University of Brasília

Institute of Exact Sciences

Department of Statistics

Dissertation

Family of multivariate distributions for modeling data with positive support: Properties and applications

by

João Victor Monteiros de Andrade

Brasília, 04 October of 2024

Family of multivariate distributions for modeling data with positive support: Properties and applications

by

João Victor Monteiros de Andrade

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, 04 October of 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
Supervisor, EST/UnB

Prof. Dr. JUVÊNCIO SANTOS NOBRE
DEMA/UFC

Prof. Dr. RAUL YUKIHIRO MATSUSHITA
EST/UnB

Earn it.

Acknowledgments

First, I would like to thank God for the gift of life and conscience. I would like to express my deepest gratitude to my parents, Maria and Edilson, for all their teaching, education and unconditional support throughout my career. I would like to express my eternal gratitude to my grandmother Ana and my late grandfathers, Gonçalo (in memoriam) and João (in memoriam), examples of integrity and character. A special thanks to my godmother Oredina (in memoriam), who was like a second mother, gifting me with unparalleled wisdom and affection.

I would also like to express my deep gratitude to my great friend Leonardo Santos, for his loyalty, companionship and constant support during much of my academic journey. I would like to thank Ana Júlia for being there for me at important moments in my life.

I am equally grateful to my colleagues at PPGEST — Melquisedec, Carlo, Estevão, Moisés, Thiago, Rebeca, Mariana and João Pedro — for the valuable exchange of knowledge, partnerships and friendship throughout this journey. To my friends Jonathan Peixoto, Matheus Garcia, Rodolfo, Braian, Ícaro, Daiane, Josyane and Rodrigo, I thank you for every moment shared and for your invaluable support.

I would like to express my eternal gratitude to the professors who were fundamental to my education and intellectual development — Raul, Cira, Felipe, Roberto, Helton Saulo, George, Luan, Thiago and Rathie. I would like to thank Professor Juvêncio, an external member of the committee, for his essential contribution. I would like to give special thanks to the team at the secretariat of the Department of Statistics and to the collaborators who always offered their sincere support, especially Karen Luíza, Tathyanna, Lucas Fernandes, Edenilson and Jader.

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

Resumo Expandido

FAMÍLIA DE DISTRIBUIÇÕES MULTIVARIADAS PARA MODELAGEM DE DADOS COM SUPORTE POSITIVO: PROPRIEDADES E APLICAÇÕES

Este trabalho tem como objetivo o desenvolvimento de uma nova família de distribuições elípticas multivariadas truncadas, ampliando as ferramentas de modelagem estatística para dados multivariados que exibem características como caudas pesadas e assimetria. As distribuições de Cauchy, t-Student e Normal servem de base para a construção dessas novas distribuições, oferecendo diferentes formas de captura de comportamentos e estruturas de dependência entre variáveis. A metodologia de construção envolve o uso de funções monotônicas crescentes, inversíveis e diferenciáveis, definidas no domínio dos números reais positivos, assegurando que as densidades resultantes tenham suporte nos reais positivos. Essa abordagem inovadora permite a preservação de propriedades fundamentais das distribuições elípticas, como a forma das caudas, ao mesmo tempo que incorpora distorções que ajustam a função densidade às características específicas dos dados analisados.

A classe de distribuições proposta oferece maior flexibilidade na modelagem de comportamentos complexos de dados multivariados, sendo particularmente relevante em cenários onde as distribuições tradicionais não capturam adequadamente a dinâmica dos dados. Ao mesmo tempo, mantém a estrutura básica das distribuições elípticas, adaptando-se de maneira robusta às particularidades observadas em fenômenos empíricos. A aplicabilidade desta nova classe de

distribuições é avaliada por meio de simulações de Monte Carlo e da aplicação a dados reais provenientes do Australian Institute of Sport (AIS) e de dados sobre os 50 estados dos Estados Unidos, como expectativa de vida e taxa de homicídios. Esses estudos fornecem uma avaliação prática da flexibilidade e desempenho das distribuições propostas.

Na seção de revisão de literatura e conceitos fundamentais, descrevem-se os conceitos básicos necessários para o desenvolvimento do trabalho. Esses conceitos são essenciais para compreender os tópicos subsequentes, que envolvem modelagem estatística de distribuições multivariadas. A discussão desses fundamentos visa proporcionar ao leitor uma base sólida para o entendimento das propriedades estatísticas e dos métodos empregados na análise e modelagem dos dados.

A seção de desenvolvimento do modelo destaca a criação de um novo modelo de distribuição assimétrica multivariada, baseado em distribuições elípticas e assimétricas. Esse modelo representa uma contribuição inédita no campo da estatística, integrando conceitos previamente explorados na literatura, mas com uma nova abordagem para lidar com a assimetria em distribuições multivariadas. A proposta amplia as possibilidades de análise e modelagem em contextos estatísticos diversos, sendo particularmente útil em aplicações onde as distribuições convencionais falham em capturar adequadamente os padrões dos dados.

No estudo de simulação, conduzimos uma avaliação detalhada do desempenho dos estimadores de máxima verossimilhança (MLEs) no caso bivariado, aplicando diferentes funções de distorção, como $G_i(x) = 1/\alpha \left(\sqrt{x/\beta} - \sqrt{\beta/x}\right)$, $G_i(x) = \log(x)$, $G_i(x) = \cosh^{-1}(x+1)$, $G_i(x) = \log(\log(x+1))$, e $G_i(x) = x - 1/x$. Essas funções permitem ajustar a distribuição para diferentes comportamentos dos dados, sendo analisadas com profundidade. Resultados específicos para a função $\log_q(x)$ estão detalhados no Apêndice.

Por fim, a aplicabilidade dos modelos propostos é avaliada em dados reais, utilizando-se uma série de métricas para verificar a qualidade dos ajustes das distribuições aos dados. Para isso, são aplicados testes estatísticos bem estabelecidos, como os testes de Kolmogorov-Smirnov, Anderson-Darling e Cramér-von Mises, com o objetivo de identificar as distribuições

que oferecem o melhor ajuste aos dados analisados. Esses resultados demonstram a robustez e a flexibilidade das novas distribuições, contribuindo para o avanço das técnicas de modelagem estatística multivariada.

Este trabalho, ao propor uma nova família de distribuições elípticas multivariadas assimétricas, oferece novas perspectivas para a análise estatística de dados complexos, permitindo modelagens mais robustas e adequadas à realidade dos dados. A abordagem inovadora traz avanços significativos na modelagem de dados multivariados, com potencial para aplicações em diversas áreas da ciência e da indústria.

Palavras-chave: Distribuições elípticas multivariadas, Modelagem estatística, Caudas pesadas, Assimetria, Análise de dados multivariados.

Abstract

This work proposes the development of a new family of multivariate truncated elliptical distributions, aiming to expand the possibilities of statistical modeling for multivariate data with specific characteristics, such as heavy tails and asymmetry. The distributions used as a basis for the construction of the new densities are Cauchy, t-Student and Normal, which allow capturing different behaviors and dependence structures between variables. The construction of these new distributions will be performed by applying monotonic increasing, invertible and differentiable functions, defined in the domain of positive real numbers. These characteristics ensure that the resulting density has support in positive real numbers. This new class of distributions offers greater flexibility to model complex behaviors of multivariate data, preserving fundamental characteristics of elliptical distributions, such as the shape of the tails, while incorporating distortions that adjust the density function to the particularities of the studied phenomenon. This development opens new perspectives for statistical analysis, providing more robust models that are appropriate to the reality of the data. A Monte Carlo simulation study will also be carried out and applied to real data from the AIS (Australian Institute of Sport) and information on the 50 states of the United States (life expectancy and homicide rate).

Keywords: Multivariate elliptical distributions, Statistical modeling, Heavy tails, Skewness, Multivariate data analysis.

Contents

List of Tables

List of Figures

List of Symbols and Notations

Γ	Gamma Function
$\boldsymbol{\mu}$	Scale parameter
$\boldsymbol{\Sigma}$	Covariance Matrix
$\boldsymbol{\lambda}$	Skewness parameter vector
τ	Generalizing or Extension Parameter
$g^{(n)}$	Density generator
ϕ_n	Multivariated cumulative normal density function
PDF	Probability density function
ECDF	Empirical Cumulative Distribution function
CDF	Cumulative distribution function
EST_1	Univariate extended skew-Student-t
SF	Survival function
$\mathrm{S}_{\mathrm{ESN}_1}$	Univariate extended skew-Normal Standart

Chapter 1

Introduction

There is partial consensus that the skew-normal distribution was initially introduced by (O'Hagan and Leonard, 1976) in the journal Biometrika. They presented a skewed distribution as a prior distribution for Bayesian estimation. However, the most detailed and elaborate studies on asymmetric distributions and their definitions were exhaustively covered by A. Azzalini, who formalized the skew-normal distribution as we know it today. Azzalini highlighted the need to explore a new class of parametric densities that included some characteristics of the normal distribution, but that could not be modeled by conventional techniques or that avoided complex and detailed mathematics. In (Azzalini, 1985) the author adapted the skewed normal distribution, creating a class that encompasses normality and provided a flexible mathematical framework for future advances by allowing a wide range of variations in the skewness and kurtosis of the distribution.

After these initial advances, other works were published, expanding knowledge on the topic and exploring different variations and applications.

Over time, since the first publication, there has been an expansion in the multivariate field through the works of (Azzalini and Valle, 1996) and (Capitanio, Azzalini, and Stanghellini, 2003), in which other probabilistic and statistical properties were discovered. This further expanded the model's application horizon, giving rise to a series of other studies, such as those

by (Nadarajah and Kotz, 2003), (Sahu, Dey, and Branco, 2003), (Arellano-Valle and Genton, 2005), (Marchenko and Genton, 2010) and (Arellano-Valle and Genton, 2010b).

Based on our studies in this area of knowledge, it was possible to find applications in several areas, such as finance, economics, biology, ecology, engineering, medicine, IT and artificial intelligence, social sciences, meteorology and environmental sciences, marketing and research. market, quality and process control.

All this development, the result of extensive research, made it possible to expand to a new family of distributions, which will be explored in more detail throughout this work.

All this development, the result of extensive research, not only made it possible to expand knowledge about asymmetric normal distributions and its generalization to asymmetric distributions, but also paved the way for a new family of distributions. This new family of distributions, inspired by theoretical and practical advances achieved over the last decades, can offer an even more flexible and comprehensive approach to modeling a wide variety of phenomena in different fields of knowledge. Throughout this work, this new family of distributions will be explored in more detail, highlighting their properties, applications and implications for understanding and analyzing data.

In order to structure this work in a detailed and organized manner, in chapter two, preliminary concepts, we will address fundamental topics that will serve as a basis for understanding the proposed model. We will begin by discussing asymmetry, exploring its variations such as positive asymmetry (or to the right), negative asymmetry (or to the left) and asymmetry with long tails, essential aspects for understanding the dispersion of data in statistical models. Next, distributions with heavy tails, which are of great relevance in modeling extreme events.

We will move on to Elliptical Distributions, with emphasis on the multivariate normal distribution, one of the best known within this group and widely used in various fields of statistics. Later, we will explore the asymmetric elliptical distribution, which expands traditional elliptical distributions by allowing the presence of asymmetries, offering greater flexibility in real applications.

We will also discuss the multivariate covariance matrix and the Cholesky decomposition, which are indispensable tools for the construction and analysis of multivariate models, providing a deeper understanding of the variance and correlation relationships between variables.

The Chapter 2 will conclude with the presentation of goodness-of-fit tests, such as the Kolmogorov-Smirnov test and the Anderson-Darling test, which are fundamental for assessing the quality of the fit of distributions to empirical data. In this context, we will detail the empirical distribution function, the Kolmogorov-Smirnov (KS) test and the Anderson-Darling (AD) test, addressing their practical applications in the validation of statistical models. These concepts will provide the essential theoretical basis for the subsequent chapters.

In the Chapter 3, the proposed distribution will be addressed, and we will explore in depth the structure and properties of the multivariate asymmetric model. We will begin with a detailed presentation of the model, including its mathematical formulation and the fundamental characteristics that define it, emphasizing its ability to capture asymmetries in multivariate distributions.

We will then address some structural properties of the model, starting with the stochastic representation, which describes the model in terms of latent variables and random components. We will also analyze special cases that illustrate specific scenarios where the model can be simplified or adjusted to different situations.

We will also explore marginal quantiles, which are essential to understanding the dispersion of individual variables, and conditional distributions, which allow us to understand how the distributions of variables are influenced by the fixation of other variables. We will also discuss the existence of marginal moments, detailing the conditions for the existence of moments such as means and variances, and the Mahalanobis distance for the particular case, a measure that quantifies the distance of a point in relation to the mean considering the covariance structure.

Finally, we will analyze the dependence measures, which assess the relationships between the model variables, capturing both linear and nonlinear correlations, essential for a comprehensive understanding of the model's behavior in different contexts.

In Chapter 4, a detailed study will be conducted using Monte Carlo simulations to verify the convergence of the model parameters under various functions. This study will be crucial to validate the model's performance in different scenarios and conditions, providing evidence of its robustness and applicability.

In Chapter 5 will focus on the practical application of the model in real data, using the presented distribution to analyze different data sets. This section will highlight the model's capabilities and identify its limitations in the analysis of observed phenomena.

In Chapter 6, we will summarize the main results achieved, highlighting the theoretical and practical contributions of the developed model. The implications of the findings will be presented, as well as suggestions for future work. The work will be complemented by a comprehensive list of bibliographical references and an appendix with additional information, technical details and possible extensions of the study, enriching the understanding of the content presented.

Chapter 2

Preliminary Concepts

In this section, some basic and important concepts for the development of the work will be described. It is important to emphasize that understanding these topics is essential, because, even though they are fundamental concepts, they are crucial for understanding the topics covered later. These concepts provide the necessary foundations for dealing with statistical properties and data, offering a clear notion of how to work appropriately with the tools and techniques involved throughout the study.

"Statistical thinking will one day be as necessary for efficient citizenship as the ability to read and write."

- H.G. Wells

2.1 Asymmetry

Asymmetry in a probability density function is a measure of its lack of symmetry with respect to its center point. In other words, a distribution is considered asymmetric if it is not symmetric about its mean. There are different types of asymmetry that can occur in a probability distribution, the three main ones being:

2.1.1 Right or positive skewness

The right tail of the distribution is longer or more extended than the left tail. This means that there are extreme or high values occurring more frequently than would be expected in a symmetric distribution. The mean is greater than the median.

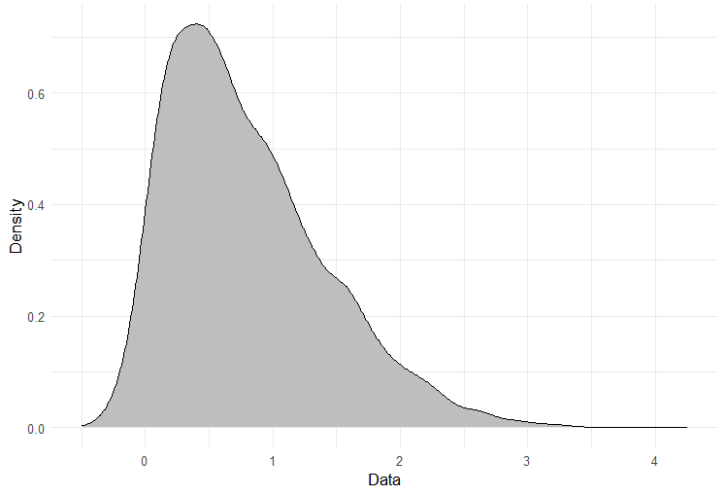


Figure 2.1: Empirical density plot of a right-skewed normal distribution.

2.1.2 Left or negative skewness

In this case, the left tail of the distribution is longer or extended than the right tail. This implies that there are extreme or low values occurring more frequently than would be expected in a symmetric distribution. The mean is smaller than the median.

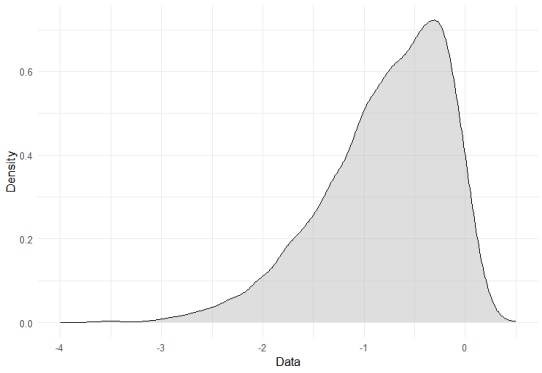


Figure 2.2: Empirical density plot of a left-skewed normal distribution.

2.1.3 Long tail skewness

Refers to a distribution in which the tails (both ends) are heavier than those of a normal distribution. This means that there is a greater likelihood of extreme values occurring than would be expected in a normal distribution.

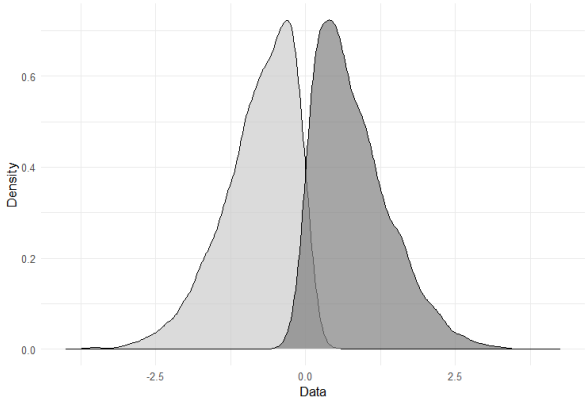


Figure 2.3: Empirical density graph of two normal distributions, one skewed to the left and the other to the right.

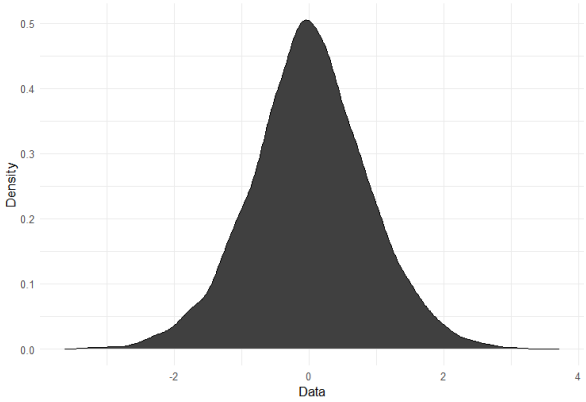


Figure 2.4: Empirical density plot of the sum of two long-tailed distributions.

