{d}{=} \varphi(G_1^{-1}(W_1), \ldots, G_n^{-1}(W_n)),$$

where $\boldsymbol{W} \sim \text{EUGSE}_n(\boldsymbol{\mu}, \boldsymbol{\Sigma}, \boldsymbol{\lambda}, \tau, g^{(n)})$. Let $\psi = \varphi \circ (G_1^{-1} \circ \pi_1, \ldots, G_n^{-1} \circ \pi_n)$ denote the composition function of φ with $(G_1^{-1} \circ \pi_1, \ldots, G_n^{-1} \circ \pi_n)$, where π_k denotes the kth projection function. The above representation is written as

$$\varphi(\boldsymbol{Y}) \stackrel{d}{=} \psi(\boldsymbol{W}),$$

which implies that

$$\mathbb{E}[\varphi(\boldsymbol{Y})] = \mathbb{E}[\psi(\boldsymbol{W})] = \int_{\mathbb{R}^n} \psi(\boldsymbol{w}) f_{\boldsymbol{W}}(\boldsymbol{w}) \mathrm{d}\boldsymbol{w}. \tag{3.2.27}$$

Consider $\boldsymbol{v} = (v_1, \ldots, v_n)^\top \in \mathbb{R}^n$ an n-dimensional vector. Upon using the multivariate Taylor expansion of function $\boldsymbol{w} \longmapsto \psi(\boldsymbol{w})$ around the point $\boldsymbol{v}$, that is (committing an abuse of notation),

$$\begin{aligned}
\psi(\boldsymbol{w} + \boldsymbol{v}) &= \left(\sum_{k=0}^{\infty} \frac{1}{k!} \sum_{i_1, \ldots, i_k = 1}^{n} w_{i_1} \cdots w_{i_k} \frac{\partial^k}{\partial v_{i_1} \cdots v_{i_k}} \right) \psi(\boldsymbol{v}) \\
&= \left(\sum_{k=0}^{\infty} \frac{1}{k!} (\boldsymbol{w}^\top \boldsymbol{\nabla}_{\boldsymbol{v}})^k \right) \psi(\boldsymbol{v}), \quad \text{with } \boldsymbol{\nabla}_{\boldsymbol{v}} = \left(\frac{\partial}{\partial v_1}, \ldots, \frac{\partial}{\partial v_n} \right)^\top, \\
&= \exp(\boldsymbol{w}^\top \boldsymbol{\nabla}_{\boldsymbol{v}}) \psi(\boldsymbol{v}),
\end{aligned} \tag{3.2.28}$$

the expectation in (3.2.27) becomes

$$\begin{aligned}
\mathbb{E}[\varphi(\boldsymbol{Y})] &= \int_{\mathbb{R}^n} \left[\psi(\boldsymbol{w} + \boldsymbol{v}) \big|_{\boldsymbol{v} = \boldsymbol{0}} \right] f_{\boldsymbol{W}}(\boldsymbol{w}) \mathrm{d}\boldsymbol{w} \\
&= \int_{\mathbb{R}^n} \left[\exp(\boldsymbol{w}^\top \boldsymbol{\nabla}_{\boldsymbol{v}}) \psi(\boldsymbol{v}) \big|_{\boldsymbol{v} = \boldsymbol{0}} \right] f_{\boldsymbol{W}}(\boldsymbol{w}) \mathrm{d}\boldsymbol{w} \\
&= \left[\int_{\mathbb{R}^n} \exp(\boldsymbol{w}^\top \boldsymbol{\nabla}_{\boldsymbol{v}}) f_{\boldsymbol{W}}(\boldsymbol{w}) \mathrm{d}\boldsymbol{w} \right] \psi(\boldsymbol{v}) \Bigg|_{\boldsymbol{v} = \boldsymbol{0}} = M_{\boldsymbol{W}}(\boldsymbol{\nabla}_{\boldsymbol{v}}) \psi(\boldsymbol{v}) \big|_{\boldsymbol{v} = \boldsymbol{0}},
\end{aligned} \tag{3.2.29}$$

where

$$\psi(\boldsymbol{v}) = \varphi(G_1^{-1}(v_1), \ldots, G_n^{-1}(v_n)) \tag{3.2.30}$$

and $M_{\boldsymbol{W}}(\boldsymbol{s})$ is the moment generating function (MGF) of the multivariate random vector $\boldsymbol{W}$, whenever it exists.

In the case that $\boldsymbol{Y}$ has a multivariate extended unit-G-skew-normal distribution (see Table 3.3) case, $\boldsymbol{W}$ follows an multivariate extended skew-normal distribution (see Table 3.4) with parameter vector $(\boldsymbol{\mu}, \boldsymbol{\Sigma}, \boldsymbol{\lambda}, \tau)^\top$. So, by using the definition of PDF $f_{\boldsymbol{W}}$ given in (3.4), we have

$$\begin{aligned} M_{\boldsymbol{W}}(\boldsymbol{s}) &= \int_{\mathbb{R}^n} \exp(\boldsymbol{s}^\top \boldsymbol{w}) f_{\boldsymbol{W}}(\boldsymbol{w}) \mathrm{d}\boldsymbol{w} \\ &= \int_{\mathbb{R}^n} \exp(\boldsymbol{s}^\top \boldsymbol{w}) \phi_n(\boldsymbol{w};\, \boldsymbol{\mu}, \boldsymbol{\Sigma}) \, \frac{\Phi\left(\boldsymbol{\lambda}^\top(\boldsymbol{w} - \boldsymbol{\mu}) + \tau\right)}{\Phi\Big(\frac{\tau}{\sqrt{1+\boldsymbol{\lambda}^\top \boldsymbol{\Sigma} \boldsymbol{\lambda}}}\Big)} \mathrm{d}\boldsymbol{w}. \end{aligned}$$

A simple observation shows that

$$\exp(\boldsymbol{s}^\top \boldsymbol{w}) \phi_n(\boldsymbol{w};\, \boldsymbol{\mu}, \boldsymbol{\Sigma}) = \exp\left(\boldsymbol{s}^\top \boldsymbol{\mu} + \frac{1}{2}\, \boldsymbol{s}^\top \boldsymbol{\Sigma} \boldsymbol{s}\right) \phi_n(\boldsymbol{w};\, \boldsymbol{\mu}^*, \boldsymbol{\Sigma}), \quad \boldsymbol{\mu}^* = \boldsymbol{\mu} + \boldsymbol{\Sigma} \boldsymbol{s}.$$

Then, upon using the above identity, the MGF of $\boldsymbol{W}$ is

$$M_{\boldsymbol{W}}(\boldsymbol{s}) = \exp\left(\boldsymbol{s}^\top \boldsymbol{\mu} + \frac{1}{2}\, \boldsymbol{s}^\top \boldsymbol{\Sigma} \boldsymbol{s}\right) \frac{\Phi\big(\frac{\tau^*}{\sqrt{1+\boldsymbol{\lambda}^\top \boldsymbol{\Sigma} \boldsymbol{\lambda}}}\big)}{\Phi\big(\frac{\tau}{\sqrt{1+\boldsymbol{\lambda}^\top \boldsymbol{\Sigma} \boldsymbol{\lambda}}}\big)} \int_{\mathbb{R}^n} \phi_n(\boldsymbol{w};\, \boldsymbol{\mu}^*, \boldsymbol{\Sigma}) \, \frac{\Phi\left(\boldsymbol{\lambda}^\top(\boldsymbol{w} - \boldsymbol{\mu}^*) + \tau^*\right)}{\Phi\Big(\frac{\tau^*}{\sqrt{1+\boldsymbol{\lambda}^\top \boldsymbol{\Sigma} \boldsymbol{\lambda}}}\Big)} \mathrm{d}\boldsymbol{w},$$

with $\tau^* = \boldsymbol{\lambda}^\top \boldsymbol{\Sigma} \boldsymbol{s} + \tau$. Let $\boldsymbol{W}^*$ be a random vector following a multivariate extended skew-normal distribution (see Table 3.3) with parameter vector $(\boldsymbol{\mu}^*, \boldsymbol{\Sigma}, \boldsymbol{\lambda}, \tau^*)$. Using this notation, the MGF of $\boldsymbol{W}$ is expressed as

$$\begin{aligned} M_{\boldsymbol{W}}(\boldsymbol{s}) &= \exp\left(\boldsymbol{s}^\top \boldsymbol{\mu} + \frac{1}{2}\, \boldsymbol{s}^\top \boldsymbol{\Sigma} \boldsymbol{s}\right) \frac{\Phi\big(\frac{\tau^*}{\sqrt{1+\boldsymbol{\lambda}^\top \boldsymbol{\Sigma} \boldsymbol{\lambda}}}\big)}{\Phi\big(\frac{\tau}{\sqrt{1+\boldsymbol{\lambda}^\top \boldsymbol{\Sigma} \boldsymbol{\lambda}}}\big)} \int_{\mathbb{R}^n} f_{\boldsymbol{W}^*}(\boldsymbol{w}) \mathrm{d}\boldsymbol{w} \\ &= \frac{1}{\Phi\big(\frac{\tau}{\sqrt{1+\boldsymbol{\lambda}^\top \boldsymbol{\Sigma} \boldsymbol{\lambda}}}\big)} \exp\left(\boldsymbol{s}^\top \boldsymbol{\mu} + \frac{1}{2}\, \boldsymbol{s}^\top \boldsymbol{\Sigma} \boldsymbol{s}\right) \Phi\left(\frac{\boldsymbol{\lambda}^\top \boldsymbol{\Sigma} \boldsymbol{s} + \tau}{\sqrt{1+\boldsymbol{\lambda}^\top \boldsymbol{\Sigma} \boldsymbol{\lambda}}}\right). \end{aligned}$$

Replacing the above formula in (3.2.29), we have

$$\mathbb{E}[\varphi(\boldsymbol{Y})] = \left[\exp(\boldsymbol{\nabla}_{\boldsymbol{v}}^\top \boldsymbol{\mu})\psi(\boldsymbol{v})\big|_{\boldsymbol{v}=\mathbf{0}}\right] \left[\exp\left(\frac{1}{2}\boldsymbol{\nabla}_{\boldsymbol{v}}^\top \boldsymbol{\Sigma}\boldsymbol{\nabla}_{\boldsymbol{v}}\right)\psi(\boldsymbol{v})\Bigg|_{\boldsymbol{v}=\mathbf{0}}\right] \left[\frac{\Phi\left(\frac{\boldsymbol{\lambda}^\top\boldsymbol{\Sigma}\boldsymbol{\nabla}_{\boldsymbol{v}}+\tau}{\sqrt{1+\boldsymbol{\lambda}^\top\boldsymbol{\Sigma}\boldsymbol{\lambda}}}\right)}{\Phi\left(\frac{\tau}{\sqrt{1+\boldsymbol{\lambda}^\top\boldsymbol{\Sigma}\boldsymbol{\lambda}}}\right)}\psi(\boldsymbol{v})\Bigg|_{\boldsymbol{v}=\mathbf{0}}\right].$$

By using the multivariate Taylor expansion (3.2.28), $\exp(\boldsymbol{\nabla}_{\boldsymbol{v}}^\top \boldsymbol{\mu})\psi(\boldsymbol{v}) = \psi(\boldsymbol{\mu}+\boldsymbol{v})$. Then, we obtain the following closed formula for the expected value of a function of $\boldsymbol{Y}$:

$$\mathbb{E}[\varphi(\boldsymbol{Y})] = \psi(\boldsymbol{\mu}) \left[\exp\left(\frac{1}{2}\boldsymbol{\nabla}_{\boldsymbol{v}}^\top \boldsymbol{\Sigma}\boldsymbol{\nabla}_{\boldsymbol{v}}\right)\psi(\boldsymbol{v})\Bigg|_{\boldsymbol{v}=\mathbf{0}}\right] \left[\frac{\Phi\left(\frac{\boldsymbol{\lambda}^\top\boldsymbol{\Sigma}\boldsymbol{\nabla}_{\boldsymbol{v}}+\tau}{\sqrt{1+\boldsymbol{\lambda}^\top\boldsymbol{\Sigma}\boldsymbol{\lambda}}}\right)}{\Phi\left(\frac{\tau}{\sqrt{1+\boldsymbol{\lambda}^\top\boldsymbol{\Sigma}\boldsymbol{\lambda}}}\right)}\psi(\boldsymbol{v})\Bigg|_{\boldsymbol{v}=\mathbf{0}}\right], \tag{3.2.31}$$

with ψ being as in (3.2.30).

Remark 3.2.4. *(i) When the extension parameter is absent, that is, $\tau = 0$, we have*

$$\mathbb{E}[\varphi(\boldsymbol{Y})] = 2\psi(\boldsymbol{\mu}) \left[\exp\left(\frac{1}{2}\boldsymbol{\nabla}_{\boldsymbol{v}}^\top \boldsymbol{\Sigma}\boldsymbol{\nabla}_{\boldsymbol{v}}\right)\psi(\boldsymbol{v})\Bigg|_{\boldsymbol{v}=\mathbf{0}}\right] \left[\Phi\left(\frac{\boldsymbol{\lambda}^\top\boldsymbol{\Sigma}\boldsymbol{\nabla}_{\boldsymbol{v}}}{\sqrt{1+\boldsymbol{\lambda}^\top\boldsymbol{\Sigma}\boldsymbol{\lambda}}}\right)\psi(\boldsymbol{v})\Bigg|_{\boldsymbol{v}=\mathbf{0}}\right].$$

(ii) When the skewness parameter is absent, that is, $\boldsymbol{\lambda} = \mathbf{0}$, we have

$$\mathbb{E}[\varphi(\boldsymbol{Y})] = \psi(\boldsymbol{\mu}) \left[\exp\left(\frac{1}{2}\boldsymbol{\nabla}_{\boldsymbol{v}}^\top \boldsymbol{\Sigma}\boldsymbol{\nabla}_{\boldsymbol{v}}\right)\psi(\boldsymbol{v})\Bigg|_{\boldsymbol{v}=\mathbf{0}}\right],$$

Remark 3.2.5. *(i) The exponential operator* $\exp\left(\boldsymbol{\nabla}_{\boldsymbol{v}}^\top \boldsymbol{\Sigma}\boldsymbol{\nabla}_{\boldsymbol{v}}/2\right)$ *that appears in* (3.2.31) *can be written as*

$$\begin{aligned}\exp\left(\frac{1}{2}\,\boldsymbol{\nabla}_{\boldsymbol{v}}^\top \boldsymbol{\Sigma}\boldsymbol{\nabla}_{\boldsymbol{v}}\right) &= \sum_{k=0}^{\infty}\frac{1}{k!}\left(\frac{1}{2}\,\boldsymbol{\nabla}_{\boldsymbol{v}}^\top \boldsymbol{\Sigma}\boldsymbol{\nabla}_{\boldsymbol{v}}\right)^k \\ &= \sum_{k=0}^{\infty}\frac{1}{k!}\,\frac{1}{2^k}\sum_{j_1,l_1,\ldots,j_k,l_k=1}^{n}\sigma_{j_1l_1}\cdots\sigma_{j_kl_k}\,\frac{\partial^{2k}}{\partial v_{j_1}\partial v_{l_1}\cdots\partial v_{j_k}\partial v_{l_k}}.\end{aligned} \tag{3.2.32}$$

(ii) By using the series representation of the Gaussian CDF:

$$\Phi(x) = \frac{1}{2} + \frac{1}{\sqrt{\pi}} \sum_{k=0}^{\infty} \frac{(-1)^{3k} 2^{-\frac{1}{2}-k}}{(1+2k)k!}\, x^{2k},$$

the operator $\Phi((\boldsymbol{\lambda}^\top \boldsymbol{\Sigma} \boldsymbol{\nabla}_{\boldsymbol{v}} + \tau)/\sqrt{1+\boldsymbol{\lambda}^\top \boldsymbol{\Sigma}\boldsymbol{\lambda}}\,)$ *that appears in* (3.2.31) *can be written as*

$$\begin{aligned}
\Phi\left(\frac{\boldsymbol{\lambda}^\top \boldsymbol{\Sigma} \boldsymbol{\nabla}_{\boldsymbol{v}} + \tau}{\sqrt{1+\boldsymbol{\lambda}^\top \boldsymbol{\Sigma}\boldsymbol{\lambda}}}\right) &= \frac{1}{2} + \frac{1}{\sqrt{\pi}} \sum_{k=0}^{\infty} \frac{(-1)^{3k} 2^{-\frac{1}{2}-k}}{(1+2k)k!} \left(\frac{\boldsymbol{\lambda}^\top \boldsymbol{\Sigma} \boldsymbol{\nabla}_{\boldsymbol{v}} + \tau}{\sqrt{1+\boldsymbol{\lambda}^\top \boldsymbol{\Sigma}\boldsymbol{\lambda}}}\right)^{2k} \\
&= \frac{1}{2} + \frac{1}{\sqrt{\pi}} \sum_{k=0}^{\infty} \frac{(-1)^{3k} 2^{-\frac{1}{2}-k}}{(1+2k)k!} \sum_{r=0}^{2k} \binom{2k}{r} \left(\frac{\tau}{\sqrt{1+\boldsymbol{\lambda}^\top \boldsymbol{\Sigma}\boldsymbol{\lambda}}}\right)^{2k-r} \\
&\quad \times \frac{\displaystyle\sum_{j_1,l_1,\ldots,j_r,l_r=1}^{n} \sigma_{l_1 j_1} \cdots \sigma_{l_r j_r} \lambda_{l_1} \cdots \lambda_{l_r} \frac{\partial^r}{\partial v_{j_1} \cdots \partial v_{j_r}}}{(\sqrt{1+\boldsymbol{\lambda}^\top \boldsymbol{\Sigma}\boldsymbol{\lambda}}\,)^r},
\end{aligned} \tag{3.2.33}$$

where in the last equality a binomial expansion was used.