Skewness in a probability density function can be identified visually through histogram plots or density plots, and can also be quantified through statistical measures such as Pearson's skewness coefficient or Fisher's skewness coefficient.

Visually, the asymmetry can be observed in histograms or by density graphs, the latter seen above. A histogram organizes a set of data into specific intervals, displaying the frequency of values in each interval. Asymmetry is evident if the histogram has a longer tail to the left or right. Similarly, a density plot, which is a smooth estimate of the probability density function of a continuous random variable, reveals asymmetry by the presence of a longer tail in one of the directions.

Quantitatively, asymmetry can be measured using Pearson's asymmetry coefficient or Fisher's asymmetry coefficient.

The equation that defines Fisher's asymmetry coefficient (G_1) according to the (Cain, Zhang, and Yuan, 2017) article is given by:

$$G_1 = \frac{\sqrt{n(n-1)}}{n-2} \frac{m_3}{m_2^{3/2}},$$

where $m_r = \sum_{i=1}^{n}(x_i - \bar{x})^r/n$ represents the r-th central moment, calculated by the average of the deviations of the values in relation to the mean sample $\bar{x}$ raised to the power r. Here, $\bar{x}$ is the sample mean and n is the sample size.

2.2 Heavy-Tailed Distributions

According to (Nair, Wierman, and Zwart, 2024), a distribution function F is said to be heavy-tailed if, and only if, for all $\mu > 0$,

$$\limsup_{x\to\infty} \frac{1-F(x)}{e^{-\mu x}} = \limsup_{x\to\infty} \frac{S(x)}{e^{-\mu x}} = \infty.$$

Otherwise, F is light-tailed.

Being F of a random variable X, that is, $F(x) = \mathbb{P}(X \leq x)$, and of the survival function S, that is , $S(x) = 1 - F(x)$.

A random variable X is said to be heavy-tailed (or light-tailed) if its distribution function is

heavy-tailed (or light-tailed).

This definition applies to the right tail of the distribution, that is, to the behavior of the probability of values greater than x as x tends to infinity. It can also be applied to the left tail by considering the right tail of $-X$.

The book ((Nair, Wierman, and Zwart, 2024)) proves and asserts about a random variable X are equivalent:

1. X is heavy-tailed.
2. $M(s) := \mathbb{E}[e^{sX}] = \infty$ for all $s > 0$.
3. $\liminf_{x\to\infty} \frac{-\log \mathbb{P}(X>x)}{x} = 0$.

Proof. (i) $\implies$ (ii). Suppose X is heavy-tailed, with distribution F. By definition, this implies that for any $s > 0$, there exists a strictly increasing sequence $(x_k)_{k\geq 1}$ satisfying $\lim_{k\to\infty} x_k = \infty$, such that

$$\lim_{k\to\infty} e^{sx_k} S(x_k) = \infty. \tag{2.2.1}$$

Now we can link $\mathbb{E}[e^{sX}]$ as follows.

$$\mathbb{E}[e^{sX}] = \int_0^\infty e^{sx} dF(x) \geq \int_{x_k}^\infty e^{sx} dF(x) \geq e^{sx_k} S(x_k).$$

As the above inequality holds for all k, it now follows from 2.2.1 that $\mathbb{E}[e^{sX}] = \infty$. Therefore, Condition (i) implies Condition (ii).

(ii) $\implies$ (iii). Suppose X satisfies Condition (ii). For the purpose of obtaining a contradiction, let us assume that Condition (iii) does not hold. Since $\liminf_{x\to\infty}[-\log \Pr(X > x)]/x \geq 0$, this means that

$$\liminf_{x\to\infty} \frac{-\log \mathbb{P}(X > x)}{x} > 0.$$

The above statement implies that there exist $\mu > 0$ and $x_0 > 0$ such that

$$\frac{-\log \mathbb{P}(X > x)}{x} \geq \mu \implies \mathbb{P}(X > x) \leq e^{-\mu x} \quad \forall x \geq x_0. \tag{2.2.2}$$

Now choose s such that $0 < s < \mu$. We can now limit the momentum generating function of X in s as follows:

$$M(s) = \mathbb{E}[e^{sX}] = \int_0^\infty \mathbb{P}(e^{sX} > x)\, dx = \int_0^{e^{sx_0}} \mathbb{P}(e^{sX} > x)\, dx + \int_{e^{sx_0}}^\infty \mathbb{P}\left(X > \frac{\log(x)}{s}\right) dx.$$

Here, we use the following representation for the expectation of a non-negative random variable Y: $\mathbb{E}[Y] = \int_0^\infty \mathbb{P}(Y > y)\, dy$. Although the first term above can be bounded from above by e^{sx_0}, we can bound the second using 2.2.2, since $x \geq e^{sx_0}$ is equivalent to $\log(x)/s \geq x_0$.

$$M(s) \leq e^{sx_0} + \int_{e^{sx_0}}^\infty e^{-\mu \frac{\log(x)}{s}}\, dx = e^{sx_0} + \int_{e^{sx_0}}^\infty x^{-\mu/s}\, dx.$$

Since $\mu/s > 1$, we have $\int_1^\infty x^{-\mu/s}\, dx < \infty$, which implies that $M(s) < \infty$, giving us a contradiction. Therefore, Condition (ii) implies Condition (iii).

(iii) $\implies$ (i). Suppose the random variable X, having distribution F, satisfies Condition (iii). Thus, there exists a strictly increasing sequence $(x_k)_{k\geq 1}$ satisfying $\lim_{k\to\infty} x_k = \infty$, such that

$$\lim_{k\to\infty} \frac{-\log S(x_k)}{x_k} = 0.$$

Given $\mu > 0$, this in turn implies that there exists $k_0 \in \mathbb{N}$ such that

$$\begin{aligned} &\frac{-\log S(x_k)}{x_k} \leq \frac{\mu}{2} \quad ,\forall k \geq k_0; \\ \implies\ &S(x_k) \geq e^{-\frac{\mu}{2}x_k} \quad ,\forall k \geq k_0; \\ \implies\ &e^{\mu x_k} S(x_k) \geq e^{\frac{\mu}{2}x_k} \quad ,\forall k \geq k_0. \end{aligned}$$

The last statement above implies that $\lim_{k\to\infty} e^{\mu x_k} S(x_k) = \infty$, which implies

$$\limsup_{x\to\infty} S(x)e^{\mu x} = \infty.$$

Since this is true for any $\mu > 0$, we conclude that Condition (iii) implies Condition (i). □

In other words, a distribution is considered heavy-tailed when its decay is slower than that of the exponential distribution or when its generating moments tend to infinity. Some notable examples of distributions with heavy tails include: the Pareto Distribution, the Stable Lévy Distribution, the Lognormal Distribution, the Weibull Distribution (with shape parameter less than 1), and the Generalized Extreme Value Distribution (GEV) .

2.3 Elliptical Distribution

Elliptical distributions represent an important and fundamental class of statistical and probabilistic models, which offer flexibility and complex analyzes for data modeling. It's distinctive graphic characteristic is its elliptical shape, providing a particular geometric representation.

According to (Fang, Kotz, and Ng, 1990), a random vector $\mathbf{X}$ of dimension $n \times 1$ follows an elliptical distribution with a location parameter $\boldsymbol{\mu}$ of dimension $n \times 1$ and a scale matrix $\boldsymbol{\Sigma}$ of dimension $n \times n$ if its density function is expressed as:

$$f_{\mathbf{X}}(\mathbf{x}) = |\boldsymbol{\Sigma}|^{-1/2} g^{(n)} \left[(\mathbf{x}-\boldsymbol{\mu})^\top \boldsymbol{\Sigma}^{-1}(\mathbf{x}-\boldsymbol{\mu})\right],$$

where $g^{(n)}$ denotes the data generating distribution, and the exponent n represents the dimension of the vector space in which the vector $\mathbf{X}$ is defined.

The data generator plays an essential role in this equation, especially when it is based on symmetric distributions. The most commonly used ones include: multivariate normal, multivariate Cauchy, and multivariate Student's t. These functions characterize the dispersion of the data and the shape of the tails of the distributions, allowing us to model a wide range of

multivariate probabilistic behaviors.

2.4 Multivariate Normal Distribution

The multivariate normal distribution, also known as elliptical normal distribution, is a generalization of the univariate normal distribution to the case in which there are multiple correlated random variables. It is considered a random vector X with n random variables, where $\mathbf{X} = (X_1, X_2, \ldots, X_n)$.

According to (Johnson and Wichern, 2007), the probability density function of a multivariate normal distribution is given by:

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

Here, $\boldsymbol{\mu}$ represents the vector of means of the random variables, $\boldsymbol{\Sigma}$ is the covariance matrix of the random variables, and $|\boldsymbol{\Sigma}|$ is the determinant of the covariance matrix.

The vector $\boldsymbol{x}$ is the multivariate random variable that follows the multivariate normal distribution, and $(\boldsymbol{x} - \boldsymbol{\mu})^\top$ represents the transposition of the vector of differences between $\boldsymbol{x}$ and $\boldsymbol{\mu}$.

This formula describes how the probability is distributed around the vector of means $\boldsymbol{\mu}$ in a n-dimensional space, taking into account the correlations between the variables represented by the covariance matrix $\boldsymbol{\Sigma}$, which is denoted by $\mathcal{N}_n(\boldsymbol{\mu}, \boldsymbol{\Sigma})$.

Below is a graphical example of the density (on the right) of a distribution $\mathcal{N}_2(\mathbf{0}, \boldsymbol{I})$ and a scatterplot (on the left) of the joint $\mathcal{N}_1(0, 1)$ and $\mathcal{N}_1(0, 1)$, with each vector having 25 thousand generated samples.

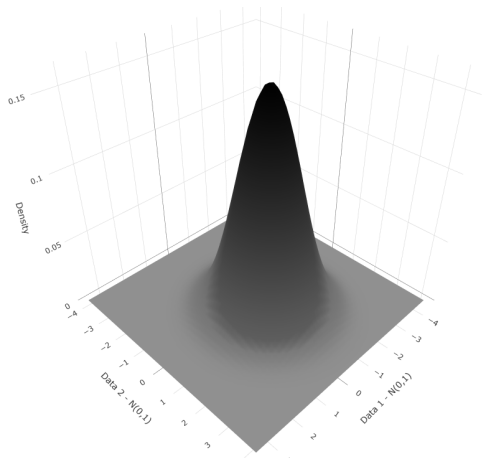


Figure 2.5: Bivariate Normal Distribution Density Chart.

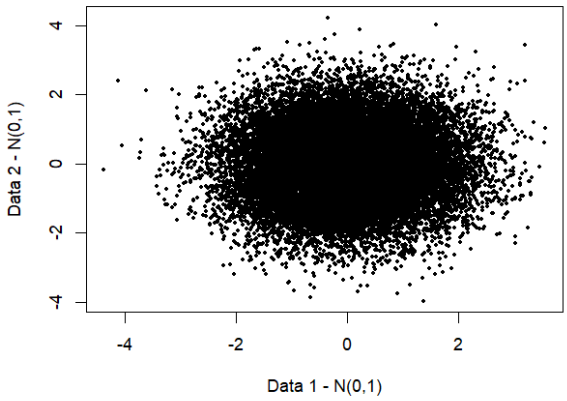


Figure 2.6: Bivariate Normal Distribution Scatter Plot.

2.5 Asymmetric Elliptical Distribution

The asymmetric elliptical distribution is an important class of multivariate distributions that is characterized by not following a perfect circular or elliptical shape. Instead, it describes a data set in which the tails of its univariate distributions are not symmetric about the median of its density. This asymmetry can be observed in different dimensions, making the asymmetric elliptical distribution a valuable tool for modeling a wide range of complex and heterogeneous phenomena.

Where $\boldsymbol{\mu}$ and $\boldsymbol{\Sigma}$ represent the location vector and the scale matrix, respectively, following the same logic as the previous equations. Furthermore, $g^{(n+1)}$ denotes the spherical density generating function with $n+1$ dimensions, as described by (Fang, Kotz, and Ng, 1990).

An essential characteristic of the asymmetric elliptical distribution is its stochastic representation, which allows the simulation of data according to this distribution. This representation can be found in studies such as that of (Zuo, Balakrishnan, and Yin, 2023), providing a valuable tool for statistical analysis and modeling in several areas.

To better understand and visualize this distribution, the density and spread of asymmetric

elliptical data are often illustrated graphically. The examples below show a density plot and a scatter plot, highlighting the distinct characteristics of this distribution.

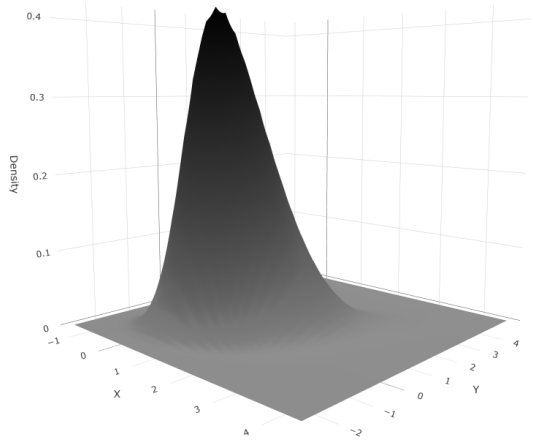


Figure 2.7: Density plot of asymmetric elliptical (skew-normal) distribution.

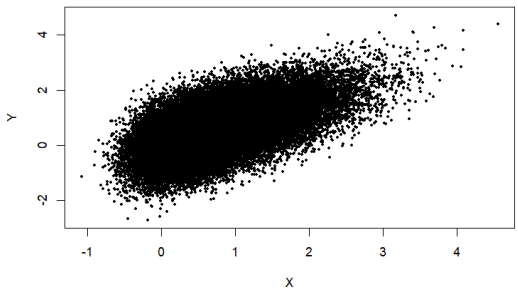


Figure 2.8: Scatterplot of asymmetric elliptical (skew-normal) distribution.

These plots provide a powerful visual representation of the properties of the skewed elliptical distribution, aiding in the interpretation and analysis of data modeled by this distribution.

2.6 Covariance Matrix

To describe the covariance matrix in a multivariate context, we use a generalization used by (Vila et al., 2023), which addresses the bivariate case. Based on this idea, we expand to the general multivariate framework, as detailed below:

$$\boldsymbol{\Sigma} = \begin{pmatrix} \sigma_1^2 & \rho_{12}\sigma_1\sigma_2 & \rho_{13}\sigma_1\sigma_3 & \cdots & \rho_{1n}\sigma_1\sigma_n \\ \rho_{21}\sigma_2\sigma_1 & \sigma_2^2 & \rho_{23}\sigma_2\sigma_3 & \cdots & \rho_{2n}\sigma_2\sigma_n \\ \rho_{31}\sigma_3\sigma_1 & \rho_{32}\sigma_3\sigma_2 & \sigma_3^2 & \cdots & \rho_{3n}\sigma_3\sigma_n \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ \rho_{n1}\sigma_n\sigma_1 & \rho_{n2}\sigma_n\sigma_2 & \rho_{n3}\sigma_n\sigma_3 & \cdots & \sigma_n^2 \end{pmatrix}.$$

In this matrix, $\boldsymbol{\Sigma}$ is the covariance matrix where the diagonal elements $(\sigma_1^2, \sigma_2^2, \ldots, \sigma_n^2)$

correspond to the variances of the respective variables. The off-diagonal elements ($\rho_{ij}\sigma_i\sigma_j$) represent the covariances between pairs of variables, weighted by the correlation coefficients ρ_{ij}. Each σ_i^2 indicates the variability of a specific variable, while ρ_{ij} reflects the degree of correlation between two distinct variables, capturing how one variable can influence the other.

This structure is fundamental to understanding the interdependence between variables in a multivariate context, as it reflects both the individual variability of the variables and the correlation relationships between them. Non-zero correlation values (ρ_{ij}) indicate a direct or inverse relationship between variables, adding complexity to the analysis of multivariate data dispersion.

2.7 Cholesky decomposition

The data generation method is one of the important parts of this work, which involves a simulation of data with multivariate normal distribution. The method chosen for this was the Cholesky decomposition, developed by André-Louis Cholesky, a French army officer and civil engineer, which was developed in the early 20th century (Cholesky, 2005). From this factorization, it becomes possible to arrive at a covariance matrix, which has the property of being symmetric and positive defined, always allowing the creation of a symmetric positive defined matrix as a product of a lower triangular matrix and its transpose. For this to work, follow these steps:

1. Generate a matrix $\boldsymbol{Z}$ of size $n \times d$ containing nd pseudo-random numbers from the distribution $N(0, 1)$.

2. Compute the decomposition $\boldsymbol{\Sigma} = \boldsymbol{Q}^\top \boldsymbol{Q}$.

3. Apply the transformation $\boldsymbol{X} = \boldsymbol{Z}\boldsymbol{Q} + \boldsymbol{J}\boldsymbol{\mu}^\top$, where $\boldsymbol{J}$ is a column vector of ones.

4. Get the matrix $\boldsymbol{X}$ of size $n \times d$.

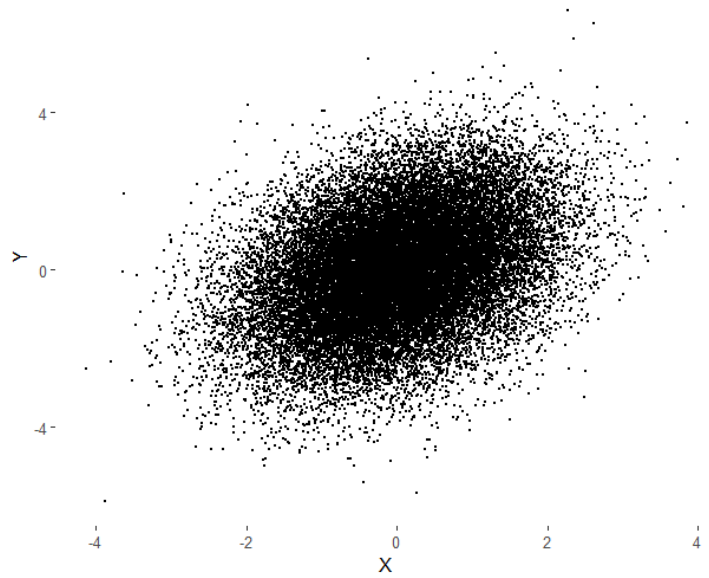


Figure 2.9: Bivariate Scatterplot Obtained through Cholesky Decomposition.

2.8 Fit Tests: Kolmogorov-Smirnov and Anderson-Darling

2.8.1 Empirical distribution function

The empirical distribution function (eCDF) is a fundamental tool in statistics for estimating the cumulative distribution function of a random variable based on an observed sample. As described in the book (Vaart, 1998) it is defined by the expression:

$$\forall x \in \mathbb{R}, \forall \omega \in \Omega, F_n(x, \omega) = \frac{1}{n} \sum_{i=1}^{n} \mathbb{1}_{\{\sigma: X_i(\sigma) \leq x\}}(\omega),$$

where $\mathbb{1}_{\{\sigma: X_i(\sigma) \leq x\}}$ is an indicator function that takes the value 1 if $X_i(\omega)$ is less than or equal to x , and 0 otherwise. This indicator function allows you to check each element in the sample to see if it satisfies the condition $X_i(\sigma) \leq x$, adding 1 for each element that meets the condition and 0 for those that do not. The sum $\sum_{i=1}^{n} \mathbb{1}_{\{\sigma: X_i(\sigma) \leq x\}}$ represents the total number of elements in the sample that are less than or equal to x. Dividing this sum by the sample size n,

we obtain the proportion of elements that satisfy the condition, thus estimating the distribution function of the random variable based on the observed data.

One of the main advantages of eCDF is its non-parametric nature, that is, there is no need for specific assumptions about the distribution of the data. This makes eCDF a flexible tool for empirical analysis. However, there are also limitations to be considered, such as sensitivity to small samples, which can result in high variability and significant fluctuations in estimating the true distribution of the population.

As the sample size increases ($n \to \infty$), the eCDF converges to the true population cumulative distribution function. This behavior is guaranteed by the Glivenko-Cantelli Theorem, which states that the eCDF is a consistent estimate, that is, it converges in probability, almost certainly (law of large numbers) and uniformly, reinforcing its theoretical basis and reliability for statistical inferences.

The eCDF concept extends from the univariate to the multivariate case, which is our context, simply considering that the random variables are of d dimensions, belonging to $\mathbb{R}^d$. In the multivariate case, the approach remains similar, but the calculation and visualization can become more complex due to the interaction between variables in multiple dimensions.

The concepts of Cumulative Distribution Function and eCDF are crucial in evaluating the quality of the fit, comparing a simulated distribution, generated from the estimated data parameters, with the observed data set itself. We will make this comparison by performing statistical tests such as Kolmogorov-Smirnov and Anderson-Darling, which use eCDF as a basis to evaluate the fit of simulated distributions to real data.

2.8.2 Kolmogorov-Smirnov test (KS)

The Kolmogorov-Smirnov (KS) test is a non-parametric statistical test used to compare two probability distributions or check whether a sample follows a specific theoretical distribution, it was introduced in (Kolmogorov, 1933) and improved in (Smirnov, 1939). It is useful for testing the equality of distributions and the adherence of data to an expected distribution. It can

be applied in two ways: with a single sample, comparing it with a theoretical distribution to evaluate how well the data fits that distribution; or with two samples, comparing them to each other to determine whether they both come from the same distribution.

The test measures the maximum difference between the cumulative distribution functions (CDF) of the compared distributions:

$$D = \sup |F_X(x) - F_Y(x)|,$$

where D is the KS statistic, $F_X(x)$ is the CDF of sample X and $F_Y(x)$ is the eCDF of sample Y.

The p-value represents the probability of observing a difference as extreme as D, assuming that the compared distributions are equal. In this work, we will adopt the significance level of 0.05 (5%), which is widely used in statistical analyses. The test hypotheses are as follows:

- H_0: The data follows the same distribution.

- H_1: The data does not follow the same distribution.

2.8.3 Anderson-Darling test (AD)

The Anderson-Darling test was developed by Theodore W. Anderson and Donald A. Darling in 1952 in (Anderson and Darling, 1952) . It is an extension of the Kolmogorov-Smirnov test and was introduced to improve the sensitivity of the test on the tail of distributions, making it more effective at detecting outliers at the tails.

The Anderson-Darling test statistic, A^2, is calculated based on the sample's empirical distribution function (eCDF) and theoretical cumulative distribution function (CDF).

For an ordered sample $X_1, X_2, \ldots, X_n$, the statistic A^2 is given by:

$$A^2 = -n - \frac{1}{n} \sum_{i=1}^{n} \left((2i-1) \left[\ln(F(X_i)) + \ln(1 - F(X_{n+1-i})) \right] \right),$$

where n is the sample size, X_i is the ordered sample values and $F(X_i)$ is the theoretical eCDF

evaluated on the sample values.

This formula weights observed discrepancies with greater weight at the ends of the distribution, making the test more sensitive to deviations in the tails. The higher the value of A^2, the greater the discrepancy between the sample and the theoretical distribution.

The interpretation of the p-value and the hypotheses of the Anderson-Darling test follow the same principles as the Kolmogorov-Smirnov test, since the AD test is based on the KS. In both tests, the p-value is used to determine whether there is enough evidence to reject the null hypothesis, which assumes that the data follows the distribution.

Chapter 3

Extended *G*-skew Distribution (EGSE$_n$).

This section is one of the most relevant of this work, as it presents a new multivariate asymmetric distribution model developed by us. Based on elliptical and asymmetric distributions, and based on a vast literature review, this model integrates important concepts previously explored, but from a new perspective. The proposal seeks to provide an innovative approach to the treatment of asymmetric data, expanding the possibilities of analysis and application in different statistical contexts.

"*If I only had an hour to chop down a tree, I would spend the first forty-five minutes sharpening my axe.*"
- Unknown

3.1 The multivariate asymmetric model

Let $G_1, \ldots, G_n : (0, \infty) \to \mathbb{R}$, $n \in \mathbb{N}$, be monotonically and strictly increasing functions (consequently Borel-measurable). We say that a n-dimensional random vector $\boldsymbol{Y} = (Y_1, \ldots, Y_n)^\top$ has a multivariate extended G-skew-elliptical (EGSE$_n$) distribution if its probability distribution function (PDF) at $\boldsymbol{y} = (y_1, \ldots, y_n)^\top \in (0, \infty)^n$, denoted by $f_{\boldsymbol{Y}}(\boldsymbol{y})$, is given by

$$f_{\boldsymbol{Y}}(\boldsymbol{y}) = \frac{1}{|\boldsymbol{\Sigma}|^{1/2} Z_{g^{(n)}}} g^{(n)}(q(\boldsymbol{y}_G)) \frac{F_{\mathrm{ELL}_1}(\boldsymbol{\lambda}^\top(\boldsymbol{y}_G - \boldsymbol{\mu}) + \tau;\, 0, 1, g_{q(\boldsymbol{y}_G)})}{F_{\mathrm{ELL}_1}(\tau;\, 0, 1 + \boldsymbol{\lambda}^\top \boldsymbol{\Sigma} \boldsymbol{\lambda}, g^{(1)})} \prod_{i=1}^{n} G_i'(y_i), \quad (3.1.1)$$

where $g^{(n)}$ is a density generator (Fang, Kotz, and Ng, 1990),

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

is a normalization constant, $\boldsymbol{\mu} = (\mu_1, \ldots, \mu_n)^\top \in \mathbb{R}^n$ is a constant vector, $\boldsymbol{\Sigma}$ is a positive definite $n \times n$ matrix, $\tau \in \mathbb{R}$ is the extension parameter, $\boldsymbol{\lambda} = (\lambda_1, \ldots, \lambda_n)^\top \in \mathbb{R}^n$ is the skewness parameter vector, and

$$q(\boldsymbol{y}_G) = (\boldsymbol{y}_G - \boldsymbol{\mu})^\top \boldsymbol{\Sigma}^{-1} (\boldsymbol{y}_G - \boldsymbol{\mu}), \quad \text{for } \boldsymbol{y}_G = (G_1(y_1), \ldots, G_n(y_n))^\top \in \mathbb{R}^n. \quad (3.1.2)$$

Moreover, in formula (3.1.1), $F_{\mathrm{ELL}_1}(\cdot;\, 0, 1, g_{q(\boldsymbol{y}_G)})$ denotes the cumulative distribution function (CDF) of a univariate elliptical (symmetric) (ELL$_1$) and standardized random variable (Fang, Kotz, and Ng, 1990) with density generator $g_{q(\boldsymbol{y}_G)}$, which is defined as

$$g_{q(\boldsymbol{y}_G)}(s) = \frac{g^{(2)}(s + q(\boldsymbol{y}_G))}{g^{(1)}(q(\boldsymbol{y}_G))}, \quad s \in \mathbb{R}. \quad (3.1.3)$$

Analogously we define $F_{\mathrm{ELL}_1}(\cdot;\, 0, 1, g^{(1)})$. For simplicity of notation, we write

$$\boldsymbol{Y} \sim \mathrm{EGSE}_n(\boldsymbol{\mu}, \boldsymbol{\Sigma}, \boldsymbol{\lambda}, \tau, g^{(n)})$$

when a vector $\boldsymbol{Y}$ has EGSE$_n$ distribution (3.1.1). Table 3.1 presents some examples of functions G_i's for use in (3.1.1).

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

$G_i(x)$	$G_i^{-1}(x)$	$G_i'(x)$	Parameters
$\frac{1}{\alpha}\left(\sqrt{\frac{x}{\beta}}-\sqrt{\frac{\beta}{x}}\right)$	$\beta\left[\frac{\alpha}{2}x+\sqrt{\left((\frac{\alpha}{2}x)^2+1\right)}\right]^2$	$\frac{1}{2\alpha x}\left(\sqrt{\frac{x}{\beta}}+\sqrt{\frac{\beta}{x}}\right)$	$\alpha,\beta>0$
$\frac{x^{1-q}-1}{1-q}$	$[1+(1-q)x]^{\frac{1}{1-q}}$	$\frac{1}{x^q}$	$q\neq 1$
$\log(x)$	$\exp(x)$	$\frac{1}{x}$	-
$\cosh^{-1}(x+1)$	$\cosh(x)-1$	$\frac{1}{(\sqrt{x+2})\sqrt{x}}$	-
$\log(\log(x+1))$	$\exp(\exp(x))-1$	$\frac{1}{(x+1)\log(x+1)}$	-
$x-\frac{1}{x}$	$\frac{1}{2}\left(x+\sqrt{x^2+4}\right)$	$\frac{1}{x^2}+1$	-
$\frac{2F_i(x)-1}{F_i(x)[1-F_i(x)]}$	$F_i^{-1}\left(\frac{-(2-x)+\sqrt{(2-x)^2+4x}}{2x}\right)$	$\frac{2f_i(x)\cdot F_i(x)[1-F_i(x)]-(2F_i(x)-1)\cdot f_i(x)[1-2F_i(x)]}{(F_i(x)[1-F_i(x)])^2}$	-

Recalling that $F_i(x)$ denotes the cumulative distribution function (CDF), $F_i^{-1}(x)$ represents its inverse (quantile function), and $f_i(x)$ is the associated probability density function (PDF).

In situations where the inverse of a function presents multiple possibilities, such as square roots, it is crucial to consider the context of the domain. In particular, when we deal with square roots in positive real numbers, we select the positive root. This choice arises from the mathematical definition of the square root function: $f(x)=\sqrt{x}$, which associates each non-negative real number with a single non-negative value.

In practice, the square root function is considered bijective when restricted to positive reals, with the range being the interval $(0,\infty)$. By defining the inverse, this unique correspondence is preserved, which means selecting the positive root. Without this restriction, the function would not be a function in the strict sense, as each input would correspond to two possible outputs.

Therefore, when solving problems involving square roots, the choice of the positive root is not just a convention, but a mathematical necessity that guarantees the precise definition of the inverse, ensuring coherence and precision in the results within the domain of positive real numbers.

Below it is possible to observe $G_i(x)$, $G_i^{-1}(x)$ and $G_i'(x)$, respectively, graphically.

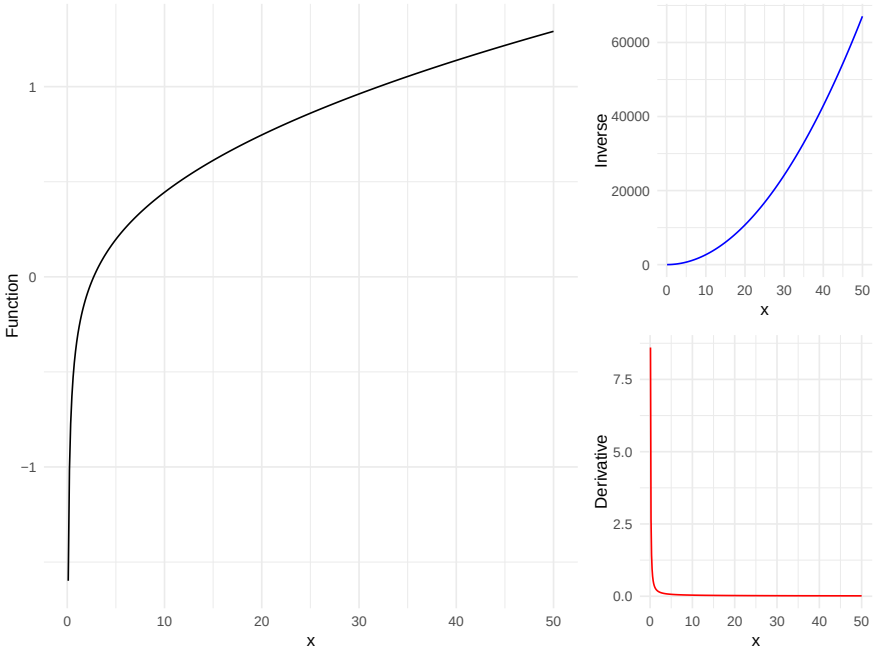


Figure 3.1: $G_i(x) = \frac{1}{\alpha}\left(\sqrt{\frac{x}{\beta}} - \sqrt{\frac{\beta}{x}}\right)$ $\alpha = \pi$ and $\beta = \mathrm{e}$.

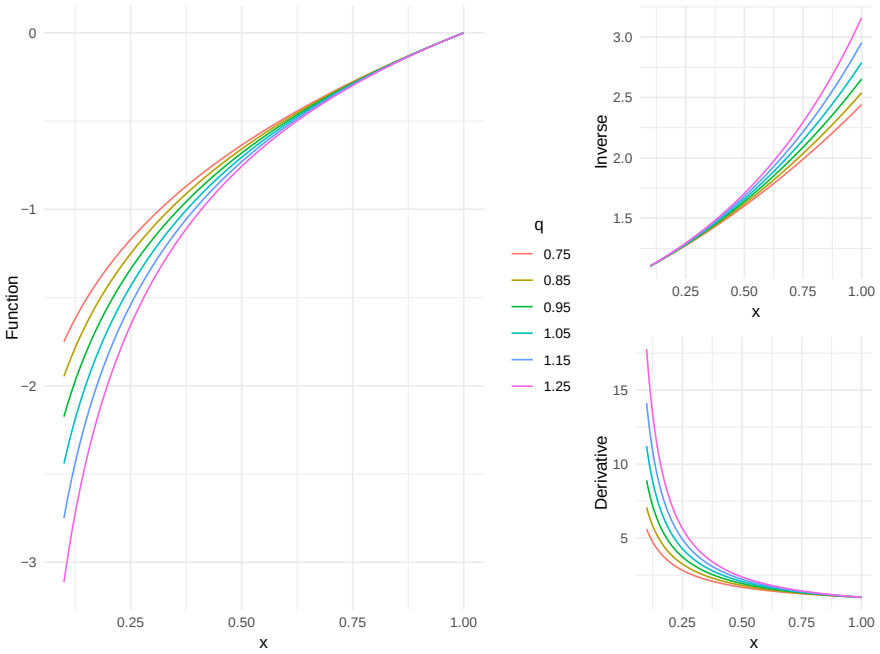


Figure 3.2: $G_i(x) = \log(\log(x+1))$.

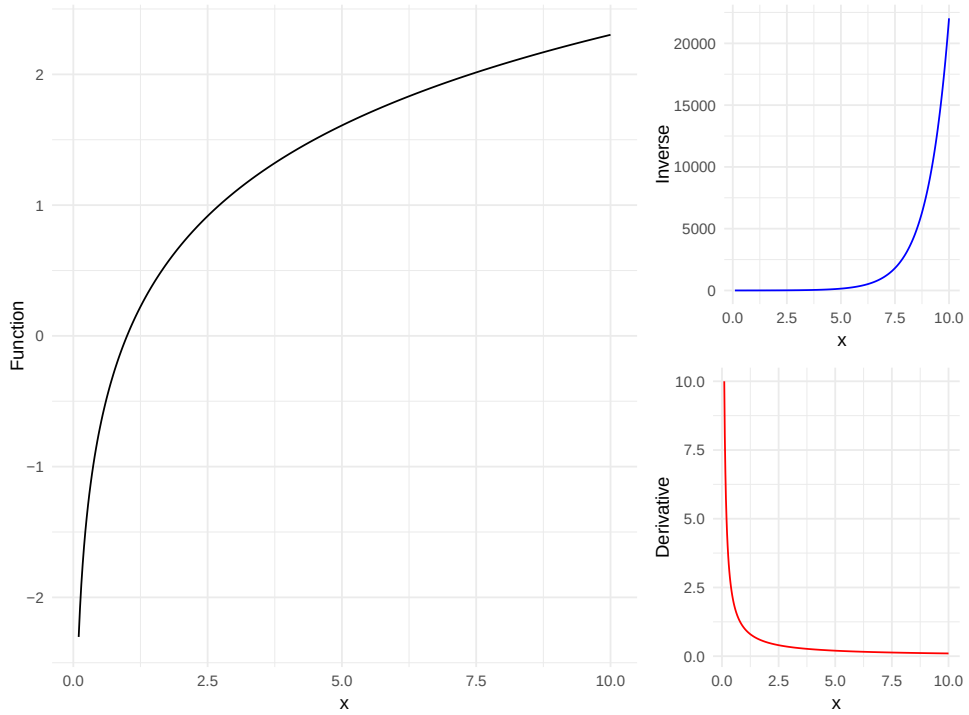


Figure 3.3: $G_i(x) = \log(x)$.