Remark 3.2.6. *Since* $\mathbb{E}[\varphi(\boldsymbol{Y})]$ *in* (3.2.31) *depends on the operator formulas in* (3.2.32) *and* (3.2.33)*, these can be used to facilitate its calculation.*

Mixed-moments

Let $\varphi(\boldsymbol{y}) = \prod_{i=1}^n \pi_i^m(\boldsymbol{y}) = \prod_{i=1}^n y_i^{m_i}$, where π_i is the ith projection function. From (3.2.29) we have the next formula for the mixed-moments of $\boldsymbol{Y}$:

$$\mathbb{E}\left(\prod_{i=1}^n Y_i^{m_i}\right) = M_{\boldsymbol{W}}(\boldsymbol{\nabla}_{\boldsymbol{v}}) \prod_{i=1}^n [G_i^{-1}(v_i)]^{m_i}\Big|_{\boldsymbol{v}=\boldsymbol{0}}.$$

In the case that $\boldsymbol{Y}$ has a multivariate extended unit-G-skew-normal distribution (see Table

3.3) case, from (3.2.31) we have

$$\begin{aligned}\mathbb{E}\left(\prod_{i=1}^{n} Y_i^{m_i}\right) &= \prod_{i=1}^{n}[G_i^{-1}(\mu_i)]^{m_i}\left[\exp\left(\frac{1}{2}\boldsymbol{\nabla}_{\boldsymbol{v}}^{\top}\boldsymbol{\Sigma}\boldsymbol{\nabla}_{\boldsymbol{v}}\right)\prod_{i=1}^{n}[G_i^{-1}(v_i)]^{m_i}\bigg|_{\boldsymbol{v}=\boldsymbol{0}}\right] \\ &\times\left[\frac{\Phi\left(\frac{\boldsymbol{\lambda}^{\top}\boldsymbol{\Sigma}\boldsymbol{\nabla}_{\boldsymbol{v}}+\tau}{\sqrt{1+\boldsymbol{\lambda}^{\top}\boldsymbol{\Sigma}\boldsymbol{\lambda}}}\right)}{\Phi\left(\frac{\tau}{\sqrt{1+\boldsymbol{\lambda}^{\top}\boldsymbol{\Sigma}\boldsymbol{\lambda}}}\right)}\prod_{i=1}^{n}[G_i^{-1}(v_i)]^{m_i}\bigg|_{\boldsymbol{v}=\boldsymbol{0}}\right].\end{aligned}\tag{3.2.34}$$

It is clear that the above formula is extremely complicated for functions G_is as chosen in Table 3.1. For illustration purposes, let us consider $G_i(x) = \log(x)$, $x > 0$. So, by using formula in (3.2.32), we have

$$\exp\left(\frac{1}{2}\,\boldsymbol{\nabla}_{\boldsymbol{v}}^{\top}\boldsymbol{\Sigma}\boldsymbol{\nabla}_{\boldsymbol{v}}\right)\prod_{i=1}^{n}[G_i^{-1}(v_i)]^{m_i} = \exp\left(\frac{1}{2}\,\boldsymbol{m}^{\top}\boldsymbol{\Sigma}\boldsymbol{m}+\boldsymbol{m}^{\top}\boldsymbol{v}\right).$$

On the other hand, by using formula in (3.2.33), we obtain

$$\Phi\left(\frac{\boldsymbol{\lambda}^{\top}\boldsymbol{\Sigma}\boldsymbol{\nabla}_{\boldsymbol{v}}+\tau}{\sqrt{1+\boldsymbol{\lambda}^{\top}\boldsymbol{\Sigma}\boldsymbol{\lambda}}}\right)\prod_{i=1}^{n}[G_i^{-1}(v_i)]^{m_i} = \Phi\left(\frac{\boldsymbol{\lambda}^{\top}\boldsymbol{\Sigma}\boldsymbol{m}+\tau}{\sqrt{1+\boldsymbol{\lambda}^{\top}\boldsymbol{\Sigma}\boldsymbol{\lambda}}}\right)\exp(\boldsymbol{m}^{\top}\boldsymbol{v}).$$

Replacing the last two expressions in (3.2.34), we obtain

$$\mathbb{E}\left(\prod_{i=1}^{n} Y_i^{m_i}\right) = \exp\left(\boldsymbol{m}^{\top}\boldsymbol{\mu}+\frac{1}{2}\,\boldsymbol{m}^{\top}\boldsymbol{\Sigma}\boldsymbol{m}\right)\frac{\Phi\left(\frac{\boldsymbol{\lambda}^{\top}\boldsymbol{\Sigma}\boldsymbol{m}+\tau}{\sqrt{1+\boldsymbol{\lambda}^{\top}\boldsymbol{\Sigma}\boldsymbol{\lambda}}}\right)}{\Phi\left(\frac{\tau}{\sqrt{1+\boldsymbol{\lambda}^{\top}\boldsymbol{\Sigma}\boldsymbol{\lambda}}}\right)}.$$

The above formula has appeared in (Marchenko and Genton, 2010) for the special case $\tau = 0$. In particular,

$$\mathbb{E}\left(Y_i^{m}\right) = \exp\left(m\mu_i+\frac{1}{2}\,m^2\sigma_{ii}\right)\frac{\Phi\left(\frac{m\sum_{k=1}^{n}\lambda_k\sigma_{ki}+\tau}{\sqrt{1+\boldsymbol{\lambda}^{\top}\boldsymbol{\Sigma}\boldsymbol{\lambda}}}\right)}{\Phi\left(\frac{\tau}{\sqrt{1+\boldsymbol{\lambda}^{\top}\boldsymbol{\Sigma}\boldsymbol{\lambda}}}\right)}.$$

Marginal moments

By letting φ as the kth power of the ith projection function, that is, $\varphi(\boldsymbol{y}) = \pi_i^k(\boldsymbol{y}) = y_i^k$, $i = 1, \ldots, n$, from (3.2.29) we have the next formula for the marginal moments:

$$\mathbb{E}(Y_i^k) = M_{\boldsymbol{T}}(\nabla_{\boldsymbol{v}})\pi_i^k(\boldsymbol{v})H(\boldsymbol{v})\big|_{\boldsymbol{v}=\boldsymbol{0}^+}.$$

Note that the moments are finite. Indeed, since $G^{-1} : \mathbb{R} \longrightarrow (0,1)$

$$(Y_1, ..., Y_n) \stackrel{d}{=} (G_1^{-1}(W_1), ..., G_n^{-1}(W_n)),$$

and

$$0 \leq \prod_{i=1}^{n} Y_i^{m_i} \leq 1,$$

we have

$$0 \leq \mathbb{E}\biggl[\prod_{i=1}^{n} Y_i^{m_i}\biggr] = \mathbb{E}\biggl[\prod_{i=1}^{n} (G_i^{-1}(W_i))^{m_i}\biggr] \leq 1.$$

Therefore the moments are finite.

Let φ be the ith projection function raised to the mth power, that is, $\varphi(\boldsymbol{y}) = \pi_i^m(\boldsymbol{y}) = y_i^m$, $i = 1, \ldots, n$. From (3.2.29) we have the next formula for the marginal moments of $\boldsymbol{Y}$:

$$\mathbb{E}(Y_i^m) = M_{\boldsymbol{W}}(\boldsymbol{\nabla}_{\boldsymbol{v}})[G_i^{-1}(v_i)]^m\big|_{v_i=0}.$$

In the case that $\boldsymbol{Y}$ has a multivariate extended unit-G-skew-normal distribution (see Table 3.3) case, from (3.2.31) we have

$$\mathbb{E}(Y_i^m) = [G_i^{-1}(\mu_i)]^m \left[\exp\left(\frac{1}{2}\boldsymbol{\nabla}_{\boldsymbol{v}}^\top \boldsymbol{\Sigma} \boldsymbol{\nabla}_{\boldsymbol{v}}\right) [G_i^{-1}(v_i)]^m\bigg|_{v_i=0}\right] \tag{3.2.35}$$

$$\times \left[\frac{\Phi\left(\frac{\boldsymbol{\lambda}^\top \boldsymbol{\Sigma} \boldsymbol{\nabla}_{\boldsymbol{v}} + \tau}{\sqrt{1+\boldsymbol{\lambda}^\top \boldsymbol{\Sigma}\boldsymbol{\lambda}}}\right)}{\Phi\left(\frac{\tau}{\sqrt{1+\boldsymbol{\lambda}^\top \boldsymbol{\Sigma}\boldsymbol{\lambda}}}\right)} [G_i^{-1}(v_i)]^m\Bigg|_{v_i=0}\right], \quad i = 1, \ldots, n. \tag{3.2.36}$$

By using formula in (3.2.32), we have

$$\exp\left(\frac{1}{2}\,\boldsymbol{\nabla}_{\boldsymbol{v}}^\top \boldsymbol{\Sigma} \boldsymbol{\nabla}_{\boldsymbol{v}}\right) [G_i^{-1}(v_i)]^m = \exp\left(\frac{\sigma_{ii}^2}{2}\,\frac{\partial^2}{\partial v_i^2}\right) [G_i^{-1}(v_i)]^m. \tag{3.2.37}$$

On the other hand, by using formula in (3.2.33), we obtain

$$\Phi\left(\frac{\boldsymbol{\lambda}^\top\boldsymbol{\Sigma}\boldsymbol{\nabla}_{\boldsymbol{v}}+\tau}{\sqrt{1+\boldsymbol{\lambda}^\top\boldsymbol{\Sigma}\boldsymbol{\lambda}}}\right)[G_i^{-1}(v_i)]^m=\Phi\left(\frac{(\sum_{l=1}^n\sigma_{li}\lambda_l)\frac{\partial}{\partial v_i}+\tau}{\sqrt{1+\boldsymbol{\lambda}^\top\boldsymbol{\Sigma}\boldsymbol{\lambda}}}\right)[G_i^{-1}(v_i)]^m. \tag{3.2.38}$$

Replacing the expressions (3.2.37) and (3.2.38) in (3.2.35), we obtain the following simple closed formula for the marginal moments of the multivariate extended unit-skew-normal random vector $\boldsymbol{Y}$:

$$\begin{aligned}\mathbb{E}(Y_i^m)=[G_i^{-1}(\mu_i)]^m&\left[\exp\left(\frac{\sigma_{ii}^2}{2}\frac{\partial^2}{\partial v_i^2}\right)[G_i^{-1}(v_i)]^m\bigg|_{v_i=0}\right]\\&\times\left[\frac{\Phi\left(\frac{(\sum_{l=1}^n\sigma_{li}\lambda_l)\frac{\partial}{\partial v_i}+\tau}{\sqrt{1+\boldsymbol{\lambda}^\top\boldsymbol{\Sigma}\boldsymbol{\lambda}}}\right)}{\Phi\left(\frac{\tau}{\sqrt{1+\boldsymbol{\lambda}^\top\boldsymbol{\Sigma}\boldsymbol{\lambda}}}\right)}[G_i^{-1}(v_i)]^m\bigg|_{v_i=0}\right].\end{aligned} \tag{3.2.39}$$

Correlation function

By considering $\varphi(\boldsymbol{y})=\pi_i(\boldsymbol{y})\pi_j(\boldsymbol{y})$, $i,j=1,\dots,n$, where π_k denotes the kth projection function, from (3.2.29) we have the following formula for the cross-moments:

$$\mathbb{E}(Y_iY_j)=M_{\boldsymbol{T}}(\nabla_{\boldsymbol{v}})\pi_i(\boldsymbol{v})\pi_j(\boldsymbol{v})H(\boldsymbol{v})\big|_{\boldsymbol{v}=\boldsymbol{0}^+}.$$

Consequently, from the above identity and from Subsubsection 3.2.6, the correlation function between Y_i and Y_j, denoted by ρ_{Y_i,Y_j}, is written as

$$\rho_{Y_i,Y_j}=\frac{M_{\boldsymbol{T}}(\nabla_{\boldsymbol{v}})\big[\pi_i(\boldsymbol{v})\pi_j(\boldsymbol{v})-\pi_i(\boldsymbol{v})\pi_j(\boldsymbol{v})\big]H(\boldsymbol{v})}{\prod_{r=i,j}\sqrt{M_{\boldsymbol{T}}(\nabla_{\boldsymbol{v}})\pi_r^2(\boldsymbol{v})H(\boldsymbol{v})-[M_{\boldsymbol{T}}(\nabla_{\boldsymbol{v}})\pi_r(\boldsymbol{v})H(\boldsymbol{v})]^2}}\Bigg|_{\boldsymbol{v}=\boldsymbol{0}^+}.$$