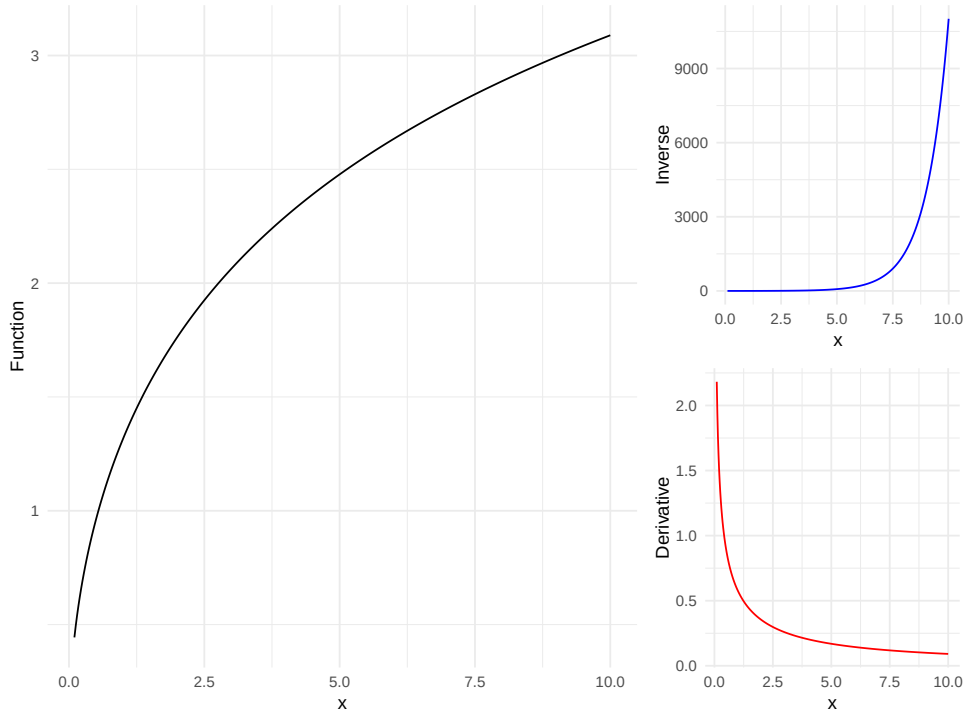


Figure 3.4: $G_i(x) = \cosh^{-1}(x+1)$.

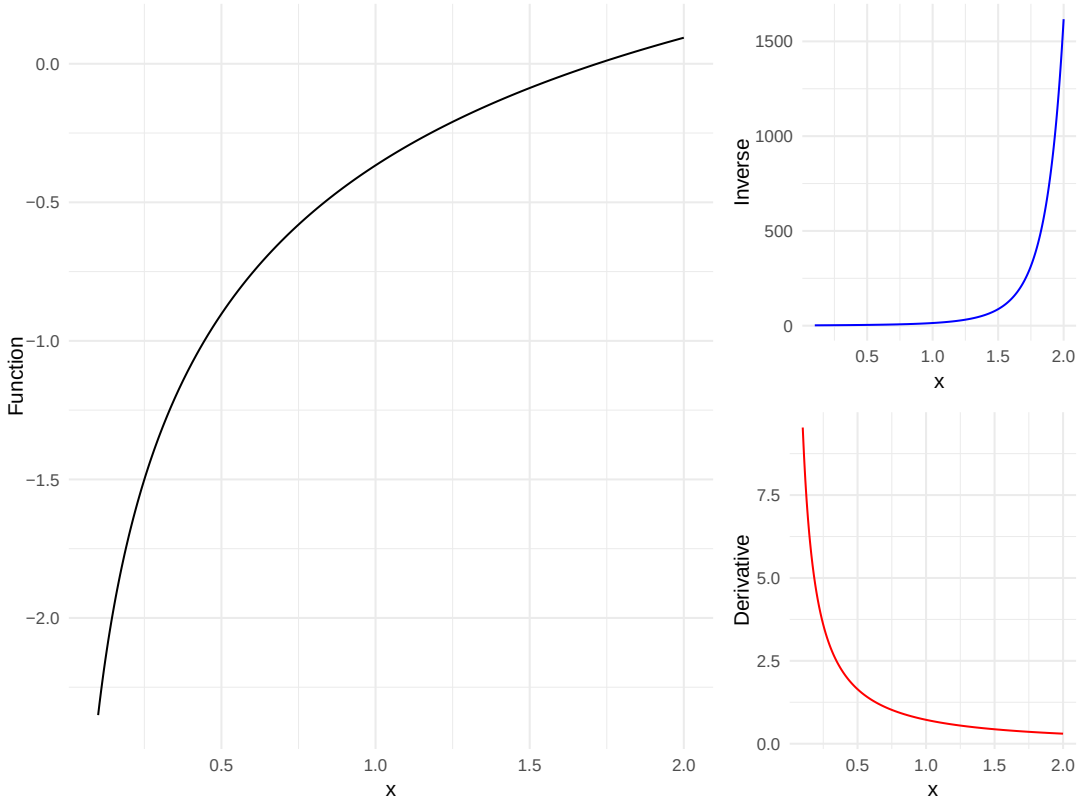


Figure 3.5: $G_i(x) = \log(\log(x+1))$.

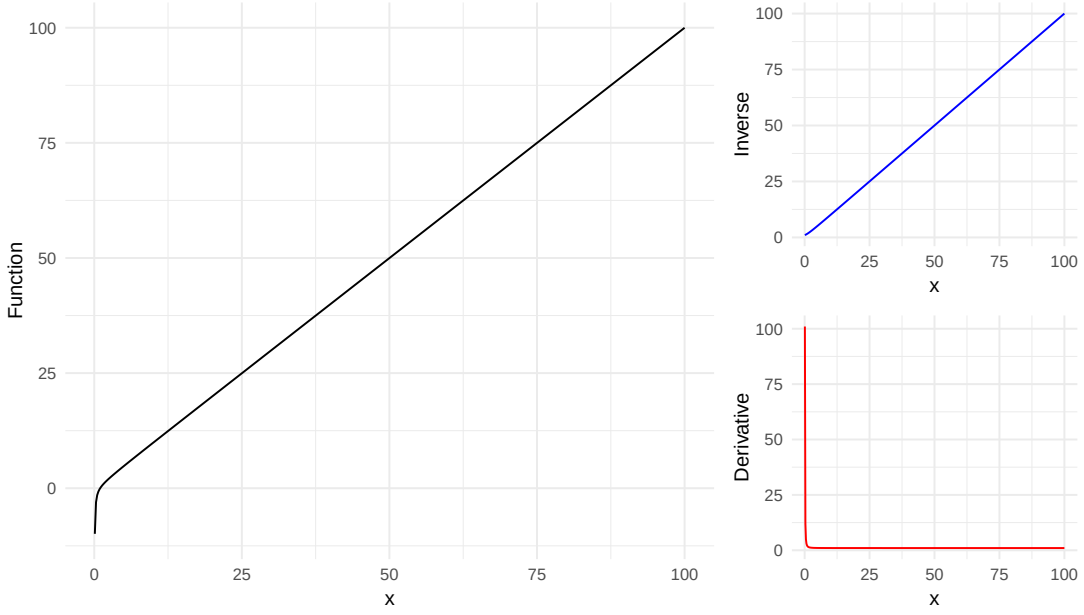


Figure 3.6: $G_i(x) = x - \frac{1}{x}$.

Table 3.2 presents some examples of generators for use in (3.1.1).

Table 3.2: Normalization functions $(Z_{g^{(n)}})$ and density generators $(g^{(n)})$.

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

Closed-form expressions for the PDF of $\boldsymbol{Y} \sim \text{EGSE}_n(\boldsymbol{\mu}, \boldsymbol{\Sigma}, \boldsymbol{\lambda}, \tau, g^{(n)})$ corresponding to multivariate extended G-skew-Student-t, multivariate extended G-skew-Cauchy and multivariate extended G-skew-normal models (see Table 3.2), are provided in Subsection 3.2.2.

The EGSE_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 distributions, with presence of asymmetry, which allows modeling date with positive support. For $\tau = 0$ and $G_i(x) = \log(x)$, $x > 0$, $i = 1, \ldots, n$, we obtain the multivariate log-skew-elliptical model studied in Marchenko and Genton, 2010. In general, for the EGSE_n model, it is not necessary to consider all G_i's equal as in Marchenko and Genton, 2010. For $g^{(n)}(x) = (1 + x/\nu)^{-(\nu+n)/2}$, $\nu > 0$, we get the multivariate extended 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.

Given the closed form of the density function, it is possible to obtain an idea of the Cumulative Distribution Function (CDF). However, in our case, the CDF does not have a closed form, requiring the use of numerical methods to estimate its values accurately. More details about the estimation procedure can be found in the annex, where we describe the approaches used.

3.2 Some structural properties

3.2.1 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

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

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

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

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

$$Z - \boldsymbol{\lambda}^\top(\boldsymbol{X} - \boldsymbol{\mu}) \sim \mathrm{ELL}_1\left(0, 1 + \boldsymbol{\lambda}^\top\boldsymbol{\Sigma}\boldsymbol{\lambda}, g^{(1)}\right), \tag{3.2.1}$$

$$\boldsymbol{X} \sim \mathrm{ELL}_n(\boldsymbol{\mu}, \boldsymbol{\Sigma}, g^{(n)}), \tag{3.2.2}$$

$$Z \sim \mathrm{ELL}_1(0, 1, g^{(1)}), \tag{3.2.3}$$

$$Z \,|\, \boldsymbol{X} = \boldsymbol{x} \sim \mathrm{ELL}_1(0, 1, g_{q(\boldsymbol{x})}), \quad \boldsymbol{x} = (x_1, \ldots, x_n)^\top \in \mathbb{R}^n, \tag{3.2.4}$$

where $q(\boldsymbol{x})$ and $g_{q(\boldsymbol{x})}$ are as in (3.1.2) and (3.1.3), respectively. From (3.2.2), $\boldsymbol{X}$ 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 (3.2.1), (3.2.2) and (3.2.4), the PDF (3.1.1) of $\boldsymbol{Y} \sim \text{EGSE}_n(\boldsymbol{\mu}, \boldsymbol{\Sigma}, \boldsymbol{\lambda}, \tau, g^{(n)})$ is written as

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

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

$$\begin{aligned} f_{\boldsymbol{Y}}(\boldsymbol{y}) &= f_{\boldsymbol{T}}(\boldsymbol{y}) \frac{\int_0^\infty f_{\boldsymbol{\lambda}^\top(\boldsymbol{X}-\boldsymbol{\mu})-Z+\tau \,|\, \boldsymbol{T}=\boldsymbol{y}}(s)\mathrm{d}s}{\mathbb{P}(Z - \boldsymbol{\lambda}^\top(\boldsymbol{X}-\boldsymbol{\mu}) < \tau)} \\ &= \frac{\int_0^\infty f_{\boldsymbol{T},\boldsymbol{\lambda}^\top(\boldsymbol{X}-\boldsymbol{\mu})-Z+\tau}(\boldsymbol{y}, s)\mathrm{d}s}{\mathbb{P}(\boldsymbol{\lambda}^\top(\boldsymbol{X}-\boldsymbol{\mu}) + \tau > Z)} = f_{\boldsymbol{T} \,|\, \boldsymbol{\lambda}^\top(\boldsymbol{X}-\boldsymbol{\mu})+\tau > Z}(\boldsymbol{y}). \end{aligned} \tag{3.2.5}$$

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

$$\boldsymbol{Y} = \boldsymbol{T} \,|\, \boldsymbol{\lambda}^\top(\boldsymbol{X} - \boldsymbol{\mu}) + \tau > Z, \tag{3.2.6}$$

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 (3.2.1)-(3.2.4).

3.2.2 Special cases

In this subsection we develop some examples of multivariate EGSE_n distributions as special cases.

Proposition 3.2.1 (Multivariate extended G-skew-Student-t)**.** *Let $\boldsymbol{Y} \sim \text{EGSE}_n(\boldsymbol{\mu}, \boldsymbol{\Sigma}, \boldsymbol{\lambda}, \tau, g^{(n)})$, 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 $\boldsymbol{y} \in (0,\infty)^n$ is given by

$$f_{\boldsymbol{Y}}(\boldsymbol{y}) = t_n(\boldsymbol{y}_G;\, \boldsymbol{\mu}, \boldsymbol{\Sigma}, \nu)\, \frac{F_{\nu+1}\left([\boldsymbol{\lambda}^\top(\boldsymbol{y}_G - \boldsymbol{\mu}) + \tau]\sqrt{\frac{\nu+1}{\nu+q(\boldsymbol{y}_G)}}\right)}{F_\nu\left(\frac{\tau}{\sqrt{1+\boldsymbol{\lambda}^\top\boldsymbol{\Sigma}\boldsymbol{\lambda}}}\right)} \prod_{i=1}^{n} G_i'(y_i), \tag{3.2.7}$$

where $\boldsymbol{y}_G$ and $q(\boldsymbol{y}_G)$ are as given in (3.1.2) and (3.1.3), 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 3.2, 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 (3.1.1), it is enough to verify that

$$F_{\mathrm{ELL}_1}(\boldsymbol{\lambda}^\top(\boldsymbol{y}_G - \boldsymbol{\mu}) + \tau;\, 0, 1, g_{q(\boldsymbol{y}_G)}) = F_{\nu+1}\left([\boldsymbol{\lambda}^\top(\boldsymbol{y}_G - \boldsymbol{\mu}) + \tau]\sqrt{\frac{\nu+1}{\nu+q(\boldsymbol{y}_G)}}\right) \tag{3.2.8}$$

and

$$F_{\mathrm{ELL}_1}(\tau;\, 0, 1 + \boldsymbol{\lambda}^\top\boldsymbol{\Sigma}\boldsymbol{\lambda}, g^{(1)}) = F_\nu\left(\frac{\tau}{\sqrt{1+\boldsymbol{\lambda}^\top\boldsymbol{\Sigma}\boldsymbol{\lambda}}}\right). \tag{3.2.9}$$

The identity (3.2.9) 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 (3.2.8). 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 generator function $g_{q(\boldsymbol{y}_G)}$ as given in (3.1.3), we have

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

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

$$= \frac{1}{Z_{g^{(2)}}/Z_{g^{(1)}}} \sqrt{\frac{\nu}{\nu+1}} \int_{-\infty}^{(\boldsymbol{\lambda}^\top(\boldsymbol{y}_G-\boldsymbol{\mu})+\tau)\sqrt{\frac{\nu+1}{\nu+q(\boldsymbol{y}_G)}}} \left(1+\frac{t^2}{\nu+1}\right)^{-(\nu+2)/2} \mathrm{d}t. \tag{3.2.10}$$

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 (3.2.10) 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 (3.2.8) follows. □

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

Proposition 3.2.2 (Multivariate extended G-skew-Cauchy)**.** *Let $\boldsymbol{Y} \sim EGSE_n(\boldsymbol{\mu}, \boldsymbol{\Sigma}, \boldsymbol{\lambda}, \tau, g^{(n)})$, where $g^{(n)}(x) = 1/(1+x)^{(n+1)/2}$, $x \in \mathbb{R}$, is the PDF generator of the multivariate Cauchy distribution. Then, the PDF of $\boldsymbol{Y}$ at $\boldsymbol{y} \in (0,\infty)^n$ is given by*

$$f_{\boldsymbol{Y}}(\boldsymbol{y}) = c_n(\boldsymbol{y}_G;\, \boldsymbol{\mu}, \boldsymbol{\Sigma}) \frac{F_2\left([\boldsymbol{\lambda}^\top(\boldsymbol{y}_G-\boldsymbol{\mu})+\tau]\sqrt{\frac{2}{1+q(\boldsymbol{y}_G)}}\right)}{F_1\left(\frac{\tau}{\sqrt{1+\boldsymbol{\lambda}^\top\boldsymbol{\Sigma}\boldsymbol{\lambda}}}\right)} \prod_{i=1}^{n} G_i'(y_i), \tag{3.2.11}$$

where $\boldsymbol{y}_G$ and $q(\boldsymbol{y}_G)$ are as given in (3.1.2) and (3.1.3), 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 3.2, 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 G-skew-normal)**.** *Let $\boldsymbol{Y} \sim$ EGSE$_n(\boldsymbol{\mu}, \boldsymbol{\Sigma}, \boldsymbol{\lambda}, \tau, g^{(n)})$, 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 $\boldsymbol{y} \in (0, \infty)^n$ is given by*

$$f_{\boldsymbol{Y}}(\boldsymbol{y}) = \phi_n(\boldsymbol{y}_G;\, \boldsymbol{\mu}, \boldsymbol{\Sigma}) \frac{\Phi\left(\boldsymbol{\lambda}^\top(\boldsymbol{y}_G - \boldsymbol{\mu}) + \tau\right)}{\Phi\left(\frac{\tau}{\sqrt{1+\boldsymbol{\lambda}^\top\boldsymbol{\Sigma}\boldsymbol{\lambda}}}\right)} \prod_{i=1}^{n} G_i'(y_i), \tag{3.2.12}$$

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 3.2, 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 3.3 summarizes the results found in Propositions 3.2.1, 3.2.2 and 3.2.3. Right after the table, several examples are presented with density and scatter plots, respectively.

Table 3.3: Densities $f_{\boldsymbol{Y}}$ of the EGSE$_n$ distributions of Table 3.2.