3.3 Maximum likelihood estimation

Let $\{\boldsymbol{Y}_k=(Y_{1k},Y_{2k},\dots,Y_{nk})^\top : k=1,\dots,m\}$ be a multivariate random sample of size m from $\boldsymbol{Y}\sim$ EUGSE$_n(\boldsymbol{\mu},\boldsymbol{\Sigma},\boldsymbol{\lambda},\tau,g^{(n)})$ with joint PDF as given in (3.1.9), and let $\boldsymbol{y}_k=(y_{1k},y_{2k},\dots,y_{nk})^\top$ be a realization of $\boldsymbol{Y}_k$. To obtain the maximum likelihood estimates (MLEs) of the model parameters with parameter vector $\boldsymbol{\theta}=(\boldsymbol{\mu},\boldsymbol{\Sigma},\boldsymbol{\lambda},\tau)^\top$, we maximize the following log-likelihood function

$$
\begin{aligned}
\ell(\boldsymbol{\theta}) = &\sum_{k=1}^{m} \log(f_{\boldsymbol{X}}(\boldsymbol{y}_{G,k})) + \sum_{k=1}^{m} \log(F_{\mathrm{ELL}_1}(\boldsymbol{\lambda}^\top(\boldsymbol{y}_{G,k} - \boldsymbol{\mu}) + \tau;\, 0, 1, g_{q(\boldsymbol{y}_{G,k})})) \\
&- m\log(F_{\mathrm{ELL}_1}(\tau;\, 0, 1 + \boldsymbol{\lambda}^\top\boldsymbol{\Sigma}\boldsymbol{\lambda}, g^{(1)})) + \sum_{k=1}^{m}\sum_{i=1}^{n} \log(G_i'(y_{ik})),
\end{aligned}
$$

where $\boldsymbol{y}_{G,k} = (G_1(y_{1k}), \ldots, G_n(y_{nk}))^\top$. As $\boldsymbol{X} \sim \mathrm{ELL}_n(\boldsymbol{\mu}, \boldsymbol{\Sigma}, g^{(n)})$, by (3.1.3), the log-likelihood function (without the additive constant) is written as

$$
\begin{aligned}
\ell(\boldsymbol{\theta}) = &\frac{m}{2}\log(|\boldsymbol{\Sigma}^{-1}|) + \sum_{k=1}^{m} \log(g^{(n+1)}((\boldsymbol{y}_{G,k} - \boldsymbol{\mu})^\top\boldsymbol{\Sigma}^{-1}(\boldsymbol{y}_{G,k} - \boldsymbol{\mu}))) \\
&+ \sum_{k=1}^{m} \log\left(\int_{-\infty}^{\boldsymbol{\lambda}^\top(\boldsymbol{y}_{G,k}-\boldsymbol{\mu})+\tau} g^{(2)}(s^2 + (\boldsymbol{y}_{G,k} - \boldsymbol{\mu})^\top\boldsymbol{\Sigma}^{-1}(\boldsymbol{y}_{G,k} - \boldsymbol{\mu}))\mathrm{d}s\right) \\
&- \sum_{k=1}^{m} \log(g^{(1)}((\boldsymbol{y}_{G,k} - \boldsymbol{\mu})^\top\boldsymbol{\Sigma}^{-1}(\boldsymbol{y}_{G,k} - \boldsymbol{\mu}))) \\
&+ \frac{m}{2}\log(1 + \boldsymbol{\lambda}^\top\boldsymbol{\Sigma}\boldsymbol{\lambda}) - m\log\left(\int_{-\infty}^{\tau} g^{(1)}\left(\frac{s^2}{1 + \boldsymbol{\lambda}^\top\boldsymbol{\Sigma}\boldsymbol{\lambda}}\right)\mathrm{d}s\right).
\end{aligned}
$$

The likelihood equations are given by

$$
\frac{\partial \ell(\boldsymbol{\theta})}{\partial \boldsymbol{\mu}} = \boldsymbol{0}_{n\times 1}, \quad \frac{\partial \ell(\boldsymbol{\theta})}{\partial \boldsymbol{\Sigma}^{-1}} = \boldsymbol{0}_{n\times n}, \quad \frac{\partial \ell(\boldsymbol{\theta})}{\partial \boldsymbol{\lambda}} = \boldsymbol{0}_{n\times 1}, \quad \frac{\partial \ell(\boldsymbol{\theta})}{\partial \tau} = 0.
$$

In what follows we determine $\partial\ell(\boldsymbol{\theta})/\partial\boldsymbol{\mu}$, $\partial\ell(\boldsymbol{\theta})/\partial\boldsymbol{\Sigma}^{-1}$, $\partial\ell(\boldsymbol{\theta})/\partial\boldsymbol{\lambda}$ and $\partial\ell(\boldsymbol{\theta})/\partial\tau$. Indeed, by using the identities

$$
\frac{\partial \boldsymbol{a}^\top\boldsymbol{x}}{\partial \boldsymbol{x}} = \boldsymbol{a}^\top, \quad \frac{\partial \boldsymbol{x}^\top\boldsymbol{A}\boldsymbol{x}}{\partial \boldsymbol{x}} = 2\boldsymbol{A}\boldsymbol{x}, \quad \frac{\partial \boldsymbol{x}^\top\boldsymbol{A}\boldsymbol{x}}{\partial \boldsymbol{A}} = \boldsymbol{x}\boldsymbol{x}^\top,
$$

$$
\frac{\partial \boldsymbol{x}^\top\boldsymbol{A}^{-1}\boldsymbol{x}}{\partial \boldsymbol{A}} = -\boldsymbol{A}^{-\top}\boldsymbol{x}\boldsymbol{x}^\top\boldsymbol{A}^{-\top}, \quad \frac{\partial \log(|\boldsymbol{A}|)}{\partial \boldsymbol{A}} = \boldsymbol{A}^{-\top},
$$

with $\boldsymbol{A}$ being a $n \times n$ invertible matrix and $\boldsymbol{x}$ an n-dimensional vector, we have

(i)

$$
\begin{aligned}
\frac{\partial \ell(\boldsymbol{\theta})}{\partial \boldsymbol{\mu}} &= -2\boldsymbol{\Sigma}^{-1} \sum_{k=1}^{m} (\boldsymbol{y}_{G,k} - \boldsymbol{\mu}) \, \frac{[g^{(n+1)}]'((\boldsymbol{y}_{G,k} - \boldsymbol{\mu})^\top \boldsymbol{\Sigma}^{-1} (\boldsymbol{y}_{G,k} - \boldsymbol{\mu}))}{g^{(n+1)}((\boldsymbol{y}_{G,k} - \boldsymbol{\mu})^\top \boldsymbol{\Sigma}^{-1} (\boldsymbol{y}_{G,k} - \boldsymbol{\mu}))} \\
&- \boldsymbol{\lambda}^\top \sum_{k=1}^{m} \frac{g^{(2)}([\boldsymbol{\lambda}^\top (\boldsymbol{y}_{G,k} - \boldsymbol{\mu}) + \tau]^2 + (\boldsymbol{y}_{G,k} - \boldsymbol{\mu})^\top \boldsymbol{\Sigma}^{-1} (\boldsymbol{y}_{G,k} - \boldsymbol{\mu}))}{\int_{-\infty}^{\boldsymbol{\lambda}^\top (\boldsymbol{y}_{G,k} - \boldsymbol{\mu}) + \tau} g^{(2)}(s^2 + (\boldsymbol{y}_{G,k} - \boldsymbol{\mu})^\top \boldsymbol{\Sigma}^{-1} (\boldsymbol{y}_{G,k} - \boldsymbol{\mu})) \mathrm{d}s} \\
&- 2\boldsymbol{\Sigma}^{-1} \sum_{k=1}^{m} (\boldsymbol{y}_{G,k} - \boldsymbol{\mu}) \, \frac{\int_{-\infty}^{\boldsymbol{\lambda}^\top (\boldsymbol{y}_{G,k} - \boldsymbol{\mu}) + \tau} [g^{(2)}]'(s^2 + (\boldsymbol{y}_{G,k} - \boldsymbol{\mu})^\top \boldsymbol{\Sigma}^{-1} (\boldsymbol{y}_{G,k} - \boldsymbol{\mu})) \mathrm{d}s}{\int_{-\infty}^{\boldsymbol{\lambda}^\top (\boldsymbol{y}_{G,k} - \boldsymbol{\mu}) + \tau} g^{(2)}(s^2 + (\boldsymbol{y}_{G,k} - \boldsymbol{\mu})^\top \boldsymbol{\Sigma}^{-1} (\boldsymbol{y}_{G,k} - \boldsymbol{\mu})) \mathrm{d}s} \\
&+ 2\boldsymbol{\Sigma}^{-1} \sum_{k=1}^{m} (\boldsymbol{y}_{G,k} - \boldsymbol{\mu}) \, \frac{[g^{(1)}]'((\boldsymbol{y}_{G,k} - \boldsymbol{\mu})^\top \boldsymbol{\Sigma}^{-1} (\boldsymbol{y}_{G,k} - \boldsymbol{\mu}))}{g^{(1)}((\boldsymbol{y}_{G,k} - \boldsymbol{\mu})^\top \boldsymbol{\Sigma}^{-1} (\boldsymbol{y}_{G,k} - \boldsymbol{\mu}))},
\end{aligned}
$$

(ii)

$$
\begin{aligned}
\frac{\partial \ell(\boldsymbol{\theta})}{\partial \boldsymbol{\Sigma}^{-1}} &= \frac{m}{2} \boldsymbol{\Sigma} + \sum_{k=1}^{m} (\boldsymbol{y}_{G,k} - \boldsymbol{\mu})(\boldsymbol{y}_{G,k} - \boldsymbol{\mu})^\top \, \frac{[g^{(n+1)}]'((\boldsymbol{y}_{G,k} - \boldsymbol{\mu})^\top \boldsymbol{\Sigma}^{-1} (\boldsymbol{y}_{G,k} - \boldsymbol{\mu}))}{g^{(n+1)}((\boldsymbol{y}_{G,k} - \boldsymbol{\mu})^\top \boldsymbol{\Sigma}^{-1} (\boldsymbol{y}_{G,k} - \boldsymbol{\mu}))} \\
&+ \sum_{k=1}^{m} (\boldsymbol{y}_{G,k} - \boldsymbol{\mu})(\boldsymbol{y}_{G,k} - \boldsymbol{\mu})^\top \, \frac{\int_{-\infty}^{\boldsymbol{\lambda}^\top (\boldsymbol{y}_{G,k} - \boldsymbol{\mu}) + \tau} [g^{(2)}]'(s^2 + (\boldsymbol{y}_{G,k} - \boldsymbol{\mu})^\top \boldsymbol{\Sigma}^{-1} (\boldsymbol{y}_{G,k} - \boldsymbol{\mu})) \mathrm{d}s}{\int_{-\infty}^{\boldsymbol{\lambda}^\top (\boldsymbol{y}_{G,k} - \boldsymbol{\mu}) + \tau} g^{(2)}(s^2 + (\boldsymbol{y}_{G,k} - \boldsymbol{\mu})^\top \boldsymbol{\Sigma}^{-1} (\boldsymbol{y}_{G,k} - \boldsymbol{\mu})) \mathrm{d}s} \\
&- \sum_{k=1}^{m} (\boldsymbol{y}_{G,k} - \boldsymbol{\mu})(\boldsymbol{y}_{G,k} - \boldsymbol{\mu})^\top \, \frac{[g^{(1)}]'((\boldsymbol{y}_{G,k} - \boldsymbol{\mu})^\top \boldsymbol{\Sigma}^{-1} (\boldsymbol{y}_{G,k} - \boldsymbol{\mu}))}{g^{(1)}((\boldsymbol{y}_{G,k} - \boldsymbol{\mu})^\top \boldsymbol{\Sigma}^{-1} (\boldsymbol{y}_{G,k} - \boldsymbol{\mu}))} \\
&- \frac{m}{2} \, \frac{\boldsymbol{\Sigma} \boldsymbol{\lambda} \boldsymbol{\lambda}^\top \boldsymbol{\Sigma}}{1 + \boldsymbol{\lambda}^\top \boldsymbol{\Sigma} \boldsymbol{\lambda}} - m \, \frac{\boldsymbol{\Sigma} \boldsymbol{\lambda} \boldsymbol{\lambda}^\top \boldsymbol{\Sigma}}{(1 + \boldsymbol{\lambda}^\top \boldsymbol{\Sigma} \boldsymbol{\lambda})^2} \, \frac{\int_{-\infty}^{\tau} s^2 \, [g^{(1)}]' \big(\frac{s^2}{1 + \boldsymbol{\lambda}^\top \boldsymbol{\Sigma} \boldsymbol{\lambda}} \big) \mathrm{d}s}{\int_{-\infty}^{\tau} g^{(1)} \big(\frac{s^2}{1 + \boldsymbol{\lambda}^\top \boldsymbol{\Sigma} \boldsymbol{\lambda}} \big) \mathrm{d}s},
\end{aligned}
$$

(iii)

$$
\begin{aligned}
\frac{\partial \ell(\boldsymbol{\theta})}{\partial \boldsymbol{\lambda}} &= \sum_{k=1}^{m} (\boldsymbol{y}_{G,k} - \boldsymbol{\mu}) \, \frac{g^{(2)}([\boldsymbol{\lambda}^\top (\boldsymbol{y}_{G,k} - \boldsymbol{\mu}) + \tau]^2 + (\boldsymbol{y}_{G,k} - \boldsymbol{\mu})^\top \boldsymbol{\Sigma}^{-1} (\boldsymbol{y}_{G,k} - \boldsymbol{\mu}))}{\int_{-\infty}^{\boldsymbol{\lambda}^\top (\boldsymbol{y}_{G,k} - \boldsymbol{\mu}) + \tau} g^{(2)}(s^2 + (\boldsymbol{y}_{G,k} - \boldsymbol{\mu})^\top \boldsymbol{\Sigma}^{-1} (\boldsymbol{y}_{G,k} - \boldsymbol{\mu})) \mathrm{d}s} \\
&+ m \, \frac{\boldsymbol{\Sigma} \boldsymbol{\lambda}}{1 + \boldsymbol{\lambda}^\top \boldsymbol{\Sigma} \boldsymbol{\lambda}} + 2m \, \frac{\boldsymbol{\Sigma} \boldsymbol{\lambda}}{(1 + \boldsymbol{\lambda}^\top \boldsymbol{\Sigma} \boldsymbol{\lambda})^2} \, \frac{\int_{-\infty}^{\tau} s^2 [g^{(1)}]' \big(\frac{s^2}{1 + \boldsymbol{\lambda}^\top \boldsymbol{\Sigma} \boldsymbol{\lambda}} \big) \mathrm{d}s}{\int_{-\infty}^{\tau} g^{(1)} \big(\frac{s^2}{1 + \boldsymbol{\lambda}^\top \boldsymbol{\Sigma} \boldsymbol{\lambda}} \big) \mathrm{d}s},
\end{aligned}
$$

(iv)

$$\frac{\partial \ell(\boldsymbol{\theta})}{\partial \tau} = \sum_{k=1}^{m} \frac{g^{(2)}([\boldsymbol{\lambda}^\top(\boldsymbol{y}_{G,k} - \boldsymbol{\mu}) + \tau]^2 + (\boldsymbol{y}_{G,k} - \boldsymbol{\mu})^\top \boldsymbol{\Sigma}^{-1}(\boldsymbol{y}_{G,k} - \boldsymbol{\mu}))}{\int_{-\infty}^{\boldsymbol{\lambda}^\top(\boldsymbol{y}_{G,k}-\boldsymbol{\mu})+\tau} g^{(2)}(s^2 + (\boldsymbol{y}_{G.k} - \boldsymbol{\mu})^\top \boldsymbol{\Sigma}^{-1}(\boldsymbol{y}_{G,k} - \boldsymbol{\mu}))\mathrm{d}s}$$
$$- m \, \frac{g^{(1)}\left(\frac{\tau^2}{1+\boldsymbol{\lambda}^\top \boldsymbol{\Sigma}\boldsymbol{\lambda}}\right)}{\int_{-\infty}^{\tau} g^{(1)}\left(\frac{s^2}{1+\boldsymbol{\lambda}^\top \boldsymbol{\Sigma}\boldsymbol{\lambda}}\right)\mathrm{d}s}.$$

No closed-form solution to the maximization problem is available. As such, the maximum likelihood (ML) estimator of $\boldsymbol{\theta}$, denoted by $\widehat{\boldsymbol{\theta}}$, can only be obtained via numerical optimization. If $I(\boldsymbol{\theta}_0)$ denotes the expected Fisher information matrix, where $\boldsymbol{\theta}_0$ is the true value of the population parameter vector, then, under well-known regularity conditions (Davison, 2008), it follows that

$$\sqrt{m}[I(\boldsymbol{\theta}_0)]^{1/2}(\widehat{\boldsymbol{\theta}} - \boldsymbol{\theta}_0) \xrightarrow{d} N(\mathbf{0}_{(n+1)^2\times 1}, I_{(n+1)^2\times(n+1)^2}), \quad \text{as } m \to \infty,$$

where $\mathbf{0}_{(n+1)^2\times 1}$ is the $(n+1)^2\times$ zero vector, and $I_{(n+1)^2\times(n+1)^2}$ is the $(n+1)^2 \times (n+1)^2$ identity matrix. Since the expected Fisher information can be approximated by its observed version (obtained from the Hessian matrix), we can use the diagonal elements of this observed version to approximate the standard errors of the ML estimates.

No closed-form solution to the maximization problem is available. As such, the maximum likelihood (ML) estimator of $\boldsymbol{\theta}$, denoted by $\widehat{\boldsymbol{\theta}}$, can only be obtained via numerical optimization.

Chapter 4

Simulation study and Applications

This chapter deals with simulation and application aspects of the family of distributions presented in the previous chapter. Simulation studies and the application of the model to real data are proposed. The simulation study was performed with versions of the probability density function representing the distribution of the data from the model. Maximum likelihood estimation was used in conjunction with the Monte Carlo algorithm. The analyzes used to evaluate the parameter estimates were relative bias and mean square error. To better illustrate the results, plots showing the behavior of these two metrics for each of the parameters are presented. In addition, several functions were provided to perform the simulation study. A small representative selection of these functions is presented in the main body of the text, while the others can be found in the appendix of this work.

The application to real data was performed with a real data set from the R software. The descriptive statistics of the data is presented and commented on. Two density functions derived from the model are then fitted and the fit is evaluated using some metrics, which are briefly presented and discussed. After discussing the data, it is indicated which distribution best fits the data set based on the criteria considered and the G functions chosen for the model. Finally, conclusions regarding the application of the data and the estimation of the parameters are presented within a general perspective of the developed work.

4.1 Monte Carlo simulation

In this section, a simulation study is conducted for evaluating the performance of the maximum likelihood estimators. The simulation study considers the estimation of model parameters in the bivariate case. Different sample sizes and parameter settings using the extended unit-G-

skew-normal distribution were evaluated.

The performance and recovery of the maximum likelihood estimators are evaluated by means of the relative bias (RB) and the root mean square error (RMSE), given by

$$\widehat{\mathrm{RB}}(\widehat{\theta}) = \frac{1}{N}\sum_{i=1}^{N}\left|\frac{(\widehat{\theta}^{(i)}-\theta)}{\theta}\right|, \quad \widehat{\mathrm{RMSE}}(\widehat{\theta}) = \sqrt{\frac{1}{N}\sum_{i=1}^{N}(\widehat{\theta}^{(i)}-\theta)^2},$$

where θ and $\widehat{\theta}^{(i)}$ are the true parameter value and its i-th estimate. The simulation scenario was configured as follows: the sample size varies between $n \in \{200, 500, 1000, 2000\}$, with the true parameters defined as

$$(\mu_1, \mu_2, \lambda_1, \lambda_2, \tau, \sigma_1, \sigma_2)^\top = (1, 1, 0.5, 0.6, 0.5, 1, 1)^\top,$$

and ρ assuming values $\{0.10, 0.25, 0.50, 0.75, 0.90\}$. In all cases 100 Monte Carlo replications were performed for each configuration. In the simulation study, two specific functions were used, as detailed in Table 3.1: $G_i(x) = \tan((x-1/2)\pi)$; $G_i(x) = \log(x^3/(1-x^3))$; $G_i(x) = \log(x^5/(1-x^5))$.

The numerical methods used to estimate the parameters were the Nelder-Mead and BFGS optimization methods. The Nelder-Mead method is a widely used technique for numerical optimization, especially effective in minimizing unconstrained nonlinear functions. Its main advantage is that it does not require the calculation of derivatives, making it ideal for situations in which these are unknown or difficult to obtain. Belonging to the class of direct search methods, Nelder-Mead is based exclusively on evaluations of the function to be minimized.

The method operates with the concept of simplex, which can be understood graphically as a geometric figure formed by $n+1$ vertices in a space of n dimensions. In two dimensions, the simplex is a triangle; in three, a tetrahedron; and, in higher dimensions, it follows the same logic. The core of the algorithm consists of iteratively adjusting the shape and position of this simplex to locate the minimum point of the function. At each iteration, the function is evaluated at the vertices of the simplex, and geometric operations such as reflection, expansion, and contraction are applied to guide the simplex toward the minimum (Mathews and Fink, 2004).

The BFGS (Broyden-Fletcher-Goldfarb-Shanno) method is a widely used numerical optimization technique for minimizing nonlinear differentiable functions. It belongs to the class of quasi-Newton methods, which seek to approximate the solution of equations derived from gradient-based methods. Its main objective is to find the minimum of a function without having to directly calculate the Hessian matrix, which contains the second derivatives of the objective function. BFGS uses objective function gradient information to iteratively construct an approximation of the inverse Hessian matrix, which is crucial for determining the most efficient descent direction. The traditional Newton method requires the exact inverse of the Hessian, which can be computationally expensive or even infeasible for high-dimensional functions. BFGS, in turn, builds an updated approximation of the inverse of the Hessian at each iteration, based on the differences between the gradients of consecutive points (Nocedal and Wright, 1999).

Figures 4.1–4.6 show maximum likelihood estimation results considering the BFGS method. In general, for both optimization methods it was possible to observe the same patterns, with little or almost no difference in the results observed. From this figures, it is possible to observe a clear convergence of the RB towards zero for all parameters as sample sizes increase. This pattern is also evident when analyzing the RMSE, indicating a decrease in the corresponding variance as the sample size increases. From Figure 4.4, it is observed that the RMSE of $\widehat{\lambda}_1$ does not consistently decrease across all possibilities for ρ. Several factors may influence this behavior, such as the sample size, the number of iterations, or the inverse transformation G_i^{-1} used.

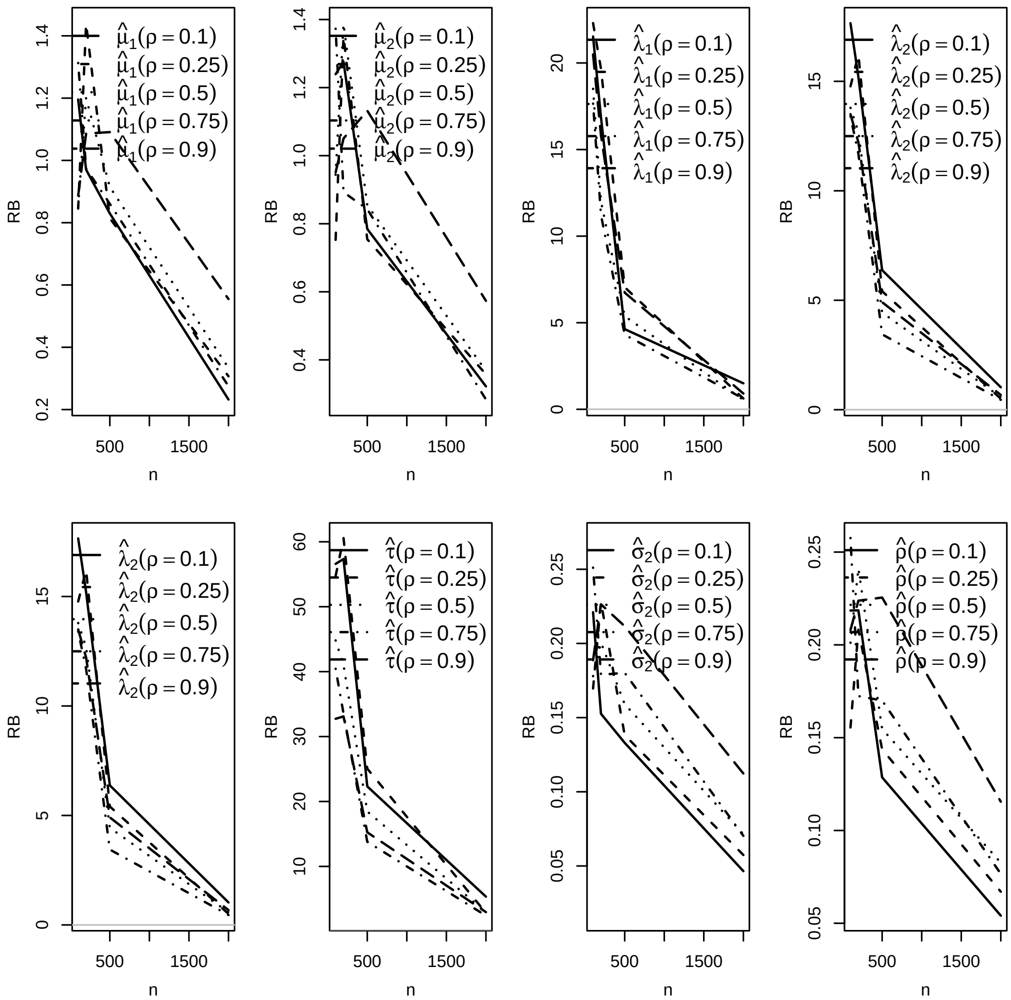


Figure 4.1: Relative bias for $G_i^{-1}(x) = \frac{1}{2} + \frac{\arctan(x)}{\pi}$.

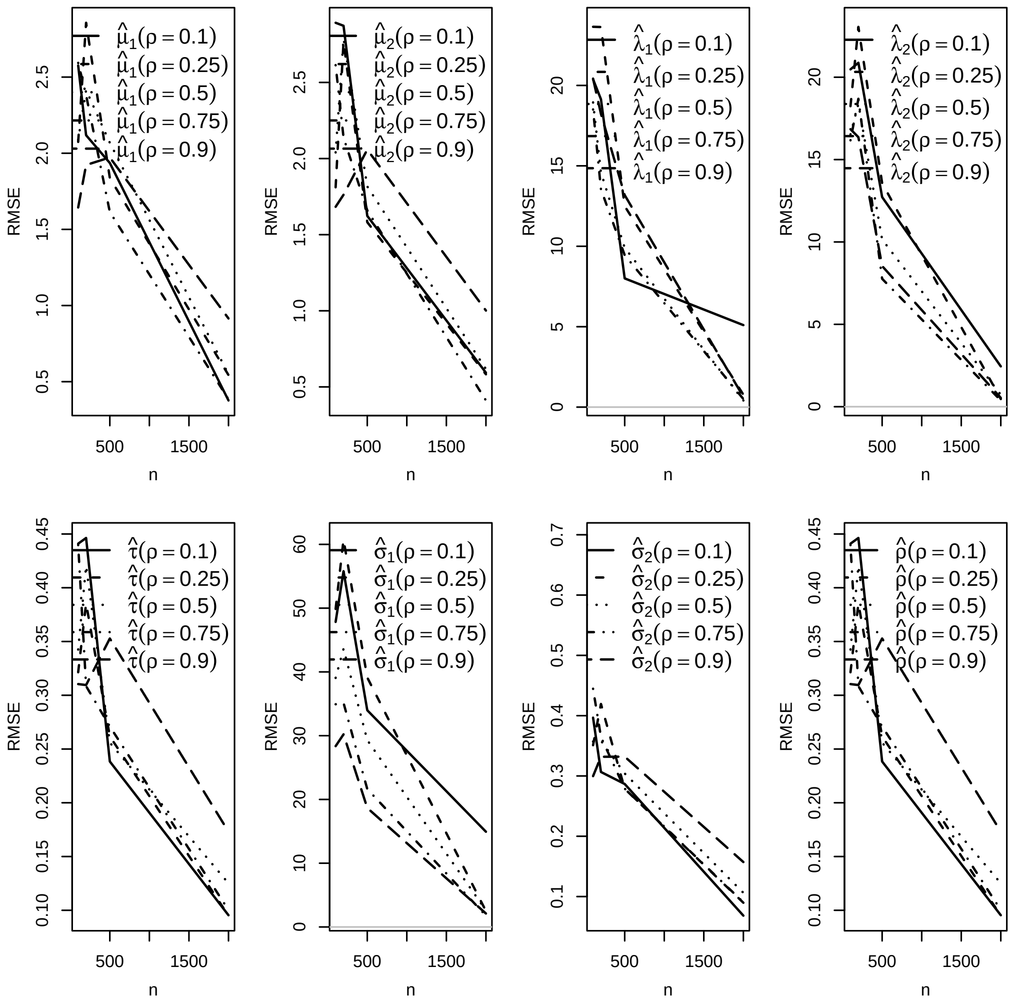


Figure 4.2: Root mean squared error for $G_i^{-1}(x) = \frac{1}{2} + \frac{\arctan(x)}{\pi}$.

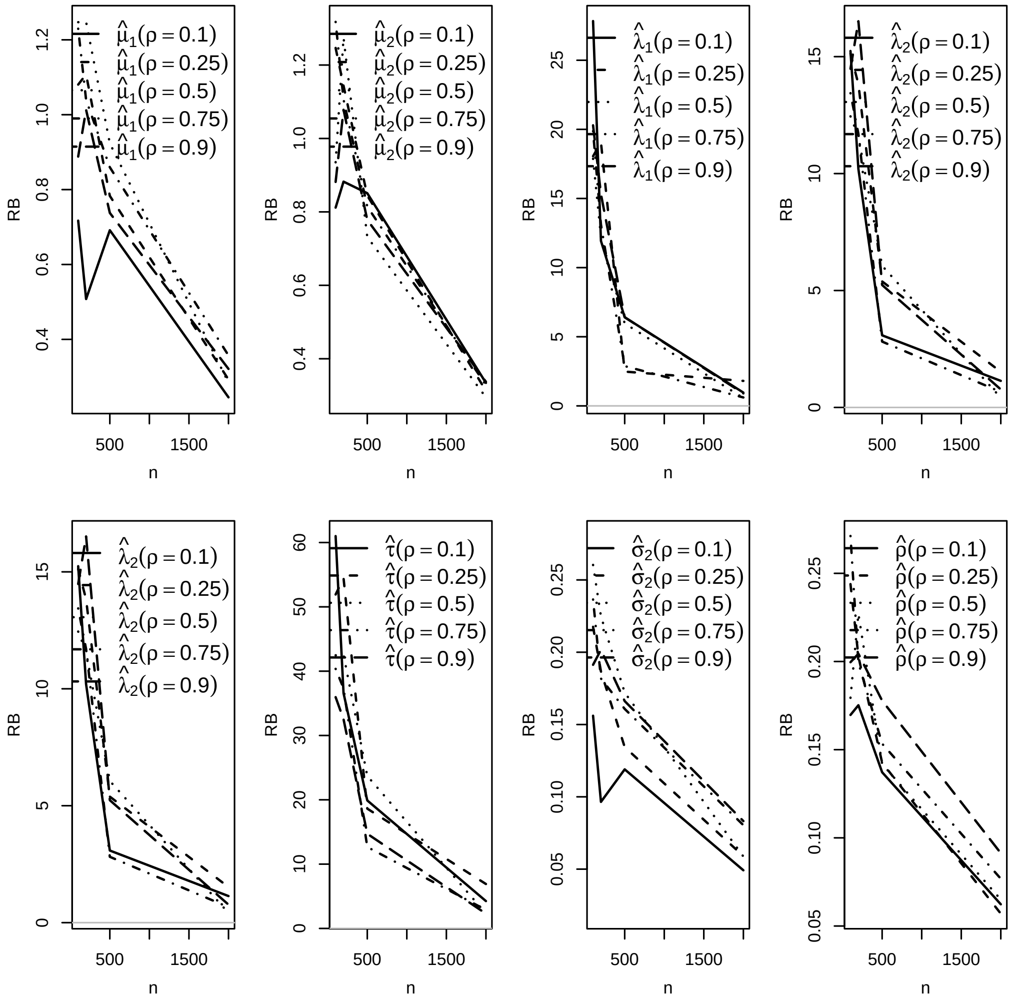


Figure 4.3: Relative bias for $G_i^{-1}(x) = \left[\frac{\exp(x)}{1+\exp(x)}\right]^{\frac{1}{3}}$.