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

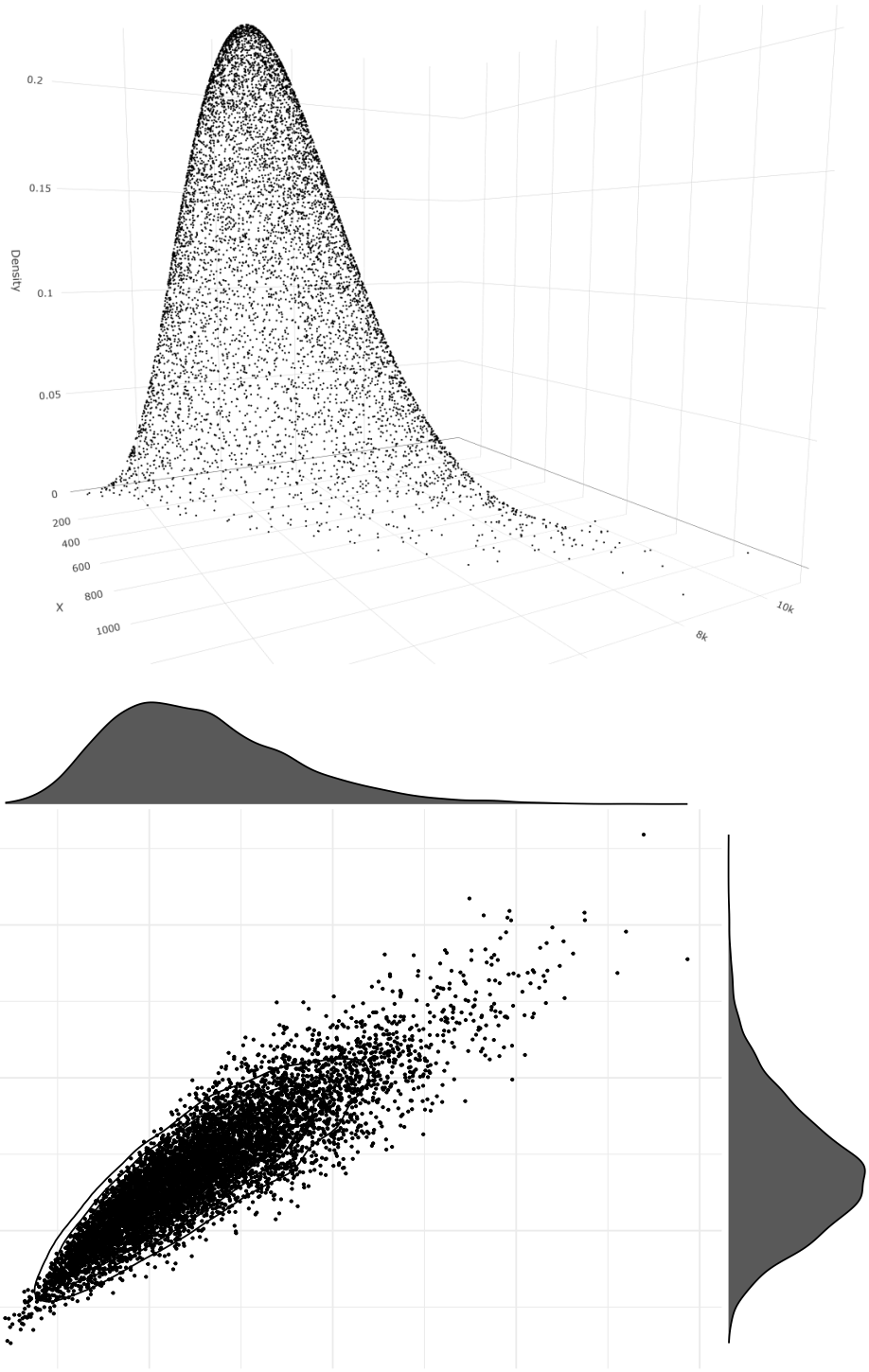


Figure 3.7: Density and scatter plots (with marginal densities) of the function $G_i(x) = \frac{1}{\alpha}\left(\sqrt{\frac{x}{\beta}} - \sqrt{\frac{\beta}{x}}\right)$ with covariance matrix $\mathbf{\Sigma} = \begin{pmatrix} \sigma_1^2 & \rho\sigma_1\sigma_2 \\ \rho\sigma_1\sigma_2 & \sigma_2^2 \end{pmatrix}$, where $\sigma_1 = 1, \sigma_2 = 2, \rho = 0.9, \mu = (4, 12), \lambda = (10, -3), \tau = 5, \alpha = 2.72, \beta = 3.14$.

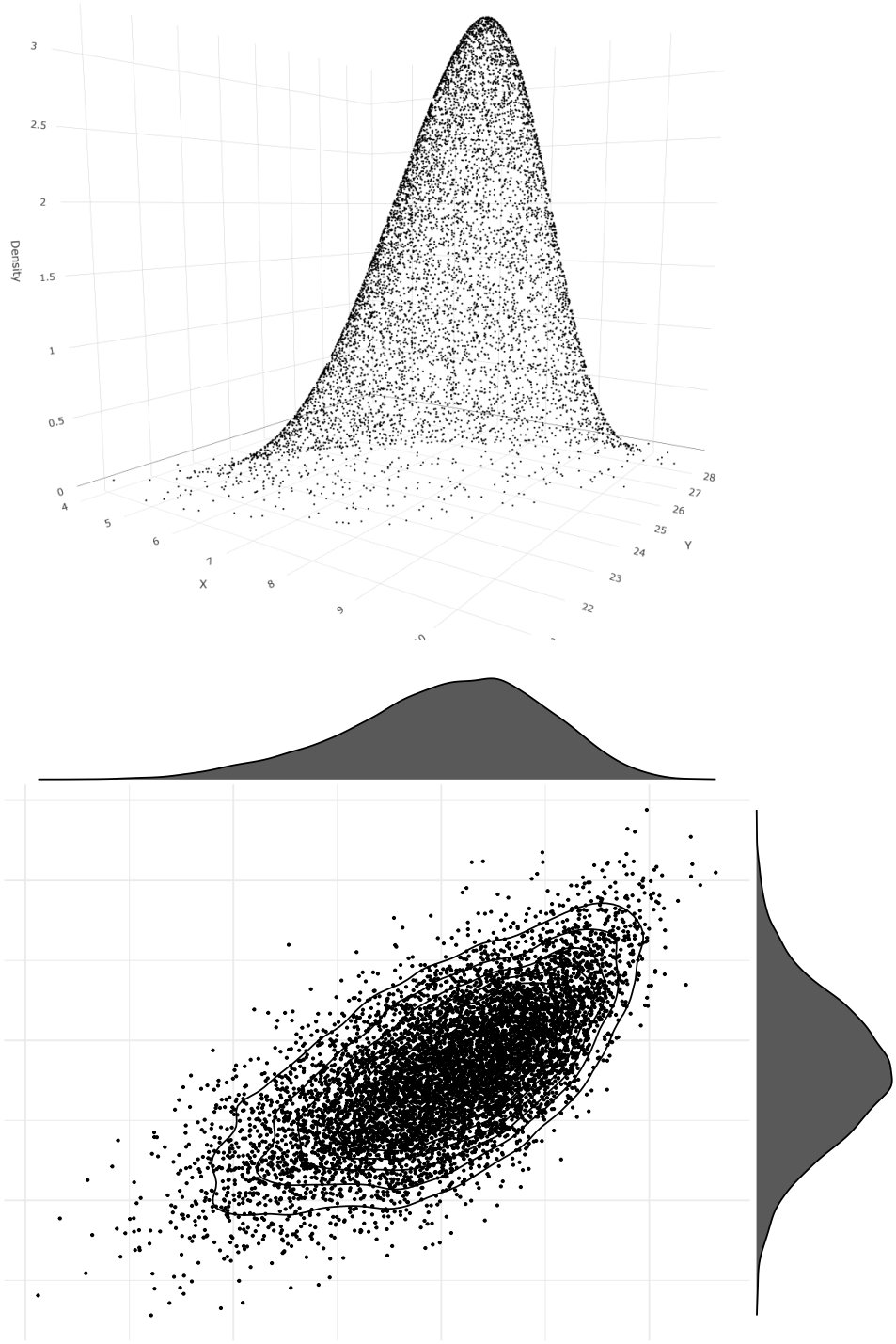


Figure 3.8: Density and scatter plots (with marginal densities) of the function $G_i(x) = \frac{x^{1-q}-1}{1-q}$ with covariance matrix $\mathbf{\Sigma} = \begin{pmatrix} \sigma_1^2 & \rho\sigma_1\sigma_2 \\ \rho\sigma_1\sigma_2 & \sigma_2^2 \end{pmatrix}$, where $\sigma_1 = 0.5, \sigma_2 = 0.25, \rho = 0.75, \mu = (4, 8), \lambda = (-10, 5), \tau = 0.5, q = 0.5$.

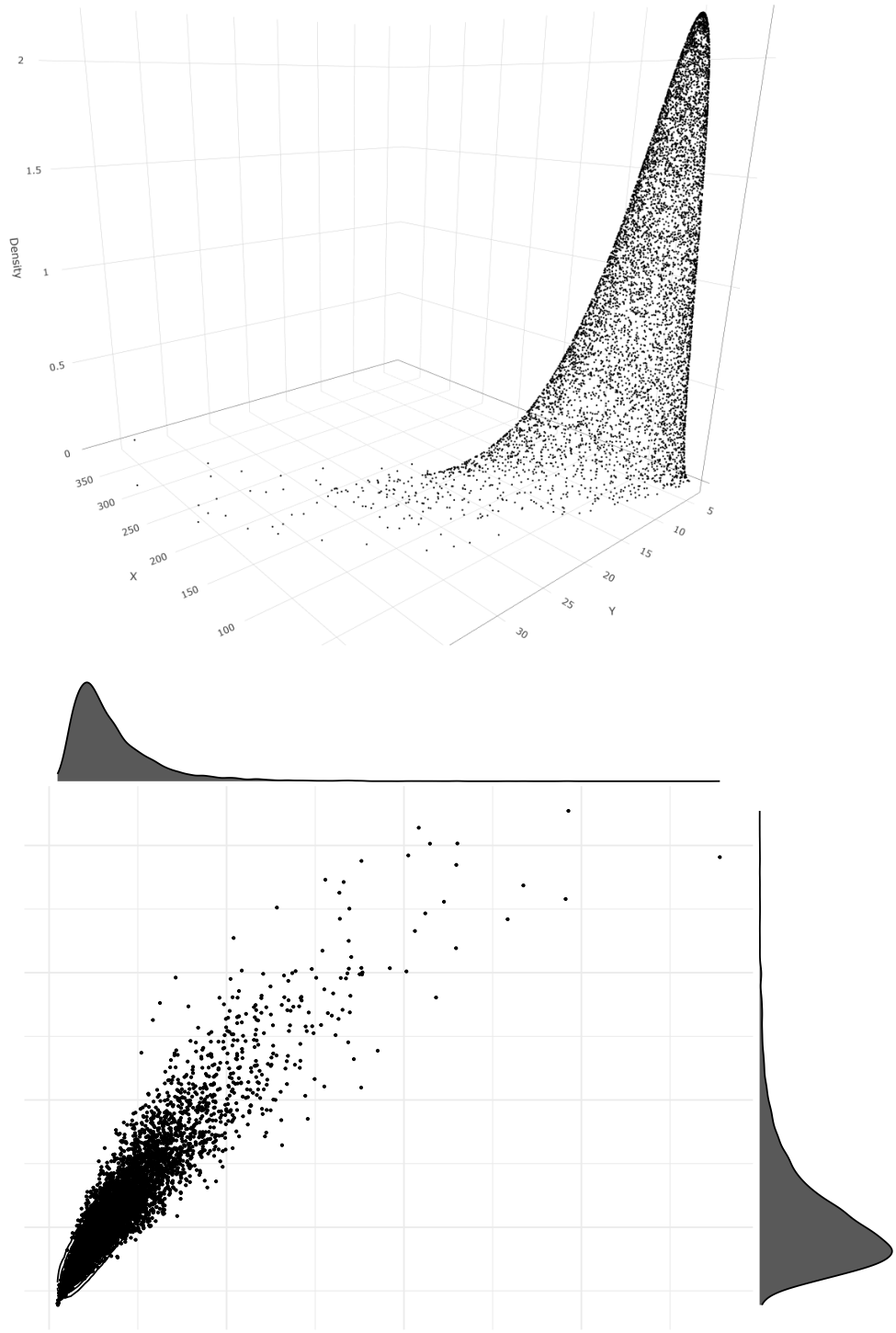


Figure 3.9: Density and scatter plots (with marginal densities) of the function $G_i(x) = \log(x)$ with covariance matrix $\mathbf{\Sigma} = \begin{pmatrix} \sigma_1^2 & \rho\sigma_1\sigma_2 \\ \rho\sigma_1\sigma_2 & \sigma_2^2 \end{pmatrix}$, where $\sigma_1 = 0.75, \sigma_2 = 0.5, \rho = 0.95, \mu = (3, 2), \lambda = (-5, 15), \tau = 1$.

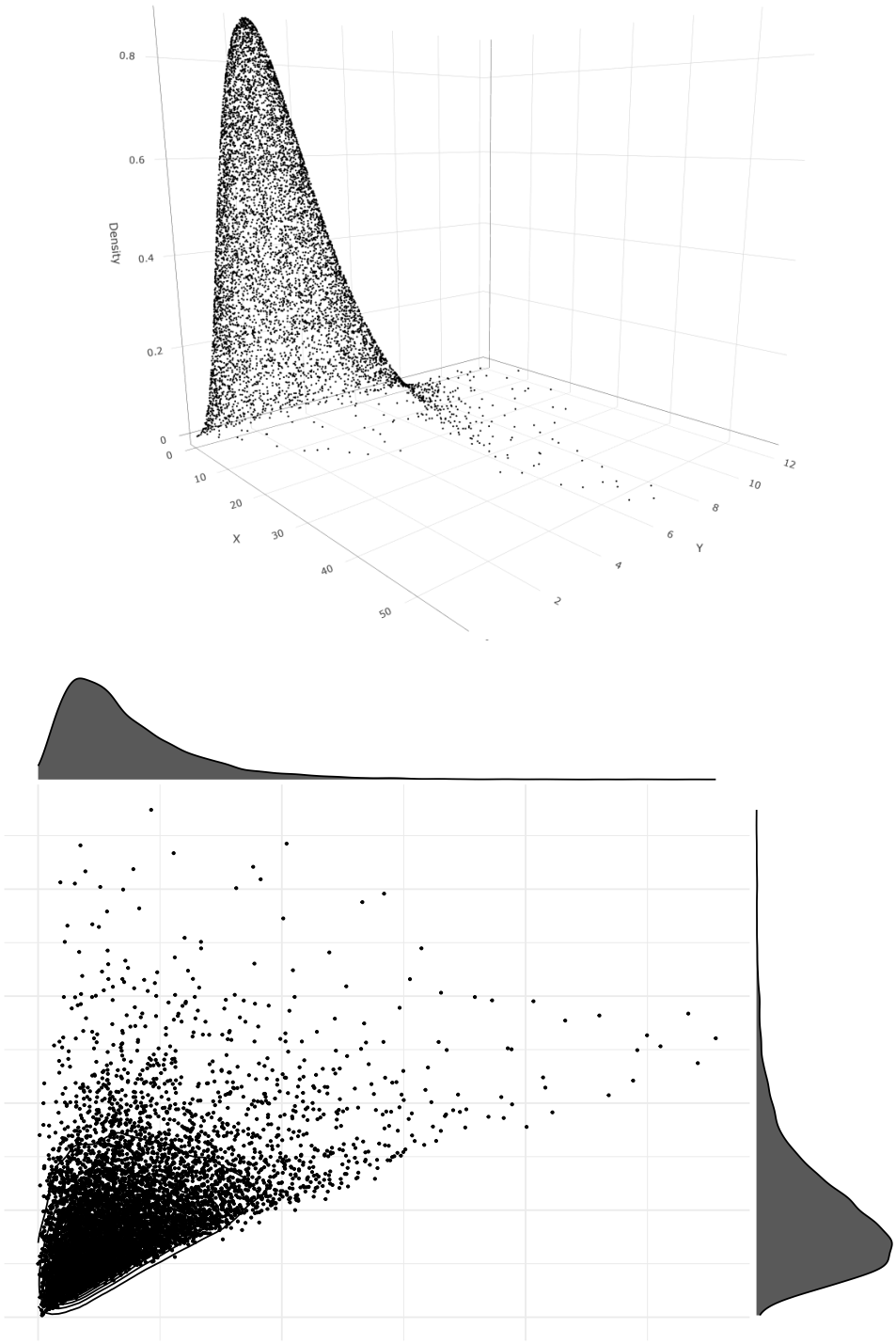


Figure 3.10: Density and scatter plots (with marginal densities) of the function $G_i(x) = \cosh^{-1}(x+1)$ with covariance matrix $\mathbf{\Sigma} = \begin{pmatrix} \sigma_1^2 & \rho\sigma_1\sigma_2 \\ \rho\sigma_1\sigma_2 & \sigma_2^2 \end{pmatrix}$, where $\sigma_1 = 0.75, \sigma_2 = 0.5, \rho = 0.1, \mu = (3, 1.5), \lambda = (-13, 22), \tau = -1$.

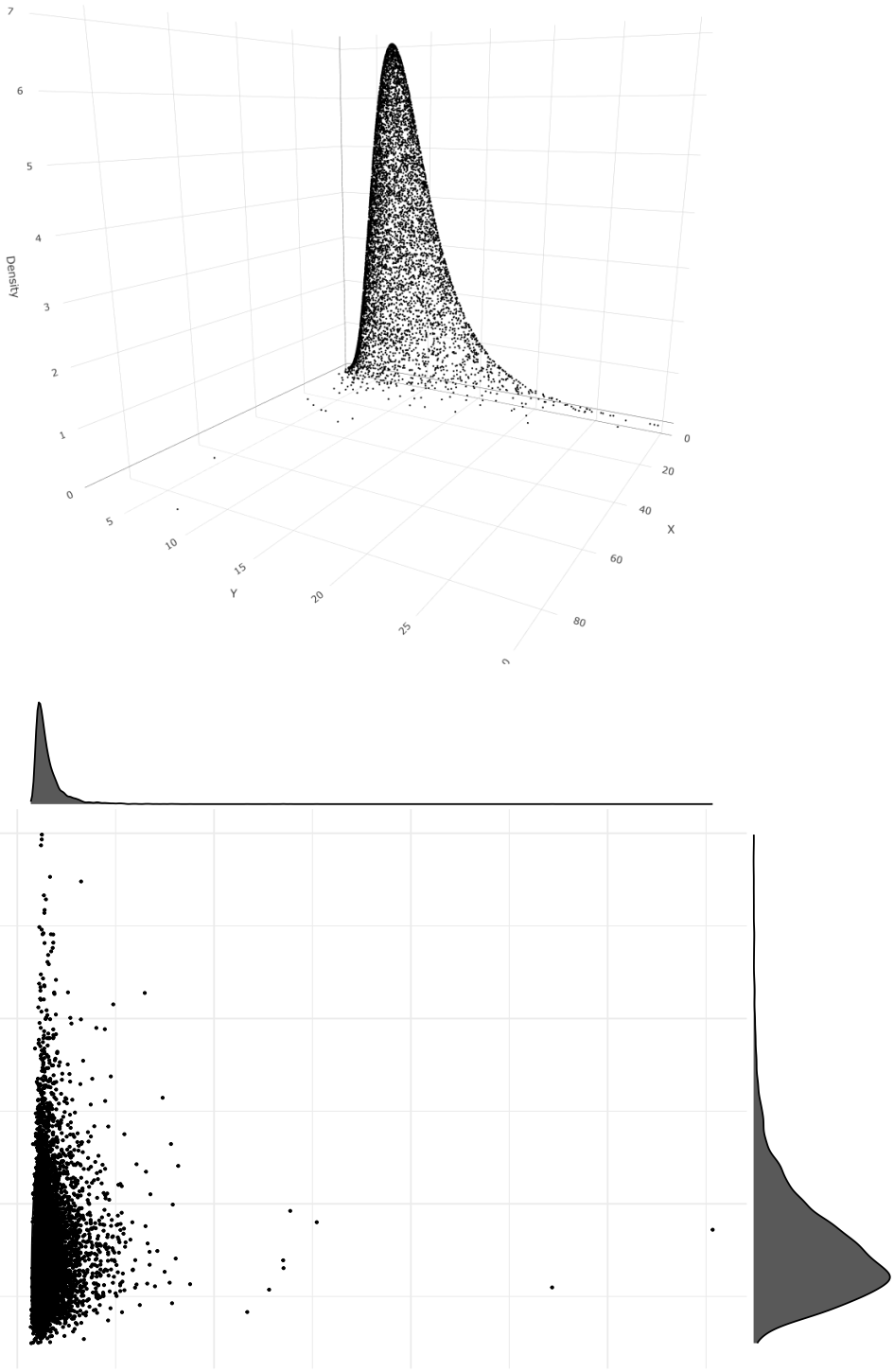


Figure 3.11: Density and scatter plots (with marginal densities) of the function $G_i(x) = \log(\log(x+1))$ with covariance matrix $\mathbf{\Sigma} = \begin{pmatrix} \sigma_1^2 & \rho\sigma_1\sigma_2 \\ \rho\sigma_1\sigma_2 & \sigma_2^2 \end{pmatrix}$, where $\sigma_1 = 0.3, \sigma_2 = 0.15, \rho = 0.01, \mu = (0.15, 0.75), \lambda = (20, -5), \tau = -1$.