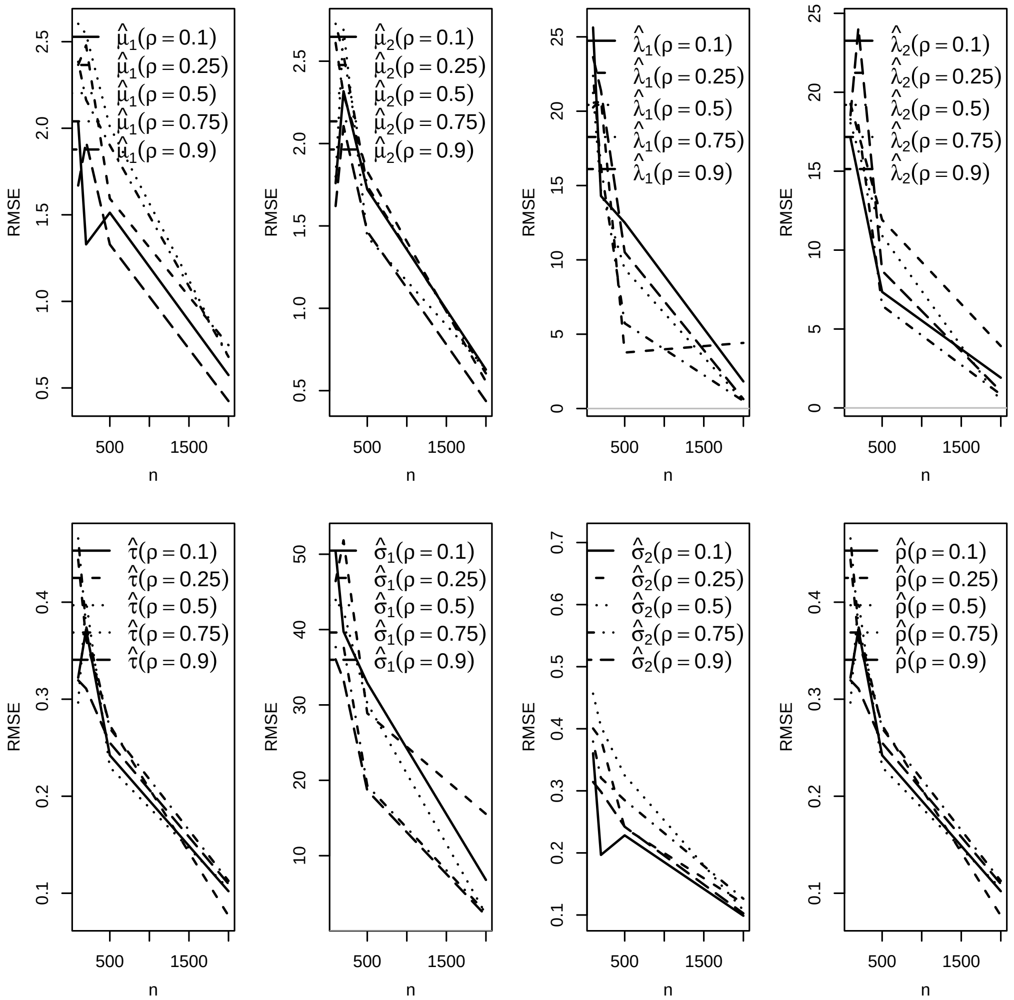


Figure 4.4: Root mean squared error for $G_i^{-1}(x) = \left[\frac{\exp(x)}{1+\exp(x)}\right]^{\frac{1}{3}}$.

It is important to note that, when $\tau = 0$ the model is simplified, allowing estimates to be obtained that characterize a particular case. This simplified model can be useful in situations where the generalization parameter does not need to be considered. Simulation studies were also carried out to evaluate the quality of parameter estimation in this context. The results of these simulations are attached to this work.

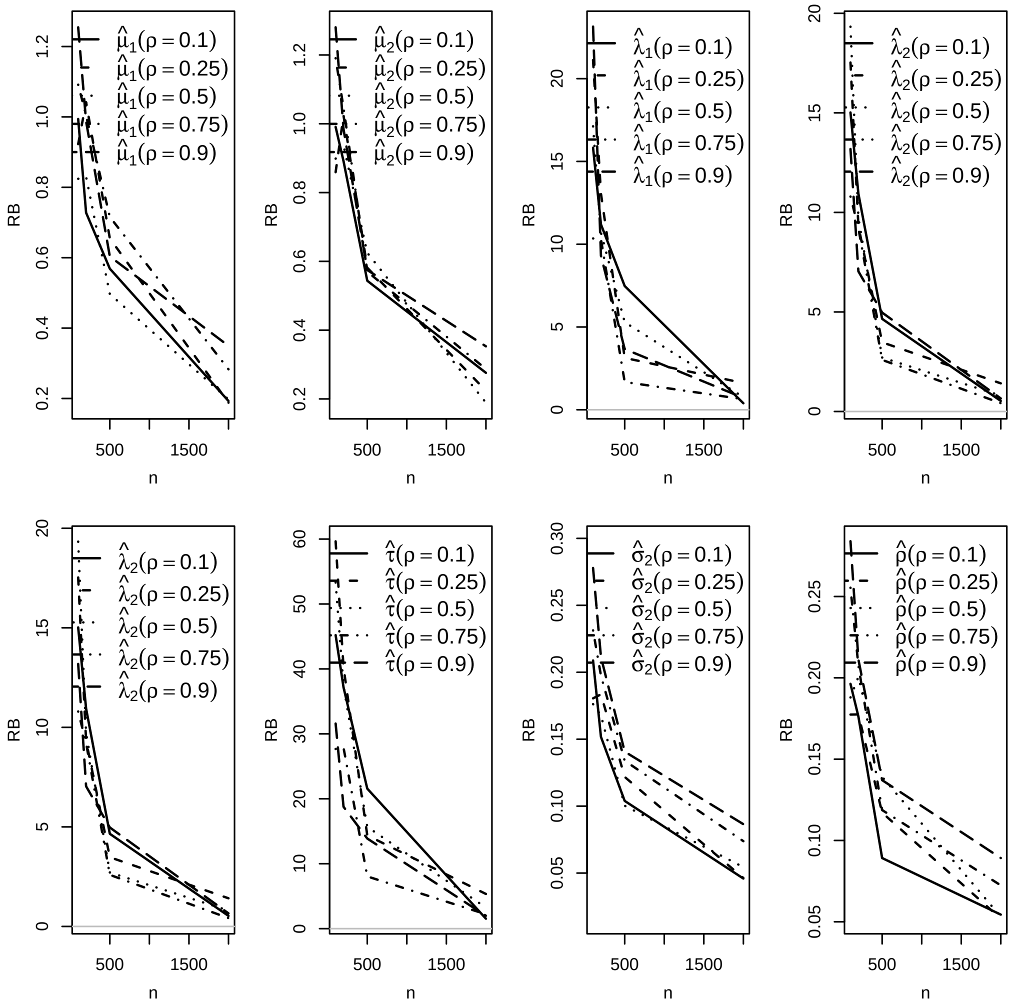


Figure 4.5: Relative bias for $G_i^{-1}(x) = \left[\frac{\exp(x)}{1+\exp(x)}\right]^{\frac{1}{5}}$.

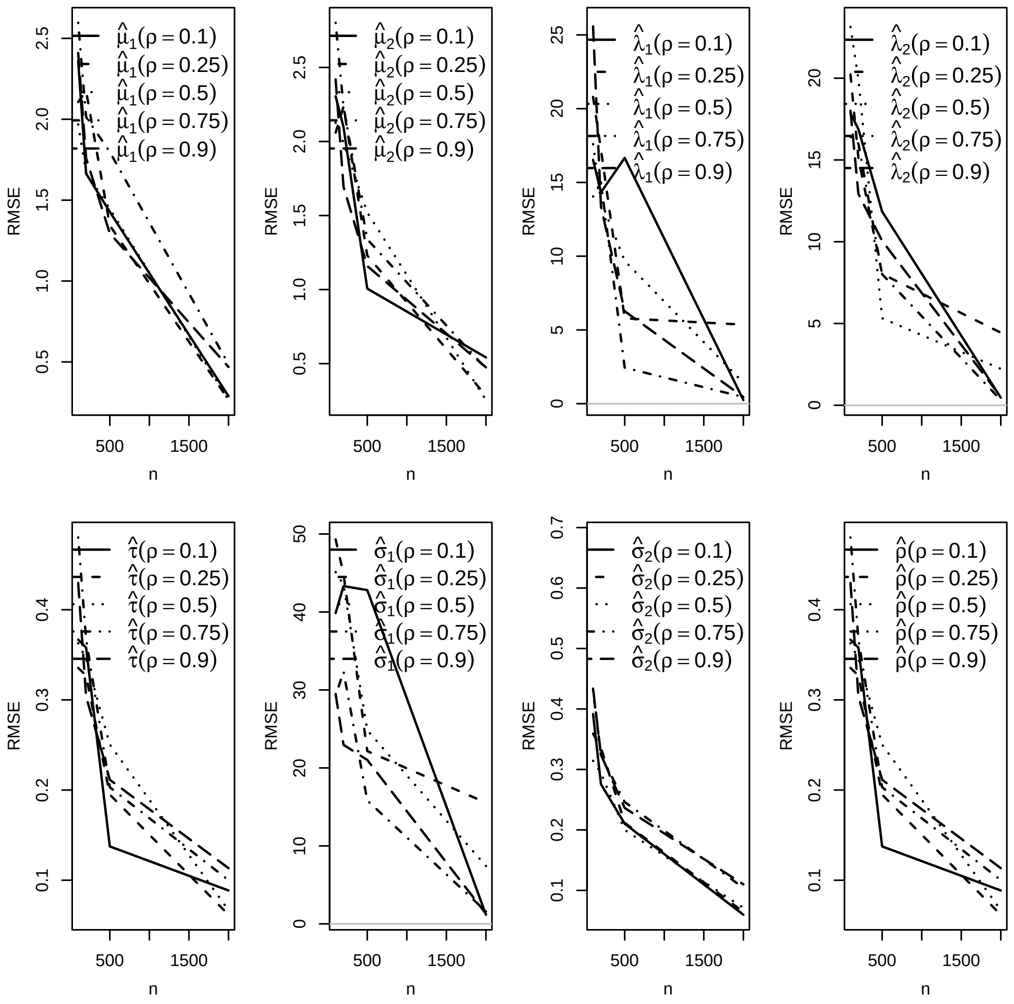


Figure 4.6: Root mean squared error for $G_i^{-1}(x) = \left[\frac{\exp(x)}{1+\exp(x)}\right]^{\frac{1}{5}}$.

4.2 Application to real data

In this section, we illustrate the proposed model and the inferential method using real data on fertility and socioeconomic indicators for each of Switzerland's 47 French-speaking provinces in 1888. This data set is called *swiss* and is available in the R software. The aim of the study was to explore the relationships between fertility (measured as the birth rate) and several other socioeconomic variables in 47 districts. The variables contained in the dataset are:

- *Fertility*: Fertility rate (average number of births per 1000 women).
- *Agriculture*: Percentage of men involved in agricultural activities.
- *Examination*: Percentage of military draftees *draftees* who received a high score on aptitude exams.
- *Education*: Percentage of men with education beyond primary education.
- *Catholic*: Percentage of Catholics (as a measure of religion and tradition).
- *Infant.Mortality*: Infant mortality rate (number of baby deaths per 1000 live births).

For the application presented here, the variables Education and Agriculture were considered. The data can be found at Swiss Fertility and Socioeconomic Indicators (1888).

Table 4.1 presents the descriptive statistics of the two variables: Education and Agriculture, both with a set of 47 observations. For the Education variable, it is observed that the minimum value recorded is 0.010, while the maximum reaches 0.530, with a median of 0.080 and an average of 0.1098. The dispersion of the Education data is reflected by the standard deviation (SD) of 0.0962, which suggests considerable variation in relation to the mean. This is further evidenced by the coefficient of variation (CV) of 87.5822, indicating a high relative variability of the data. Positive skewness, with a skewness coefficient (CS) of 2.3428, suggests that the data distribution is skewed to the right, which is reinforced by the kurtosis coefficient (CK) of 6.5414, indicating a more elongated distribution with heavy tails. Considering the Agriculture variable, the minimum value is 0.012 and the maximum is 0.897, with a median of 0.541, very close to the average of 0.5066, which suggests a more balanced distribution. The standard deviation is higher, 0.2271, reflecting greater data dispersion compared to Education. The coefficient of variation is 44.8311, less high than that of Education, suggesting less relative variability. The Agriculture distribution presents negative skewness, with an asymmetry coefficient of -0.3309,

Variables	n	Minimum	Median	Mean	Maximum	SD	CV	CS	CK
Education	47	0.010	0.08	0.11	0.53	0.096	87.58	2.33	6.54
Agriculture	47	0.012	0.54	0.51	0.90	0.230	44.83	-0.33	-0.79

Table 4.1: Summary statistics.

indicating a slight leftward bias. The negative kurtosis coefficient (-0.7926) suggests a flatter distribution with lighter tails, in contrast to the more elongated distribution of Education.

The extended unit-G-skew-normal and extended unit-G-skew-Student-t distributions were used to fit the data. The model parameters were estimated according to the methodology presented in Section 3.3 – for simplification purposes τ was set to zero. The estimation of the ν parameter of the extended unit-G-skew-Student-t distribution was carried out by using the profile likelihood method. First, an initial grid of values was defined for $\nu \in \{1, 2, \ldots, 50\}$, then for each fixed value of ν it is computed the maximum likelihood estimates of the remaining parameters and also the log-likelihood function. The final estimate of ν is the one that maximizes the log-likelihood function and the associated estimates of the remaining parameters are then the final ones (see Saulo et al., 2021).

Tables 4.2-4.5 report the Kolmogorov-Smirnov (KS) and Anderson-Darling (AD) tests, the maximum likelihood estimates, and the standard errors for the extended unit-G-skew-normal and extended unit-G-skew-Student-t distributions. Moreover, Figures 4.7-4.22 display the quantile versus quantile (QQ) plots of the randomized quantile (Saulo et al., 2022b) residuals for these models. From these results, we observe that the extended unit-G-skew-normal model provides better adjustment compared to the unit-G-skew-Student-t model. Note that the results of the QQ plots indicate that $G_i(x) = \log(\frac{x}{1-x})$ shows better agreement with the expected standard normal distribution; note also that the p-values of the KS and AD tests favor the extended unit-G-skew-normal with $G_i(x) = \log(\frac{x}{1-x})$.

Table 4.2: KS and AD test results.

Extended unit-G-skew-Student-t		
$G_i(x)$	p-value.KS	p-value.AD
$\tan\left(\pi\left(x-\frac{1}{2}\right)\right)$	0.18	0.08
$\log\left(\frac{x^3}{1-x^3}\right)$	0.18	0.07
$\log\left(\frac{x^5}{1-x^5}\right)$	0.18	0.02
$\log(-\log(1-x))$	0.17	0.03
$\frac{2}{\pi}\ln\left(\tan(\frac{\pi}{2}x)\right)$	0.18	0.03
$1-\log(-\log(x))$	0.18	0.04
$\log\left(\log\left(\frac{1}{-x+1}\right)+1\right)$	0.00	0.00
$\log\left(\frac{x}{1-x}\right)$	0.16	0.03

Table 4.3: KS and AD test results.

Extended unit-G-skew-normal		
$G_i(x)$	p-value.KS	p-value.AD
$\tan\left(\pi\left(x-\frac{1}{2}\right)\right)$	0.03	0.01
$\log\left(\frac{x^3}{1-x^3}\right)$	0.23	0.03
$\log\left(\frac{x^5}{1-x^5}\right)$	0.23	0.04
$\log(-\log(1-x))$	0.35	0.03
$\frac{2}{\pi}\ln\left(\tan(\frac{\pi}{2}x)\right)$	0.35	0.06
$1-\log(-\log(x))$	0.35	0.06
$\log\left(\log\left(\frac{1}{-x+1}\right)+1\right)$	0.00	0.00
$\log(\frac{x}{1-x})$	0.35	0.05

Table 4.4: Parameters estimates (with standard errors in parentheses).

Extended unit-G-skew-Student-t								
$G_i(x)$	$\hat{\mu}_1$	$\hat{\mu}_2$	$\hat{\lambda}_1$	$\hat{\lambda}_2$	$\hat{\sigma}_1$	$\hat{\sigma}_2$	$\hat{\rho}$	$\hat{\nu}$
$\tan(\pi(x-\frac{1}{2}))$	-1.63	-0.06	-2.23	-2.72	3.77	0.85	-0.31	2
	(0.41)	(0.27)	(0.94)	(1.57)	(0.87)	(0.14)	(0.30)	-
$\log\left(\frac{x^3}{1-x^3}\right)$	-4.68	-4.10	-0.65	-0.10	4.53	4.01	-0.88	31
	(1.04)	(1.67)	(0.29)	(0.28)	(1.96)	(2.31)	(0.13)	-
$\log\left(\frac{x^5}{1-x^5}\right)$	-5.21	-8.19	-0.92	-0.20	10.45	6.89	-0.92	16
	(1.56)	(1.85)	(0.87)	(0.20)	(3.28)	(2.84)	(0.06)	-
$\log(-\log(1-x))$	-1.46	-0.61	-5.51	-3.22	1.34	0.90	-0.49	46
	(0.22)	(0.39)	(3.05)	(1.61)	(0.11)	(0.05)	(0.28)	-
$\frac{2}{\pi}\ln(\tan\frac{\pi}{2}x))$	-1.14	0.41	-7.45	-7.27	0.48	0.68	-0.38	11
	(0.23)	(0.27)	(5.42)	(4.37)	(0.04)	(0.14)	(0.13)	-
$1-\log(-\log(x))$	0.08	1.52	0.39	-0.08	0.32	0.71	-0.67	15
	(0.25)	(0.51)	(3.46)	(1.91)	(0.03)	(0.08)	(0.10)	-
$\log(\log(\frac{1}{-x+1})+1)$	0.04	0.93	0.73	0.19	-0.10	0.46	0.76	23
	(0.02)	(0.06)	(2.25)	(0.32)	(0.01)	(0.01)	(0.03)	-
$\log(\frac{x}{1-x})$	-3.12	1.20	0.26	-1.06	1.18	1.72	-0.84	24
	(0.40)	(0.34)	(1.15)	(0.91)	(0.34)	(0.37)	(0.10)	-

Table 4.5: Parameters estimates (with standard errors in parentheses).

Extended unit-G-skew-normal							
$G_i(x)$	$\hat{\mu}_1$	$\hat{\mu}_2$	$\hat{\lambda}_1$	$\hat{\lambda}_2$	$\hat{\sigma}_1$	$\hat{\sigma}_2$	$\hat{\rho}$
$\tan(\pi(x-\frac{1}{2}))$	-1.26	0.32	-2.75	-3.02	6.67	3.75	-0.14
	(0.39)	(0.51)	(2.41)	(3.80)	(0.63)	(0.35)	(0.14)
$\log\left(\frac{x^3}{1-x^3}\right)$	-3.88	-4.36	-1.12	-0.42	6.18	4.90	-0.91
	(0.12)	(0.69)	(0.44)	(0.27)	(1.47)	(1.57)	(0.06)
$\log\left(\frac{x^5}{1-x^5}\right)$	-5.40	-6.97	-2.02	-0.43	9.66	4.76	-0.78
	(0.52)	(0.96)	(1.75)	(0.33)	(1.51)	(0.78)	(0.10)
$\log(-\log(1-x))$	-2.59	0.14	-0.62	-1.57	0.79	1.08	-0.58
	(0.70)	(1.27)	(0.90)	(3.60)	(0.07)	(0.65)	(0.06)
$\frac{2}{\pi}\ln(\tan\frac{\pi}{2}x))$	-1.52	0.64	-1.93	-3.45	0.57	0.95	-0.75
	(0.20)	(0.14)	(2.58)	(2.07)	(0.10)	(0.14)	(0.14)
$1-\log(-\log(x))$	0.34	1.01	-0.75	0.58	0.42	0.91	-0.78
	(0.12)	(0.49)	(1.57)	(1.61)	(0.07)	(0.23)	(0.02)
$\log(\log(\frac{1}{-x+1})+1)$	0.06	0.93	-0.23	0.30	-0.17	0.87	0.88
	(0.15)	(0.79)	(1.72)	(5.11)	(0.52)	(3.58)	(0.80)
$\log(\frac{x}{1-x})$	-2.36	0.02	-0.14	-0.12	0.89	1.21	-0.71
	(1.05)	(1.02)	(3.02)	(1.89)	(0.09)	(0.12)	(0.02)

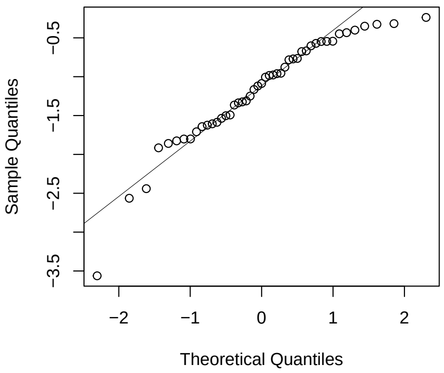


Figure 4.7: Extended unit-G-skew-Student-t with $G_i(x) = \tan\left((x - \frac{1}{2})\pi\right)$.

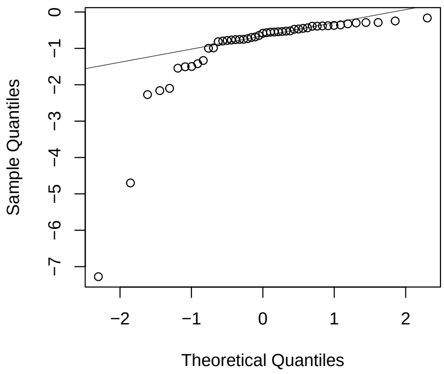


Figure 4.8: Extended unit-G-skew-normal with $G_i(x) = \tan\left((x - \frac{1}{2})\pi\right)$.

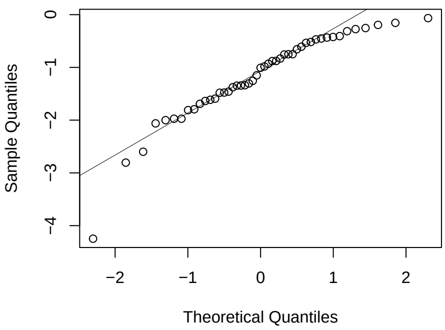


Figure 4.9: Extended unit-G-skew-Student-t with $G_i(x) = \log\left(\frac{x^3}{1-x^3}\right)$.

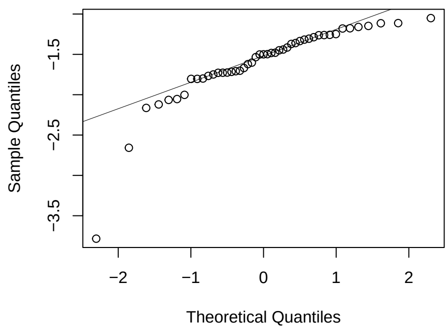


Figure 4.10: Extended unit-G-skew-normal with $G_i(x) = \log\left(\frac{x^3}{1-x^3}\right)$.

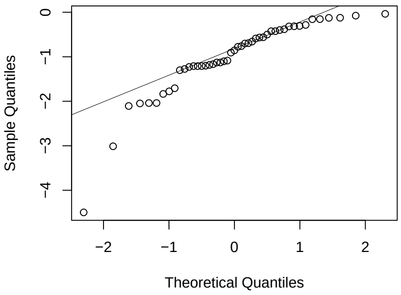


Figure 4.11: Extended unit-G-skew-Student-t with $G_i(x) = \log\left(\frac{x^5}{1-x^5}\right)$.

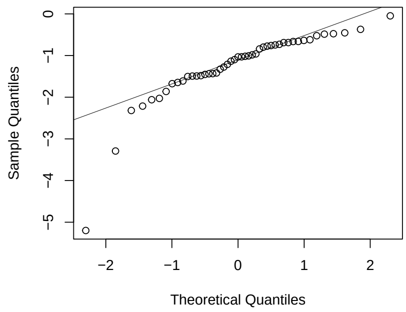


Figure 4.12: Extended unit-G-skew-normal with $G_i(x) = \log\left(\frac{x^5}{1-x^5}\right)$.

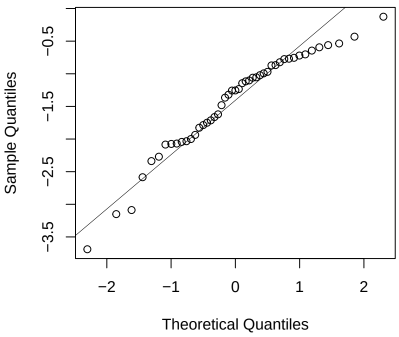


Figure 4.13: Extended unit-G-skew-Student-t with $G_i(x) = \log(\log(\frac{1}{-x+1}) + 1)$.

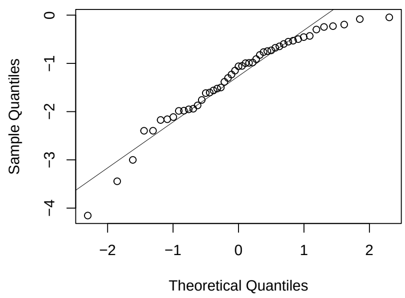


Figure 4.14: Extended unit-G-skew-normal with $G_i(x) = \log(\log(\frac{1}{-x+1}) + 1)$.

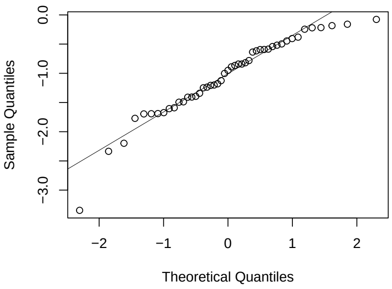


Figure 4.15: Extended unit-G-skew-Student-t with $G_i(x) = \frac{2}{\pi}\ln(\tan\frac{\pi}{2}x))$.

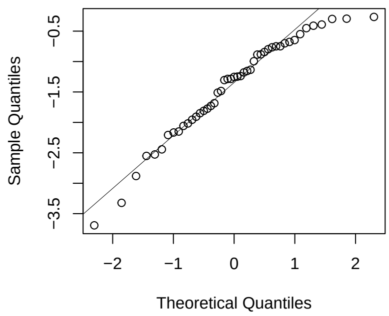


Figure 4.16: Extended unit-G-skew-normal with $G_i(x) = \frac{2}{\pi}\ln(\tan\frac{\pi}{2}x))$.

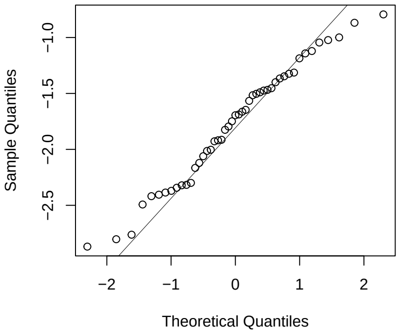


Figure 4.17: Extended unit-G-skew-Student-t with $G_i(x) = 1 - \log(-\log(x))$.

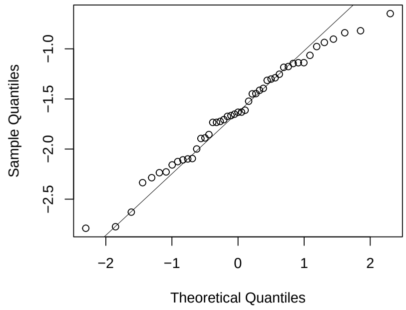


Figure 4.18: Extended unit - G - skew - normal with $G_i(x) = 1 - \log(-\log(x))$.

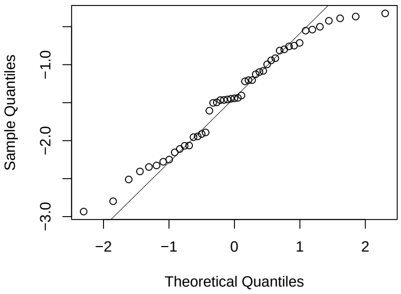


Figure 4.19: Extended unit-G-skew-Student-t with $G_i(x) = \log(-\log(1-x))$.

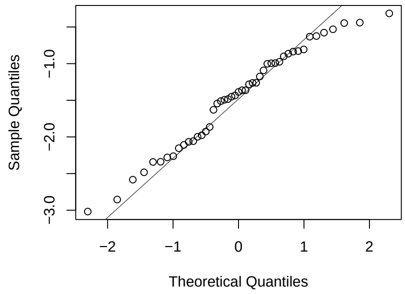


Figure 4.20: Extended unit-G-skew-normal with $G_i(x) = \log(-\log(1-x))$.

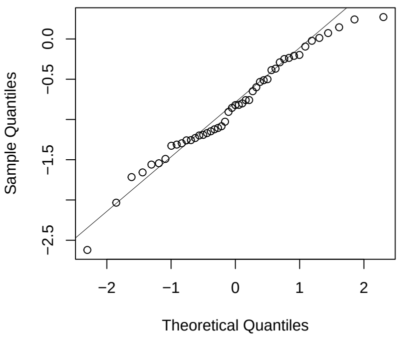


Figure 4.21: Extended unit-G-skew-Student-t with $G_i(x) = \log(\frac{x}{1-x})$.

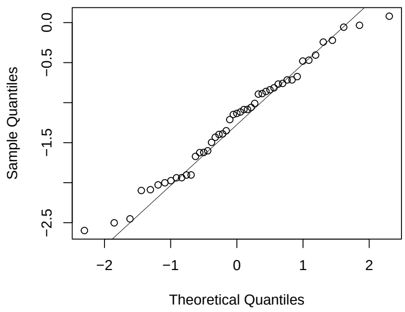


Figure 4.22: Extended unit-G-skew-normal with $G_i(x) = \log(\frac{x}{1-x})$.

Considering that the function $G_i(x) = \log(\frac{x}{1-x})$ based on the normal model presented the best results in terms of fit, one now can compute the estimated CDF given the observed data: see Table 4.6. From this table, it is observed an inverse relationship between the percentage of men with education beyond primary level and involvement in agriculture. Regions with a high percentage of men involved in agriculture tend to have a low percentage of education and vice versa. This is evident in the high values of **Agriculture**, which generally correspond to low values of **Education** (Line 7: 7% education, 70.2% agriculture; Line 45: 53% education, 1.2% agriculture). There are some cases in the data that have a high percentage of education, such as row 45, which has 53% education and one of the lowest values of men in agriculture (1.2%). These cases may be related to more urbanized regions or with broader access to education. Several observations have the percentage of agriculture above 60%, suggesting that a significant portion of the population still depends on the agricultural sector, despite there being variations in education rates.

Table 4.6: Estimated probabilities.