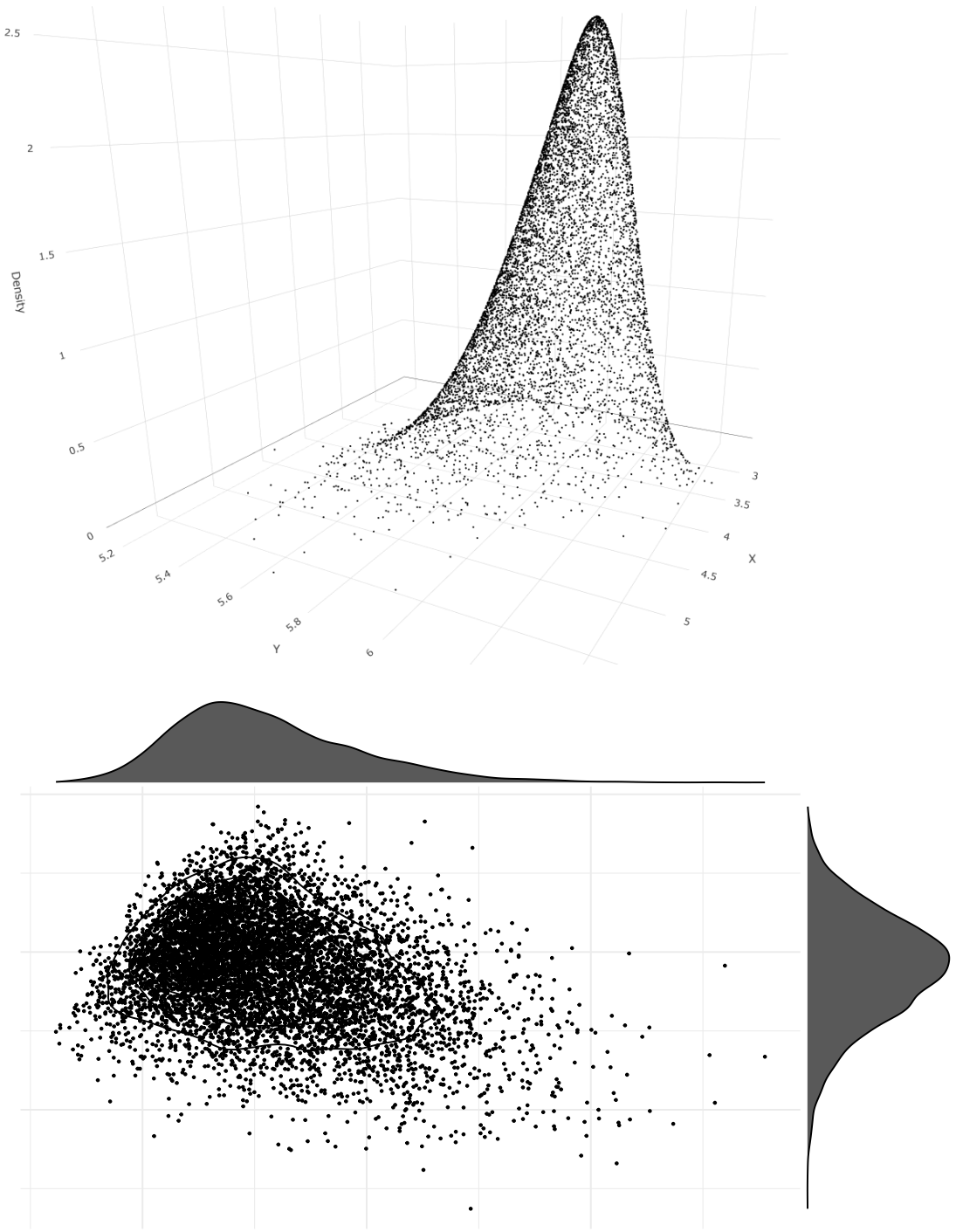


Figure 3.12: Density and scatter plots (with marginal densities) of the function $G_i(x) = x - \frac{1}{x}$ with covariance matrix $\mathbf{\Sigma} = \begin{pmatrix} \sigma_1^2 & \rho\sigma_1\sigma_2 \\ \rho\sigma_1\sigma_2 & \sigma_2^2 \end{pmatrix}$, where $\sigma_1 = 0.75, \sigma_2 = 0.25, \rho = 0.70, \mu = (2.5, 6), \lambda = (7.5, -10), \tau = -3$.

Given the context of the distributions discussed previously, we will conduct simulations using the Extended G-skew-Cauchy and Extended G-skew-Student-t distributions. The simulations will be carried out varying the degrees of freedom from 1 to 9. It is worth noting that, when the degree of freedom is equal to 1, the distribution reduces to the particular case of Cauchy, and for greater degrees of freedom, the distribution approaches the Student-t. To simplify the analysis, we will maintain the same parameters as the previous graphs for each applied function. The only function that will not be demonstrated is $loq_q(x)$, as this is generalized by the function $\log(x)$ when $q = 1$.

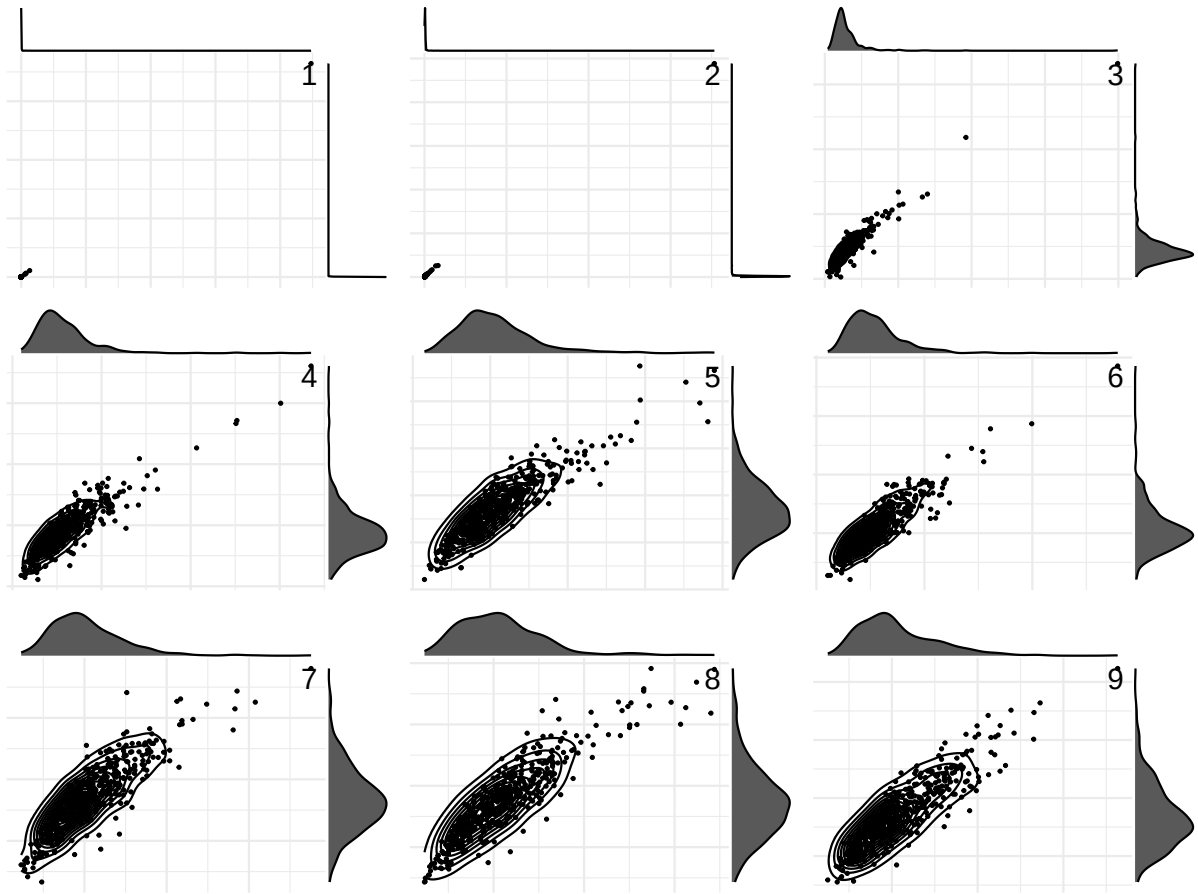


Figure 3.13: $G_i(x) = \frac{1}{\alpha}\left(\sqrt{\frac{x}{\beta}} - \sqrt{\frac{\beta}{x}}\right)$.

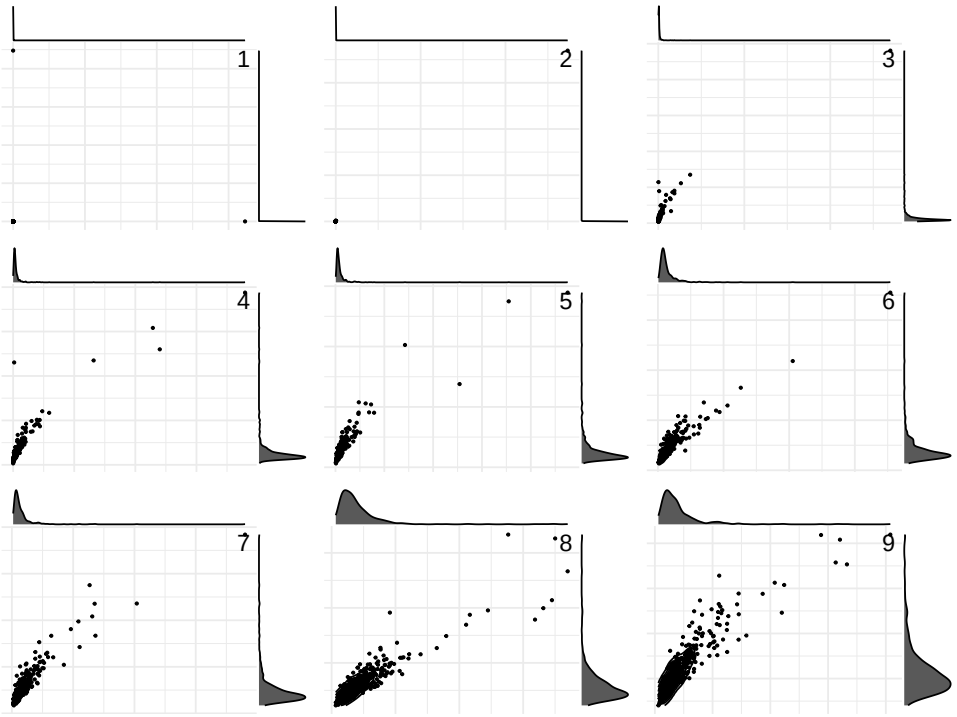


Figure 3.14: $G_i(x) = \log(x)$.

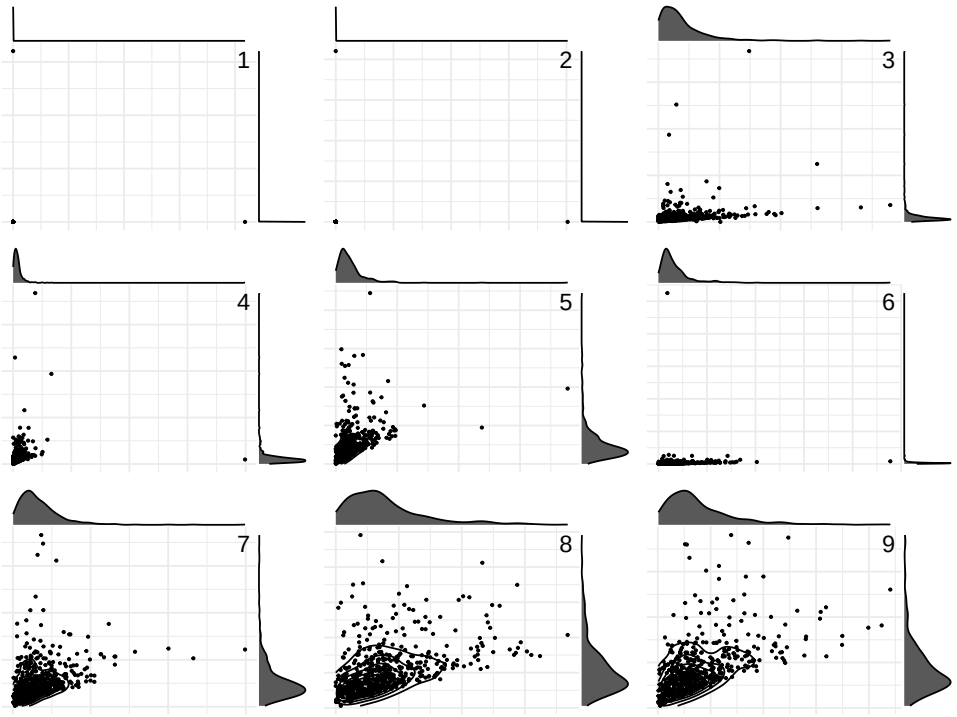


Figure 3.15: $G_i(x) = \cosh^{-1}(x+1)$.

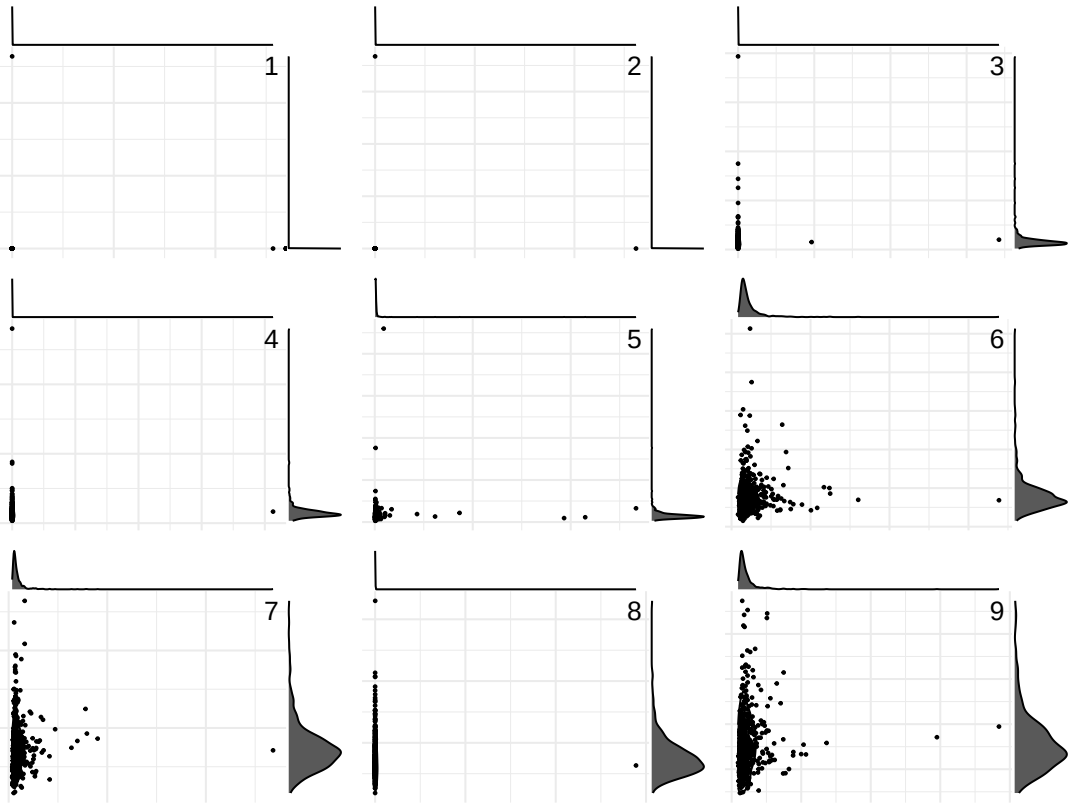


Figure 3.16: $G_i(x) = \log(\log(x+1))$.