Education	**Agriculture**	CDF_values
0.12	0.17	0.01
0.09	0.45	0.11
0.05	0.40	0.02
0.07	0.36	0.03
0.15	0.43	0.23
0.07	0.35	0.03
0.07	0.70	0.22
0.08	0.68	0.25
0.07	0.53	0.10
0.13	0.45	0.21
0.06	0.64	0.13
0.12	0.62	0.35
0.07	0.68	0.20
0.12	0.61	0.34
0.05	0.69	0.11
0.02	0.73	0.01
0.08	0.34	0.04
0.28	0.19	0.09
0.20	0.15	0.03
0.09	0.73	0.34
0.10	0.60	0.26
0.03	0.55	0.01
0.12	0.51	0.24

Education	**Agriculture**	CDF_values
0.06	0.54	0.08
0.01	0.71	0.00
0.08	0.58	0.17
0.03	0.64	0.02
0.10	0.61	0.27
0.19	0.27	0.11
0.08	0.49	0.11
0.02	0.86	0.02
0.06	0.85	0.27
0.02	0.90	0.03
0.06	0.78	0.22
0.03	0.65	0.02
0.09	0.76	0.37
0.03	0.85	0.07
0.13	0.63	0.40
0.12	0.38	0.13
0.11	0.08	0.00
0.13	0.17	0.02
0.32	0.18	0.08
0.07	0.38	0.04
0.07	0.19	0.00
0.53	0.01	0.00
0.29	0.47	0.42
0.29	0.28	0.18

Concluding Remarks

Based on the studies developed in this work, it was possible to observe that the model, despite having a considerable level of complexity, also presents a high degree of versatility. Depending on the configuration of its parameters, the model can take other forms, which have already been discussed in previous publications. The model introduces innovations by considering transformations or generic functions in the density function, in addition to adding a generalization parameter, which can significantly influence the data fit.

Invertible and differentiable functions with domains in the open interval $(0, 1)$ proved to be fundamental for defining relevant properties, such as marginal quantiles and moments. Through these functions, it was possible to observe, for example, that the moments established for the density function are necessarily finite, which could be different when considering functions with different domains. It is important to highlight that a way of defining the functions that can be used was presented, which is based on the consideration of functions such as the accumulated functions of continuous probability distributions, where the corresponding derivative is a known density function and its inverse can be obtained through algebraic manipulations. Furthermore, from the deduction of relevant mathematical properties, it was possible to observe that the model contributes significantly to the theory of probability distributions, and can be fundamentally used in the treatment of asymmetric data. Now considering the computational aspects, it was possible to estimate parameters using the maximum likelihood function, jointly employing the Monte Carlo algorithm, through which all model parameters were estimated. From the model presented in this work, it was also possible, through computational implementation, to observe that, as the sample size increased and the interactions in the Monte Carlo algorithm were carried out, the relative bias and the mean squared error decreased. significantly, showing convergence towards zero. This indicates a reliable fit of the estimated parameters with respect to the original parameter values. Still through computation, it was possible to apply the model to real data, in which the data set was used in two distribution functions. The application has favored the use of the extended unit-G-skew-normal model over the unit-G-skew-Student-t model.

Appendix A

Shape of distribuitions

Next, we will present the graphs of the density functions for the distributions of Extended unit-G-skew-Cauchy and Extended unit-G-skew-Student-t. For both graphs, we used the same parameters in data simulation, with a degree of freedom value $\nu = 10$ in Extended unit-G-skew-Student-t. The values of the reinforced partnerships were: $\boldsymbol{\mu} = (2, 3)^\top$; $\boldsymbol{\lambda} = (0.5, 0.6)^\top$; $\sigma_1 = 1$; $\sigma_2 = 1$; $\rho = 0.5$; $\tau = 10$

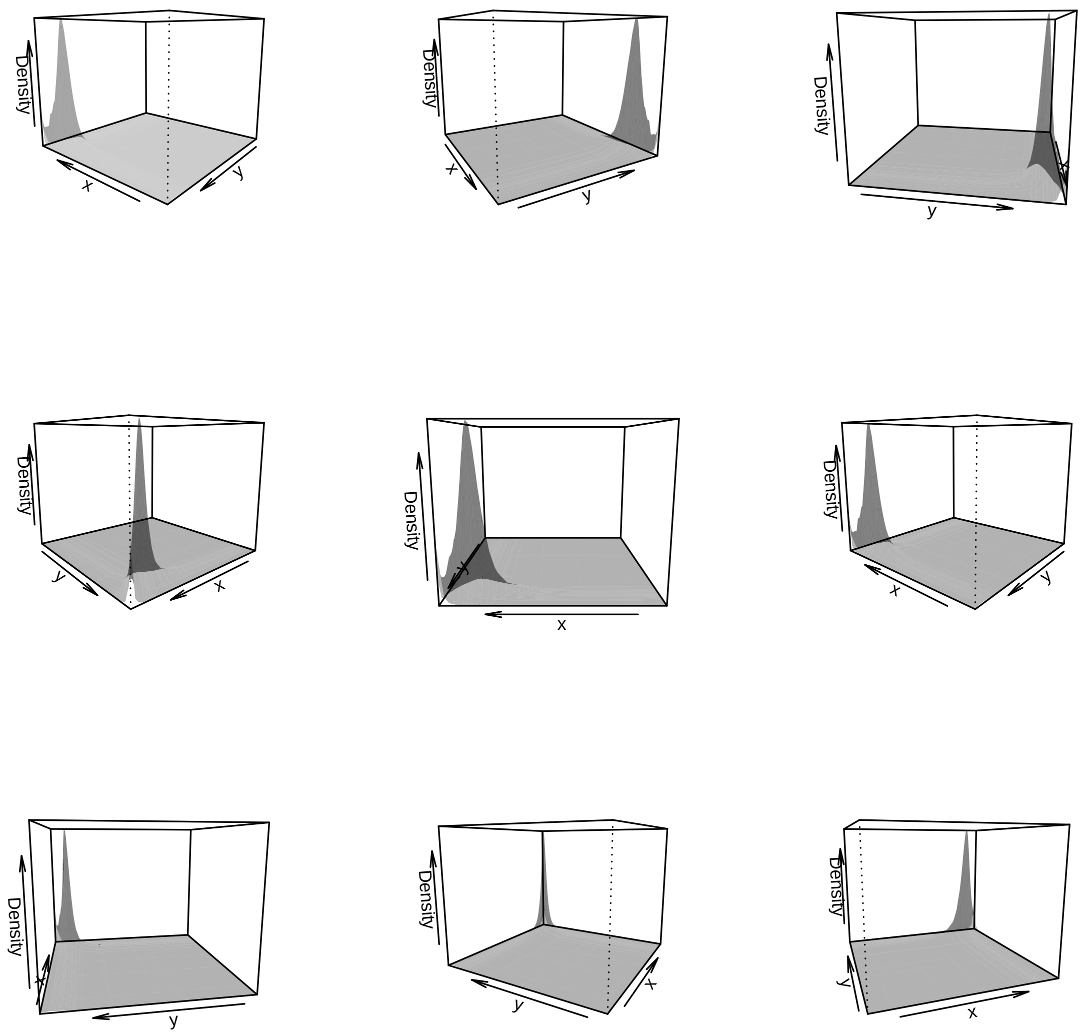


Figure A.1: Extended unit-G-skew-Cauchy density function with $G_i^{-1}(x) = (1/2) + \arctan(x)/\pi$.

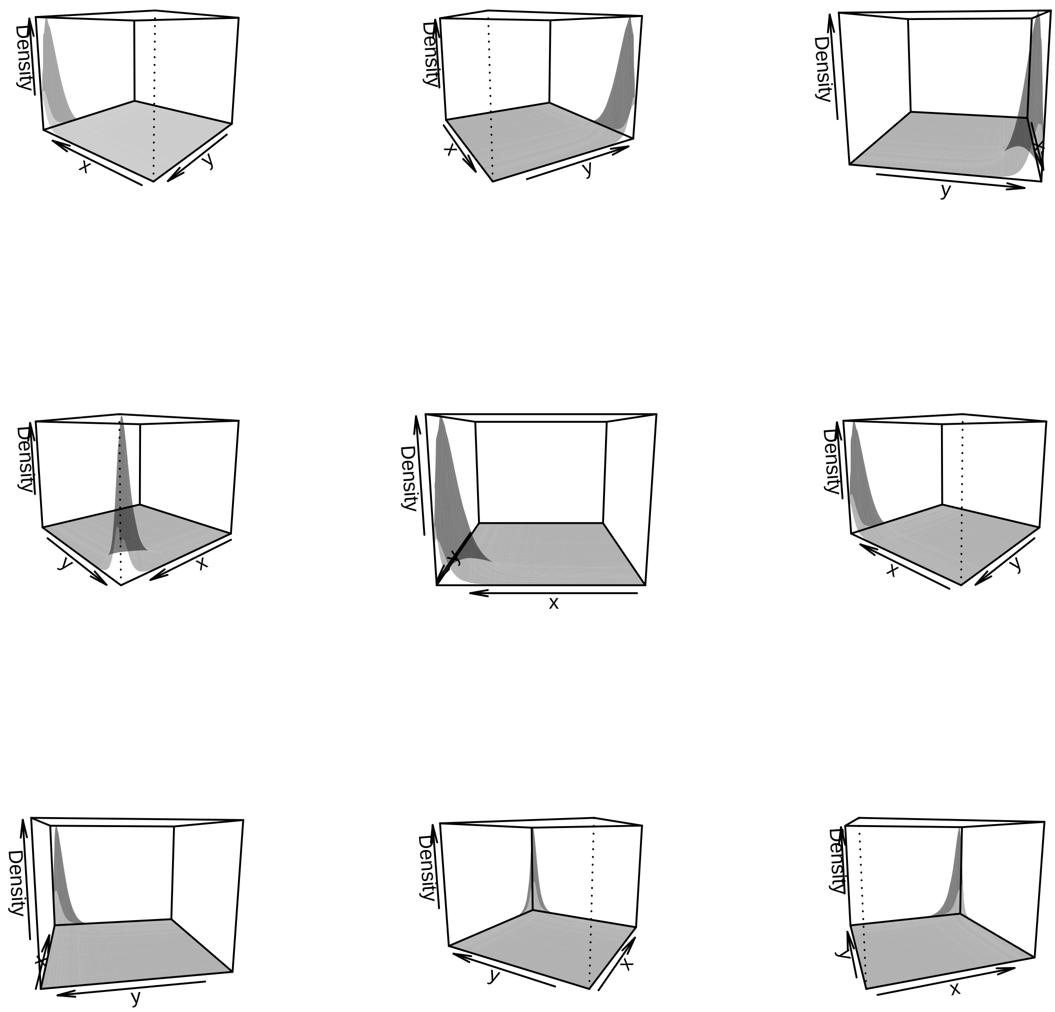


Figure A.2: Extended unit-G-skew-Student-t density function with $G_i^{-1}(x) = (1/2) + \arctan(x)/\pi$.

Appendix B

Monte Carlo simulation

This appendix presents the results of parameter estimation considering $\tau = 0$. When the parameter τ takes this value, the configuration of the density function of the Extended unit-G-skew distribution is simplified, resulting in a more direct expression. Depending on the choice of the G_i function, this density function may coincide with an already known expression. A specific example occurs when $\boldsymbol{\lambda} = \mathbf{0}, \tau = 0$, $G_1(x) = G_2(x) = \log(-\log(1-x)), 0 < x < 1$, and $n = 2$, we obtain the bivariate unit model studied in reference (Vila et al., 2023b). For this case, we only consider the fixed parameter τ, and perform iterations varying the number of samples in a range of 100 to 1000.

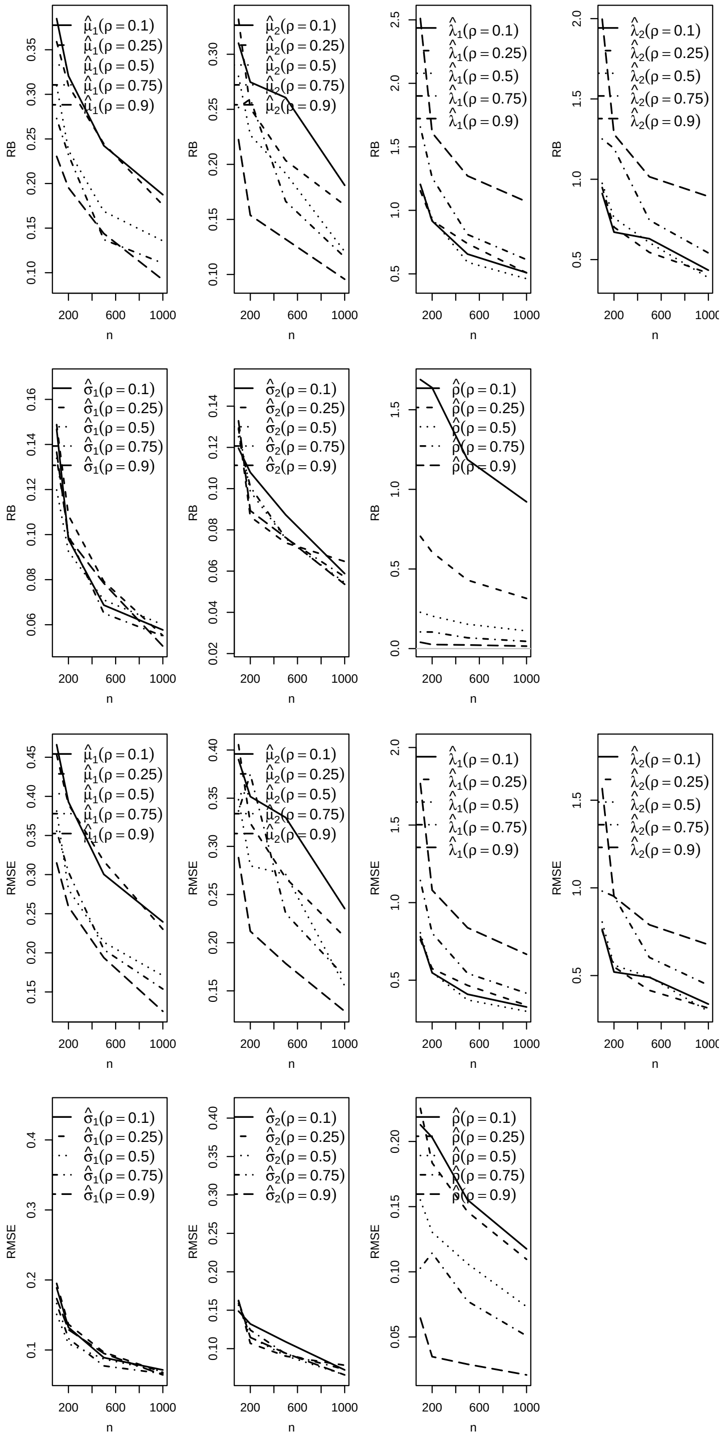


Figure B.1: Relative bias for and root mean squared error for $G_i^{-1}(x) = \frac{1}{2} + \frac{\arctan(x)}{\pi}$.

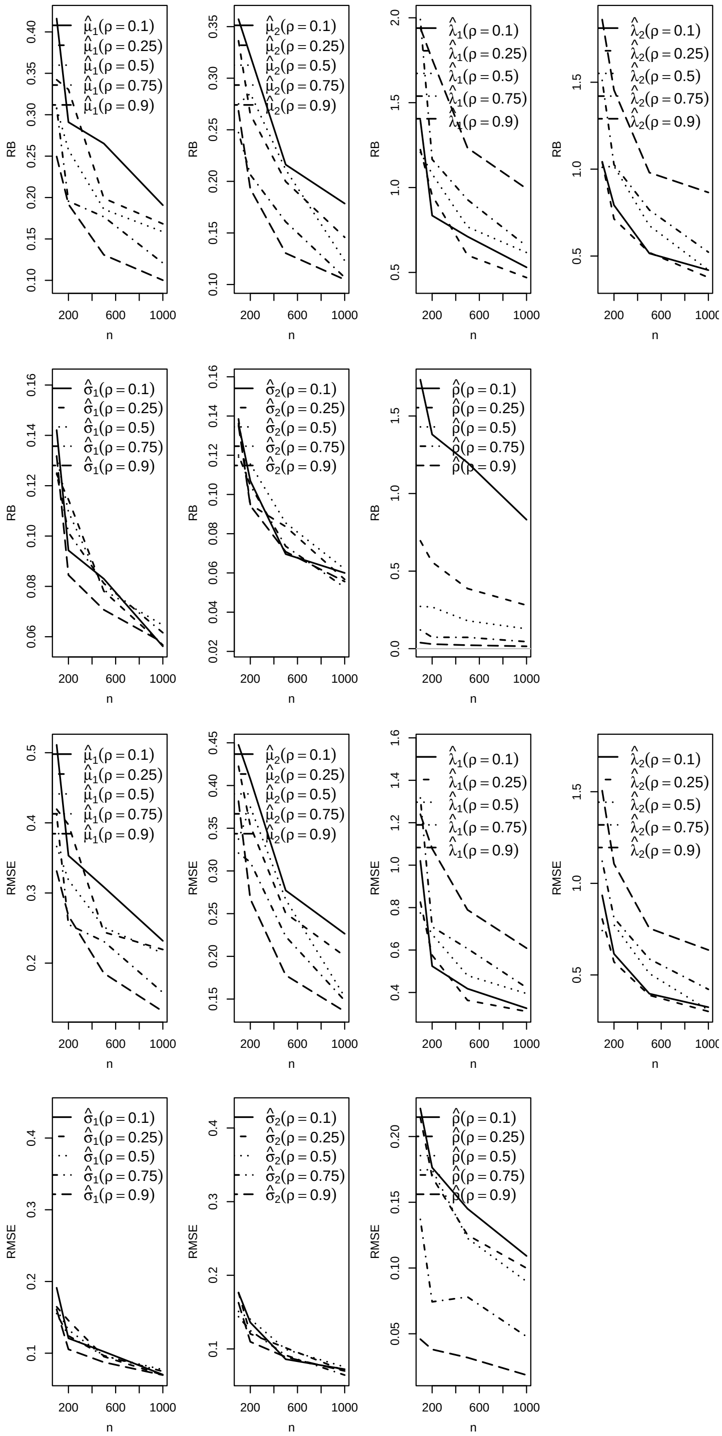


Figure B.2: Relative bias for and root mean squared error for $G_i^{-1}(x) = \left[\frac{\exp(x)}{1+\exp(x)}\right]$.

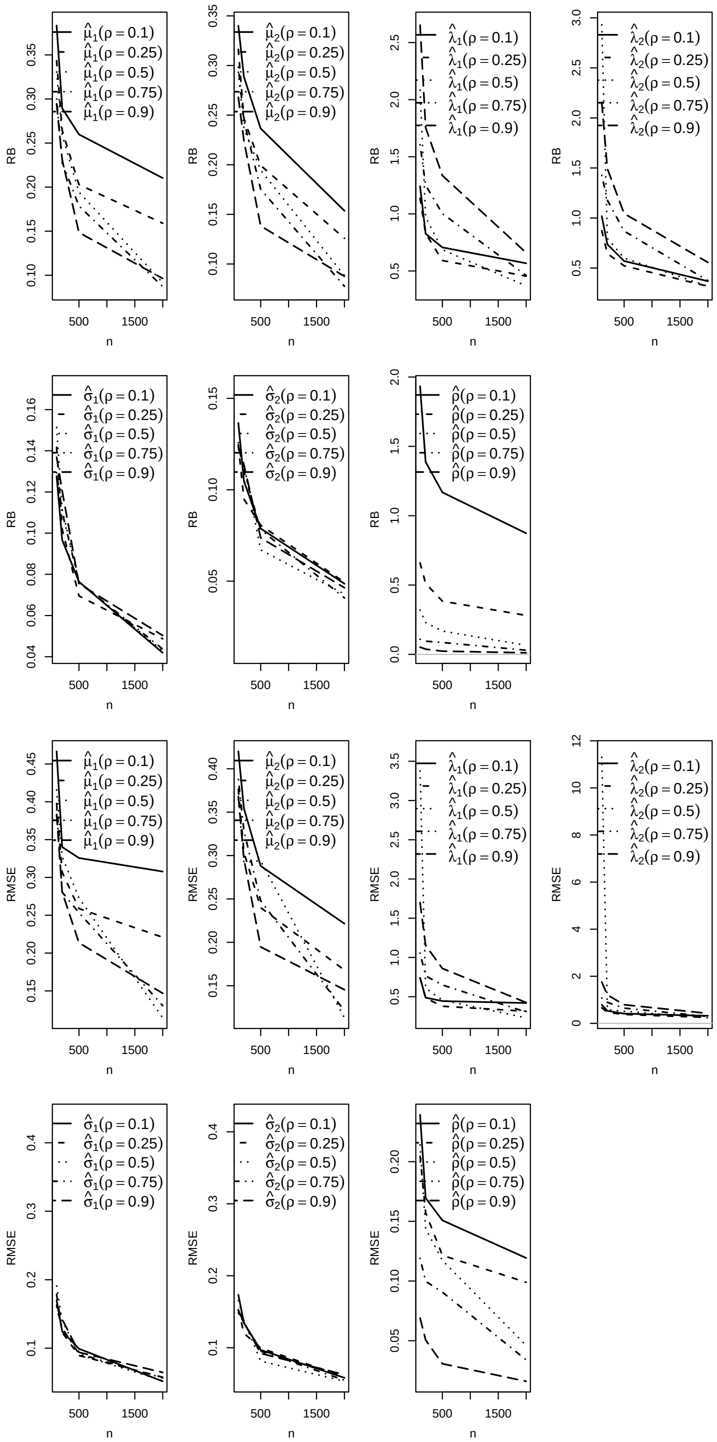


Figure B.3: Relative bias for and root mean squared error for $G_i^{-1}(x) = \left[\frac{\exp(x)}{1+\exp(x)}\right]^{\frac{1}{3}}$.

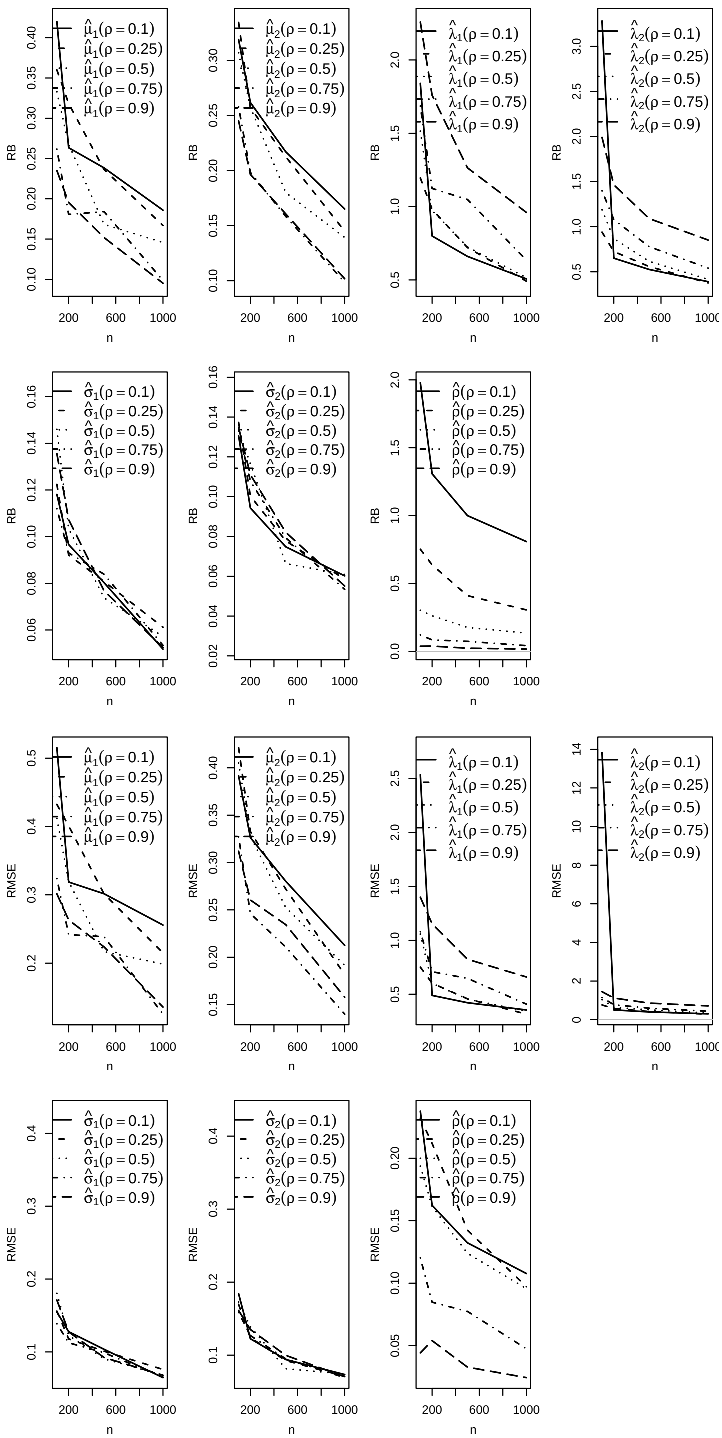


Figure B.4: Relative bias for and root mean squared error for $G_i^{-1}(x) = \frac{x-2}{2x} + \frac{\sqrt{x^2+4}}{2x}$.

References

Arellano-Valle, R.B. and Genton, M.G. (2010). “Multivariate extended skew-t distributions and related families”. *METRON* 68, pp. 201–234.

Armerin, F. (2017). *On some properties of elliptical distributions*.

Azzalini, A. (1985). “A Class of Distributions Which Includes the Normal Ones”. *Scandinavian Journal of Statistics* 12.2, pp. 171–178. URL: http://www.jstor.org/stable/4615982.

Azzalini, A. and Capitanio, A. (1999). “Statistical Applications of the Multivariate Skew Normal Distribution”. *Journal of the Royal Statistical Society: Series B* 61, pp. 579–602. DOI: 10.1111/1467-9868.00194.

Azzalini, A. and Dalla Valle, A. (1996). “The Multivariate Skew-Normal Distribution”. *Biometrika* 83.4, pp. 715–726. URL: http://www.jstor.org/stable/2337278.

Balakrishnan, N. and Lai, C. D. (2009). *Continuous Bivariate Distributions*. New York, NY: Springer-Verlag.

Barry, C. A. and Robert, J. B. (2000). “The skew-Cauchy distribution”. *Statistics Probability Letters* 49.3, pp. 285–290. ISSN: 0167-7152. DOI: https://doi.org/10.1016/S0167-7152(00)00059-6. URL: https://www.sciencedirect.com/science/article/pii/S0167715200000596.

Bingham, N.H. and Kiesel, R. (2002). “Semi-parametric modelling in finance: theoretical foundation”. *Quantitative Finance* 2, pp. 241–250.

Bolfarine, H. and Sandoval, M. C. (2010). *Introdução a Inferência Estatística*. Acesso em: 23 ago. 2024. Rio de Janeiro: SBM.

Branco, M.D. and Dey, D.K. (2001). “A General Class of Multivariate Skew-Elliptical Distributions”. *Journal of Multivariate Analysis* 79, pp. 99–113.

Castro, L. M., San Martín, E., and Arellano-Valle, R. B. (2013). “A note on the parameterization of multivariate skewed-normal distributions”. *Brazilian Journal of Probability and Statistics* 27, pp. 110–115.

Choy, S. T. B. (1995). “Robust Bayesian Analysis Using Scale Mixture of Normals Distributions”. Ph.D. dissertation. London: Department of Mathematics, Imperial College.

Davison, A. C. (2008). *Statistical Models*. Cambridge, England: Cambridge University Press.

Dickey, J. M. and Chen, C. H. (1985). *Direct subjective-probability modelling using ellipsoidal distributions*. Ed. by J. M. Bernardo et al. Elsevier and Valencia University Press.

Díaz-García, J.A. and Vera, J.F. (2011). “Multivariate Three-Parameter Log-Elliptical Distributions”. *Communications in Statistics-Theory and Methods* 40, pp. 1964–1976.

Fang, K. T., Kotz, S., and Ng, K. W. (1990). *Symmetric Multivariate and Related Distributions*. London, UK: Chapman and Hall.

Florens, J. P., Mouchart, M., and Rolin, J. M. (1990). *Elements of Bayesian Statistics*. New York: Marcel Dekker.

Genton, M. G. and Loperfido, N. M. R. (2002). *Generalized Skew-Elliptical Distributions and Their Quadratic Forms*. Tech. rep. 2539. North Carolina State University. Dept. of Statistics. DOI: http://www.lib.ncsu.edu/resolver/1840.4/209. URL: http://www.stat.ncsu.edu/library/papers/skewellrev.ps.

Genton, M.G. and Loperfido, N.M.R. (2005). "Generalized skew-elliptical distributions and their quadratic forms". *Ann Inst Stat Math* 57, pp. 389–401.

Heckman, J.J. (1976). "The common structure of statistical models of truncation, sample selection and limited dependent variables and a simple estimator for such models". *Annals of Economic and Social Measurement* 5, pp. 475–492.

Hill, M. A. and Dixon, W. J. (1982). "Robustness in real life: A study of clinical laboratory data". *Biometrics* 38, pp. 377–396.

Kotz, S. (1975). *Multivariate distributions: Statistical at a cross-road*. Ed. by G.P. Patil, S. Kotz, and J.K. Ord. Vol. 1. D. Reidel Publ. Co.

Kundu, D., Balakrishnan, N., and Jamalizadeh, A. (2013). "Generalized multivariate Birnbaum–Saunders distributions and related inferential issues". *Journal of Multivariate Analysis* 116, pp. 230–244.

Liu, J. and Dey, D. (July 2004). *Skew-Elliptical Distributions*. ISBN: 978-1-58488-431-6. DOI: 10.1201/9780203492000.ch3.

Marchenko, Y.V. and Genton, M.G. (2010). "Multivariate log-skew-elliptical distributions with applications to precipitation data". *Environmetrics* 21, pp. 318–340.

Mathews, John H. and Fink, Kurtis K. (2004). *Numerical Methods Using Matlab*. 4th. Prentice Hall. ISBN: 0-13-065248-2.

Nocedal, Jorge and Wright, Stephen J (1999). *Numerical optimization*. Springer.

Saulo, H. et al. (2021). "A class of asymmetric regression models for left-censored data". *Brazilian Journal of Probability and Statistics* 35.1, pp. 62 –84. DOI: 10.1214/20-BJPS494. URL: https://doi.org/10.1214/20-BJPS494.

Saulo, H. et al. (2022a). "Bivariate symmetric Heckman models and their characterization". *Journal of Multivariate Analysis*, p. 105097.

Saulo, Helton et al. (2022b). "Log-symmetric quantile regression models". *Statistica Neerlandica* 76.2, pp. 124–163.

Vernic, R. (2005). "On the multivariate Skew-Normal distribution and its scale mixtures". *An. Şt. Univ. Ovidius Constanţa* 13, pp. 83–96.

Vila, R. et al. (2023a). "Bivariate log-symmetric models: Distributional properties, parameter estimation and an application to public spending data". *Brazilian Journal of Probability and Statistics* 37.3, pp. 619 –642. DOI: 10.1214/23-BJPS584. URL: https://doi.org/10.1214/23-BJPS584.

Vila, R. et al. (2023b). "Family of bivariate distributions on the unit square: theoretical properties and applications". *Journal of Applied Statistics* 0.0, pp. 1–27. DOI: 10.1080/02664763.2023.2232127.

Vilca, F., Balakrishnan, N., and Zeller, C.B. (2014). "The bivariate Sinh-Elliptical distribution with applications to Birnbaum–Saunders distribution and associated regression and measurement error models". *Computational Statistics & Data Analysis* 80, pp. 1–16.