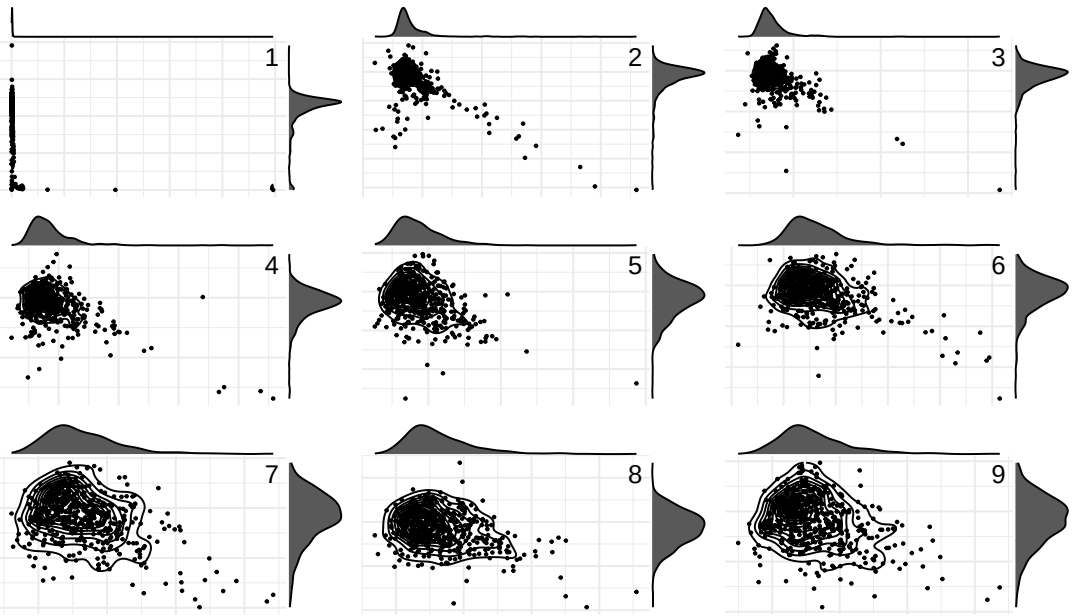


Figure 3.17: $G_i(x) = x - \frac{1}{x}$.

In the Cauchy distribution, samples are heavily concentrated around the center, making central points difficult to see and allowing extreme values to have a significant influence on the tails of marginal distributions. This behavior is especially useful for modeling data with heavy clustering and the presence of prominent outliers. As the degrees of freedom increase, there is a greater dispersion of the points, reducing the central concentration and causing the distribution to gradually approach normal. With higher degrees of freedom, the distribution becomes almost identical to the Extended G-skew-normal, with less pronounced tails and less impact of extreme values. This versatility demonstrates the potential of the distribution to adjust to a wide variety of data modeling situations, from scenarios with high concentration to contexts that present normal variability with asymmetries present.

3.2.3 Marginal Quantiles

Let $Y_i = T_i|\lambda^T(X-\mu)+\tau > Z; i = 1, ..., n$ be, 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 $\mathbf{Y} = (Y_1, ..., Y_n)^T$) denoted by $Q_{Y_i}(p)$, is a real number such that:

$$\mathbb{P}(Y_i \leq Q_{Y_i}(p)) = p, \quad i = 1, ..., n. \tag{3.2.13}$$

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, following way:

$$p = \mathbb{P}(Y_i \leq Q_{Y_i}(p)) = \mathbb{P}(T_i \leq Q_{Y_i}(p)|\lambda^T(X-\mu)+\tau > Z)$$

$$= \mathbb{P}(X_i \leq G_i(Q_{Y_i}(p))|\lambda^T(X-\mu)+\tau > Z)$$

$$= \mathbb{P}(W_i \leq G_i(Q_{Y_i}(p))).$$

Hence,

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

In other words, the relationship between the p quantiles for Y_i and W_i is bidirectional: the p quantile for Y_i is determined by the corresponding p quantile of W_i, and vice versa. Thus, it is sufficient to investigate the distribution of W_i to establish the p quantile for Y_i. Fortunately, the distribution of W_i for the cases studied in this work has already been characterized, as discussed in the Section dedicated to Conditional Distributions.

3.2.4 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 $\boldsymbol{Y} = (Y_1, \ldots, Y_n)^\top \sim \text{EGSE}_n(\boldsymbol{\mu}, \boldsymbol{\Sigma}, \boldsymbol{\lambda}, \tau, g^{(n)})$, with $\kappa > 0$. 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 (3.2.5), Bayes' rule gives

$$f_{Y_i \,|\, Y_j > \kappa}(y) = f_{Y_i}(y) \, \frac{\displaystyle\int_\kappa^\infty f_{Y_j \,|\, Y_i = y}(s) \mathrm{d}s}{\mathbb{P}(Y_j > \kappa)}, \quad y > 0, \ \kappa > 0. \tag{3.2.14}$$

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

$$f_{Y_j \,|\, Y_i = y}(s) = f_{W_j \,|\, W_i = G_i(y)}(G_j(s)) \, G_j'(s). \tag{3.2.15}$$

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 (3.2.14) and (3.2.15) 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,

$$f_{Y_i \,|\, Y_j>\kappa}(y) = f_{W_i}(G_i(y))G_i'(y)\,\frac{S_{W_j \,|\, W_i=G_i(y)}(G_j(\kappa))}{S_{W_j}(G_j(\kappa))}, \quad y>0,\ \kappa>0, \tag{3.2.16}$$

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 (3.2.16) 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 3.2), be the Student-t density generator of the EGSE$_n$ (multivariate extended 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 see Arellano-Valle and Genton, 2010a*

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

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, 2010a,

the unconditional and conditional distributions of $\boldsymbol{W}$ are respectively given by

$$W_i \sim \mathrm{EST}_1\left(\mu_i,\, \sigma_i^2,\, \frac{\lambda_i\sigma_i + \lambda_j\sigma_j\rho_{ij}}{\sigma_i\sqrt{1+\lambda_j^2\sigma_j^2(1-\rho_{ij}^2)}},\, \nu,\, \frac{\tau}{\sqrt{1+\lambda_j^2\sigma_j^2(1-\rho_{ij}^2)}}\right), \tag{3.2.17}$$

$$W_j \sim \mathrm{EST}_1\left(\mu_j,\, \sigma_j^2,\, \frac{\lambda_j\sigma_j + \lambda_i\sigma_i\rho_{ij}}{\sigma_j\sqrt{1+\lambda_i^2\sigma_i^2(1-\rho_{ij}^2)}},\, \nu,\, \frac{\tau}{\sqrt{1+\lambda_i^2\sigma_i^2(1-\rho_{ij}^2)}}\right), \tag{3.2.18}$$

and

$$W_j \,|\, W_i = y \sim \mathrm{EST}_1\left(\boldsymbol{\mu}_y,\, \boldsymbol{\sigma}_{y;\nu}^2,\, \lambda_j\sigma_j\sqrt{1-\rho_{ij}^2},\, \nu+1,\, \boldsymbol{\tau}_{y;\nu}\right), \tag{3.2.19}$$

where we are adopting the following notation:

$$\begin{aligned}
\boldsymbol{\mu}_y &= \mu_j + \sigma_j\rho_{ij}\left(\frac{y-\mu_i}{\sigma_i}\right);\\
\boldsymbol{\sigma}_{y;\nu}^2 &= \frac{\nu + \left(\frac{y-\mu_{1i}}{\sigma_i}\right)^2}{\nu+1}\,\sigma_j^2(1-\rho_{ij}^2);\\
\boldsymbol{\tau}_{y;\nu} &= \left[(\lambda_i\sigma_i + \lambda_j\sigma_j\rho_{ij})\left(\tfrac{y-\mu_i}{\sigma_i}\right) + \tau\right]\sqrt{\tfrac{\nu+1}{\nu+\left(\frac{y-\mu_i}{\sigma_i}\right)^2}}.
\end{aligned} \tag{3.2.20}$$

Hence, by combining (3.2.16) with (3.2.18), (3.2.19) and (3.2.20), we obtain

$$\begin{aligned}
f_{Y_i\,|\,Y_j>\kappa}(y) = f_{\mathrm{EST}_1}&\left(G_i(y);\, \mu_i,\, \sigma_i^2,\, \frac{\lambda_i\sigma_i + \lambda_j\sigma_j\rho_{ij}}{\sigma_i\sqrt{1+\lambda_j^2\sigma_j^2(1-\rho_{ij}^2)}},\, \nu,\, \frac{\tau}{\sqrt{1+\lambda_j^2\sigma_j^2(1-\rho_{ij}^2)}}\right) G_i'(y)\\
&\times \frac{S_{\mathrm{EST}_1}\left(G_j(\kappa);\, \boldsymbol{\mu}_{G_i(y)},\, \boldsymbol{\sigma}^2_{G_i(y);\nu},\, \lambda_j\sigma_j\sqrt{1-\rho_{ij}^2},\, \nu+1,\, \boldsymbol{\tau}_{G_i(y);\nu}\right)}{S_{\mathrm{EST}_1}\left(G_j(\kappa);\, \mu_j,\, \sigma_j^2,\, \frac{\lambda_j\sigma_j + \lambda_i\sigma_i\rho_{ij}}{\sigma_j\sqrt{1+\lambda_i^2\sigma_i^2(1-\rho_{ij}^2)}},\, \nu,\, \frac{\tau}{\sqrt{1+\lambda_i^2\sigma_i^2(1-\rho_{ij}^2)}}\right)},
\end{aligned} \tag{3.2.21}$$

Example 3.2.1. *By taking $G_i(x) = \log(x)$, $x > 0$, $i = 1, \ldots, n$ (see Table 3.2), we get the* for $y > 0$ and $\kappa > 0$.
multivariate asymmetric version of the Student-t model addressed in Marchenko and Genton, 2010. So, from (3.2.21) and Table 3.1, we have

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

for $y > 0$, $\kappa > 0$, and $\boldsymbol{\mu}_y$, $\boldsymbol{\sigma}^2_{y;\nu}$ and $\boldsymbol{\tau}_{y;\nu}$ being as in (3.2.20).

Cauchy density generator

Let $g^{(n)}(x) = 1/(1+x)^{(n+1)/2}$, $x \in \mathbb{R}$ (see Table 3.2), be the Cauchy density generator of the EGSE$_n$ (multivariate extended G-skew-Cauchy) distribution.

By taking $\nu = 1$ in formula (3.2.21), we have

$$f_{Y_i \,|\, Y_j > \kappa}(y) = f_{\mathrm{EST}_1}\left(G_i(y);\, \mu_i,\, \sigma_i^2,\, \frac{\lambda_i\sigma_i + \lambda_j\sigma_j\rho_{ij}}{\sigma_i\sqrt{1+\lambda_j^2\sigma_j^2(1-\rho_{ij}^2)}},\, 1,\, \frac{\tau}{\sqrt{1+\lambda_j^2\sigma_j^2(1-\rho_{ij}^2)}}\right)G_i'(y)$$
$$\times \frac{S_{\mathrm{EST}_1}\left(G_j(\kappa);\, \boldsymbol{\mu}_{G_i(y)},\, \boldsymbol{\sigma}^2_{G_i(y);1},\, \lambda_j\sigma_j\sqrt{1-\rho_{ij}^2},\, 2,\, \boldsymbol{\tau}_{G_i(y);1}\right)}{S_{\mathrm{EST}_1}\left(G_j(\kappa);\, \mu_j,\, \sigma_j^2,\, \frac{\lambda_j\sigma_j + \lambda_i\sigma_i\rho_{ij}}{\sigma_j\sqrt{1+\lambda_i^2\sigma_i^2(1-\rho_{ij}^2)}},\, 1,\, \frac{\tau}{\sqrt{1+\lambda_i^2\sigma_i^2(1-\rho_{ij}^2)}}\right)}, \tag{3.2.22}$$

for $y > 0$ and $\kappa > 0$.

Example 3.2.2. *By taking $G_i(x) = x - 1/x$, $x > 0$, $i = 1, \ldots, n$, from (3.2.22) and Table 3.1,*

we have

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

for $y>0$, $\kappa>0$, and $\boldsymbol{\mu}_y$, $\boldsymbol{\sigma}^2_{y;\nu}$ and $\boldsymbol{\tau}_{y;\nu}$ being as in (3.2.20).

Gaussian density generator

Let $g^{(n)}(x)=\exp(-x)$, $x\in\mathbb{R}$ (see Table 3.2), be the Gaussian density generator of the EGSE_n (multivariate extended 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 see Vernic, 2005; Arellano-Valle and Genton, 2010a*

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

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 (3.2.21), we

obtain

$$f_{Y_i \,|\, Y_j>\kappa}(y) = f_{\mathrm{ESN}_1}\left(G_i(y);\, \mu_i,\, \sigma_i^2,\, \frac{\lambda_i\sigma_i + \lambda_j\sigma_j\rho_{ij}}{\sigma_i\sqrt{1+\lambda_j^2\sigma_j^2(1-\rho_{ij}^2)}},\, \nu,\, \frac{\tau}{\sqrt{1+\lambda_j^2\sigma_j^2(1-\rho_{ij}^2)}}\right) G_i'(y)$$
$$\times \frac{S_{\mathrm{ESN}_1}\left(G_j(\kappa);\, \mu_j + \sigma_j\rho_{ij}\left(\frac{G_i(y)-\mu_i}{\sigma_i}\right),\, \sigma_j^2(1-\rho_{ij}^2),\, \lambda_j\sigma_j\sqrt{1-\rho_{ij}^2},\, (\lambda_i\sigma_i + \lambda_j\sigma_j\rho_{ij})\left(\frac{G_i(y)-\mu_i}{\sigma_i}\right) + \tau\right)}{S_{\mathrm{ESN}_1}\left(G_j(\kappa);\, \mu_j,\, \sigma_j^2,\, \frac{\lambda_j\sigma_j + \lambda_i\sigma_i\rho_{ij}}{\sigma_j\sqrt{1+\lambda_i^2\sigma_i^2(1-\rho_{ij}^2)}},\, \nu,\, \frac{\tau}{\sqrt{1+\lambda_i^2\sigma_i^2(1-\rho_{ij}^2)}}\right)}, \tag{3.2.23}$$

for $y > 0$ and $\kappa > 0$.

Example 3.2.3. *By taking $G_i(x) = (\sqrt{\beta/x} - \sqrt{x/\beta})/\alpha$, $x > 0$, $i = 1, \ldots, n$, from (3.2.23) and Table 3.1, we have*

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

for $y > 0$ and $\kappa > 0$.

3.2.5 Existence of marginal moments

The objective of this subsection is to provide sufficient conditions to ensure the existence of the real moments of the random variable $Y_i = T_i \,|\, \boldsymbol{\lambda}^\top(\boldsymbol{X} - \boldsymbol{\mu}) + \tau > Z$, with $T_i = G_i^{-1}(X_i)$, $i = 1, \ldots, n$. To do this, we will consider the notation $W_i = X_i \,|\, \boldsymbol{\lambda}^\top\boldsymbol{X} + \tau > Z$, $i = 1, \ldots, n$, used in Subsection 3.2.13.

Indeed, by using the well-known identity

$$\mathbb{E}(Y^p) = p\int_0^\infty y^{p-1}\mathbb{P}(Y > y)\mathrm{d}y, \quad Y > 0,\ p > 0, \tag{3.2.24}$$

and by employing the relation

$$Y_i \stackrel{d}{=} G_i^{-1}(W_i), \quad i = 1, \ldots, n,$$

with $\stackrel{d}{=}$ denoting equality in distribution, it follows that

$$\begin{aligned}
\mathbb{E}(Y_i^p) &= p \int_0^\infty y^{p-1} \mathbb{P}(W_i > G_i(y)) \mathrm{d}y \\
&= p \int_0^a y^{p-1} \mathbb{P}(W_i > G_i(y)) \mathrm{d}y + p \int_a^\infty y^{p-1} \mathbb{P}(W_i > G_i(y)) \mathrm{d}y \\
&\leqslant a^p + p \int_a^\infty y^{p-1} \mathbb{P}(W_i > G_i(y)) \mathrm{d}y,
\end{aligned}$$

for some $a \in (0, \infty)$. Therefore, a sufficient condition for the existence of positive order moments of Y_i is that

$$I = \int_a^\infty y^{p-1} \mathbb{P}(W_i > G_i(y)) \mathrm{d}y < \infty, \quad i = 1, \ldots, n. \tag{3.2.25}$$

In what remains of this subsection we will analyze condition in (3.2.25) in the special case that (see Table 3.1)

$$G_i(x) = \frac{2F_i(x) - 1}{F_i(x)[1 - F_i(x)]}, \quad x > 0, \ i = 1, \ldots, n, \tag{3.2.26}$$

with F_i being the CDF of a continuous random variable with positive support. Indeed, as $\{W_i > G_i(y)\} \subset \{|W_i| > G_i(y)\}$, the integral in (3.2.25) is

$$I \leqslant \int_a^\infty y^{p-1} \mathbb{P}(|W_i| > G_i(y)) \mathrm{d}y.$$

By Markov's inequality, the above expression is at most

$$\mathbb{E}(|W_i|^p)\int_a^\infty \frac{y^{p-1}}{G_i^p(y)}\mathrm{d}y = \mathbb{E}(|W_i|^p)\int_a^\infty \frac{y^{p-1}}{G_i^{p-1}(y)}\frac{F_i(y)[1-F_i(y)]}{[2F_i(y)-1]}\mathrm{d}y.$$

As G_i and F_i are increasing, for $p>1$, the above expression is

$$\begin{aligned}
&\leqslant \frac{\mathbb{E}(|W_i|^p)}{G_i^{p-1}(a)[2F_i(a)-1]}\int_a^\infty y^{p-1}[1-F_i(y)]\mathrm{d}y\\
&\leqslant \frac{\mathbb{E}(|W_i|^p)}{G_i^{p-1}(a)[2F_i(a)-1]}\int_0^\infty y^{p-1}[1-F_i(y)]\mathrm{d}y,
\end{aligned}$$

provided $F_i(a)\neq 1/2$ and $G_i(a)\in(0,\infty)$. If $S>0$ is a continuous random variable such that $S_i \overset{d}{=} F_i$, by (3.2.24), the above integral is

$$= \frac{\mathbb{E}(|W_i|^p)\mathbb{E}(S_i^p)}{pG_i^{p-1}(a)[2F_i(a)-1]}.$$

Therefore, for the choice of G_i as in (3.2.26), we have verified that

$$I \leqslant \frac{\mathbb{E}(|W_i|^p)\mathbb{E}(S_i^p)}{pG_i^{p-1}(a)[2F_i(a)-1]}.$$

Hence, if G_i as in (3.2.26), $a>0$ is such that $F_i(a)\neq 1/2$ and $G_i(a)\in(0,\infty)$, $\mathbb{E}(|W_i|^p)<\infty$ and $\mathbb{E}(S_i^p)<\infty$ for some $p>1$, then $\mathbb{E}(Y_i^p)$, $i=1,\ldots,n$, exists.

3.2.6 Mahalanobis distance

To evaluate the residuals of the model, we propose a metric that uses the Mahalanobis distance. This approach consists of individually checking whether each data point belongs to the proposed distribution, based on a threshold distance ϵ. If the distance of a point in relation to the center of the distribution is within this threshold, the point is considered part of the distribution, indicating that the data follow the distribution EGSE$_n(\boldsymbol{\mu},\boldsymbol{\Sigma},\boldsymbol{\lambda},\tau,g^{(n)})$. Thus, this metric allows

a robust assessment of whether the observed points are aligned with the theoretical distribution, ensuring a more detailed analysis.

$$\boldsymbol{Y} = \boldsymbol{T}|\boldsymbol{\lambda}^T(\boldsymbol{X} - \boldsymbol{\mu}) + \tau > Z, \quad \boldsymbol{T} = (G_1^{-1}(X_1), ..., G_n^{-1}(X_n)).$$

Let be $Z^* = \boldsymbol{\lambda}^T(\boldsymbol{X} - \boldsymbol{\mu}) + \tau - Z$, then $\boldsymbol{Y} = T|Z^* > 0$.

We can define the Mahalanobis distance in terms of $\boldsymbol{Y}$:

$$d^2 = (\boldsymbol{Y} - \boldsymbol{\mu}_Y)^T \boldsymbol{\Sigma}_y^{-1}(\boldsymbol{Y} - \boldsymbol{\mu}_Y); \quad \boldsymbol{\mu}_Y = (\mu_{y_1}, ..., \mu_{y_n})^T.$$

We goal is calculate of accumulate function of d^2. In this sense it's can be define

$$F_{d^2}(x) = \mathbb{P}(d^2 \leq x); \quad f.d.a; \quad \forall x \in \mathbb{R}.$$

We can to use the Cholesky decomposition to describe the covariance matrix:

$$\boldsymbol{\Sigma_Y} = \boldsymbol{L}\boldsymbol{L}^T,$$

remember that covariance matrix is symmetric and positive and L is inferior triangular matrix. Therefore

$$\boldsymbol{\Sigma}_Y^{-1} = (\boldsymbol{L}\boldsymbol{L}^T)^{-1} = (\boldsymbol{L}^T)^{-1}\boldsymbol{L}^{-1} = (\boldsymbol{L}^{-1})^T\boldsymbol{L}^{-1}.$$

Hence

$$d^2 = (\boldsymbol{Y}-\boldsymbol{\mu}_y)^T(\boldsymbol{L}^{-1})^T\boldsymbol{L}^{-1}(\boldsymbol{Y}-\boldsymbol{\mu}_y) = \boldsymbol{L}^{-1}(\boldsymbol{Y}-\boldsymbol{\mu}_Y)^T\boldsymbol{L}^{-1}(\boldsymbol{Y}-\boldsymbol{\mu}_Y) = \boldsymbol{S}^T\boldsymbol{S} = \sum_{i=1}^{n} S_i^2 = ||S||^2,$$

where $\boldsymbol{S} = (S_1, ..., S_n)^T = \boldsymbol{L}^{-1}(\boldsymbol{Y} - \boldsymbol{\mu}_Y)$ and $||\cdot||$ denote the euclidean norm.

Let $\mathbf{L^{-1}}$ the triangular matrix originating the Cholesky decomposition $n \times n$ and let $\boldsymbol{Y} - \boldsymbol{\mu}_Y$ be a vetor $n \times 1$:

$$\mathbf{L^{-1}} \cdot (\mathbf{Y} - \mu_{\mathbf{Y}}) = \begin{bmatrix} a_{11} & 0 & \cdots & 0 \\ a_{21} & a_{22} & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ a_{n1} & a_{n2} & \cdots & a_{nn} \end{bmatrix} \cdot \begin{bmatrix} Y_1 - \mu_{Y_1} \\ Y_2 - \mu_{Y_2} \\ \vdots \\ Y_n - \mu_{Y_n} \end{bmatrix}. \tag{3.2.27}$$

The product in (3.2.27) results in a vector $\mathbf{S}$ given by:

$$\mathbf{S} = \begin{bmatrix} S_1 \\ S_2 \\ \vdots \\ S_n \end{bmatrix} = \begin{bmatrix} a_{11}(Y_1 - \mu_{Y_1}) \\ a_{21}(Y_1 - \mu_{Y_1}) + a_{22}(Y_2 - \mu_{Y_2}) \\ \vdots \\ a_{n1}(Y_1 - \mu_{Y_1}) + ... + a_{nn}(Y_n - \mu_{Y_n}) \end{bmatrix}. \tag{3.2.28}$$

Then by (3.2.28) there is $c_{ij} : 1 \leq j \leq i \leq n$ such that

$$S_1 = c_{11}(Y_1 - \mu_{Y_1}),$$

$$S_2 = c_{21}S_1 + c_{22}(Y_2 - \mu_{Y_2}),$$

$$\vdots$$

$$S_{n-1} = \sum_{i=1}^{n-2} c_{(n-1)i}S_i + c_{(n-1)(n-1)}(Y_{n-1} - \mu_{Y_{n-1}}),$$

$$S_n = \sum_{i=1}^{n-1} c_{ni}S_i + c_{nn}(Y_n - \mu_{Y_n}).$$

Now we can define the accumulate function by Mahalanobis square distance

$$F_{d^2}(x) = \mathbb{P}(d^2 \leq x)$$

$$= \mathbb{P}\Big(\sum_{i=1}^{n} S_i^2 \leq x\Big) = \mathbb{E}[\mathbb{1}_{\{\sum_{i=1}^{n} S_i^2 \leq x\}}].$$

Applying the Law of total expectation we have

$$\mathbb{P}\Big(\sum_{i=1}^{n} S_i^2 \leq x\Big) = \mathbb{E}[\mathbb{1}_{\{\sum_{i=1}^{n} S_i^2 \leq x\}}] = \mathbb{E}[\mathbb{E}(\mathbb{1}_{\{\sum_{i=1}^{n} S_i^2 \leq x\}}|S_1, ..., S_{n-1})]$$

$$= \mathbb{E}[\mathbb{E}(\mathbb{1}_{\{\sum_{i=1}^{n} S_i^2 \leq x\}}|S_1, ..., S_{n-1})] = \int ... \int_{\mathbb{R}^{n-1}} \mathbb{E}(\mathbb{1}_{\{\sum_{i=1}^{n} S_i^2 \leq x\}}|S_1, ..., S_{n-1}) dF_{S_1,...,S_{n-1}}$$

$$= \int ... \int_{\mathbb{R}^{n-1}} \mathbb{E}(\mathbb{1}_{\{S_n^2 \leq x - \sum_{i=1}^{n-1} S_i^2\}}|S_1, ..., S_{n-1}) dF_{S_1,...,S_{n-1}}$$

$$= \int ... \int_{\sum_{i=1}^{n-1} \leq x} \mathbb{E}(\mathbb{1}_{\{-\sqrt{x - \sum_{i=1}^{n-1} S_i} \leq S_n \leq \sqrt{x - \sum_{i=1}^{n-1} S_i}\}}|S_1, ..., S_{n-1}) dF_{S_1,...,S_{n-1}}.$$

Although the mathematical development is solid and well-founded, the computational implementation of this concept has not yet been achieved. However, we have registered this proposal so that, in the future, with the advancement of new studies or techniques, it will be possible to overcome these limitations and evaluate the model as presented. It is important to emphasize that the main challenge lies in the calculation of multiple integrals of the conditional expectation, characterized as a highly complex and non-trivial expression. This record not only serves as a starting point for future investigations and improvements, but also as a reference for other researchers who may be interested in exploring or developing this methodology.

3.3 Maximum likelihood estimation

Consider $\{\boldsymbol{Y}_k = (Y_{1k}, Y_{2k}, \ldots, Y_{nk})^\top : k = 1, \ldots, m\}$ to be a multivariate random sample of size m from the distribution $\boldsymbol{Y} \sim \text{EUGSE}_n(\boldsymbol{\mu}, \boldsymbol{\Sigma}, \boldsymbol{\lambda}, \tau, g^{(n)})$, with the joint PDF given in the expression (3.1.1). Suppose $\boldsymbol{y}_k = (y_{1k}, y_{2k}, \ldots, y_{nk})^\top$ is an observed realization of $\boldsymbol{Y}_k$. To find the maximum likelihood estimates (MLEs) of the parameters of the model, whose parameter vector is $\boldsymbol{\theta} = (\boldsymbol{\mu}, \boldsymbol{\Sigma}, \boldsymbol{\lambda}, \tau)^\top$, it is necessary to maximize the following log-likelihood function

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

where $\boldsymbol{y}_{G,k} = (G_1(y_{1k}), \ldots, G_n(y_{nk}))^\top$. Since $\boldsymbol{X} \sim \text{ELL}_n(\boldsymbol{\mu}, \boldsymbol{\Sigma}, g^{(n)})$, by the expression 3.1.1, the log-likelihood function (neglecting the additive constant) is written as:

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

The likelihood equations are given by

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

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)

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

(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.

Chapter 4

Simulation study

In this section, the results obtained from Monte Carlo simulations will be presented and discussed, with the aim of evaluating the quality of the estimates related to the distribution under study. Simulations allow examining the performance of estimators under different scenarios, providing a detailed analysis of their accuracy, bias and variance. Through this approach, it will be possible to validate the robustness of the methods employed, as well as identify possible limitations, contributing to a more comprehensive understanding of the behavior of the distribution under different conditions.

"*All models are wrong, but some are useful*"

- George E. P. Box

4.1 Simulation Study

In this section, we conduct a simulation study to evaluate the performance of maximum likelihood estimators (MLEs). The simulation focuses on the estimation of the model parameters in the bivariate case. We present the results for the extended unit-G-skew normal distribution with the following G_i functions: $G_i(x) = \frac{1}{\alpha}\left(\sqrt{\frac{x}{\beta}} - \sqrt{\frac{\beta}{x}}\right)$, $G_i(x) = \log(x)$, $G_i(x) = \cosh^{-1}(x+1)$, $G_i(x) = \log(\log(x+1))$, and $G_i(x) = x - \frac{1}{x}$. In the context of the function $\log_q(x)$, the detailed estimates are available in the Appendix A.

The efficiency of MLEs is evaluated using two key metrics: relative bias (RB) and root mean square error (RMSE). These are calculated as follows:

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

where $\widehat{\theta}^{(i)}$ is the estimated value in the ith iteration and N is the number of simulations. Similarly, the RMSE is calculated by:

$$\widehat{RMSE}(\widehat{\theta}) = \sqrt{\frac{1}{N}\sum_{i=1}^{N}\left(\widehat{\theta}^{(i)} - \theta\right)^2}.$$

The simulation scenario employed samples of the following orders $n \in \{100, 200, 500, 1000\}$, where the true parameters are specified as follows

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

and ρ takes on values in $\{0.10, 0.25, 0.50, 0.75, 0.90\}$. In each case, 100 Monte Carlo replicates were performed for each configuration.

Figures 4.1–4.10 show the simulation results of the maximum likelihood estimation. It can be observed that the relative bias decreases with increasing sample size, which indicates that the

estimates are approaching the real values. For smaller ρ, the bias is larger, while larger values of ρ have a smaller bias, proving that the choice of ρ affects the precision of the estimates. The curves show a rapid decrease in bias with the increase in n, confirming the consistency of the estimators and emphasising the importance of large samples to improve accuracy.

Figures 4.1–4.10 also show that the RMSE decreases with increasing sample size, which indicates that the estimates become more accurate and are closer to the true values of the parameters. Larger ρ values tend to show smaller MSEs, while smaller ρ values show a larger error, reflecting the influence of ρ on the accuracy of the estimates. The curves show a strong decrease in RMSE with increasing n, especially for small samples, emphasising the importance of larger samples for more accurate and reliable estimates.

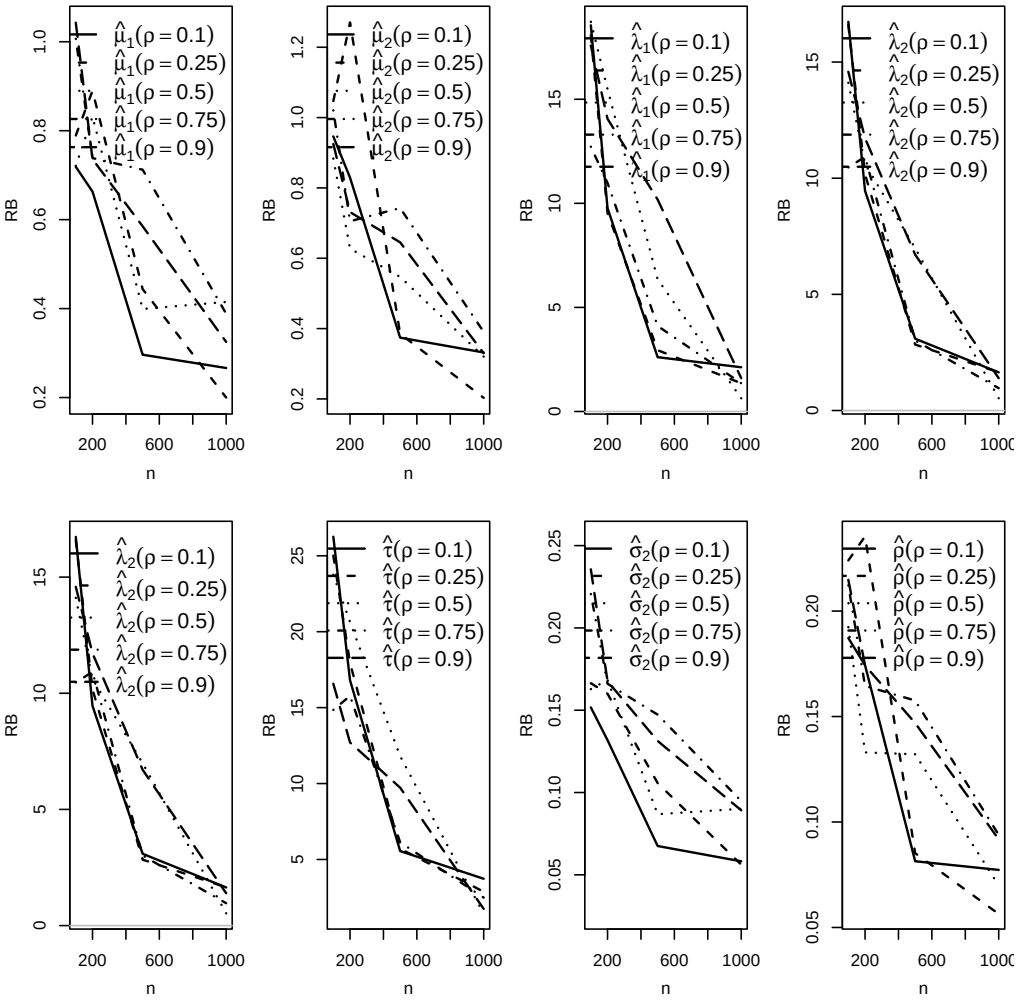


Figure 4.1: Relative bias for $G_i(x) = \frac{1}{\alpha}\left(\sqrt{\frac{x}{\beta}} - \sqrt{\frac{\beta}{x}}\right)$.

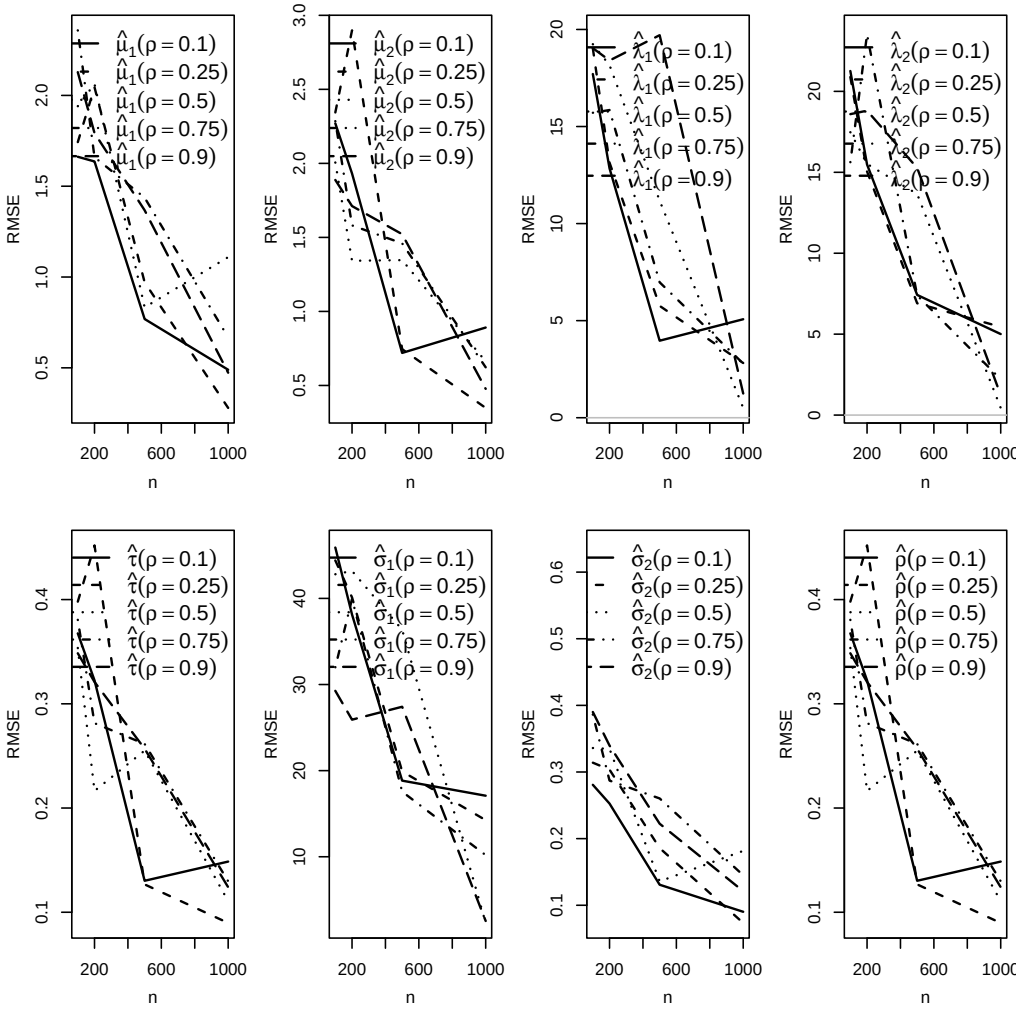


Figure 4.2: Root mean squared error for $G_i(x) = \frac{1}{\alpha}\left(\sqrt{\frac{x}{\beta}} - \sqrt{\frac{\beta}{x}}\right)$.

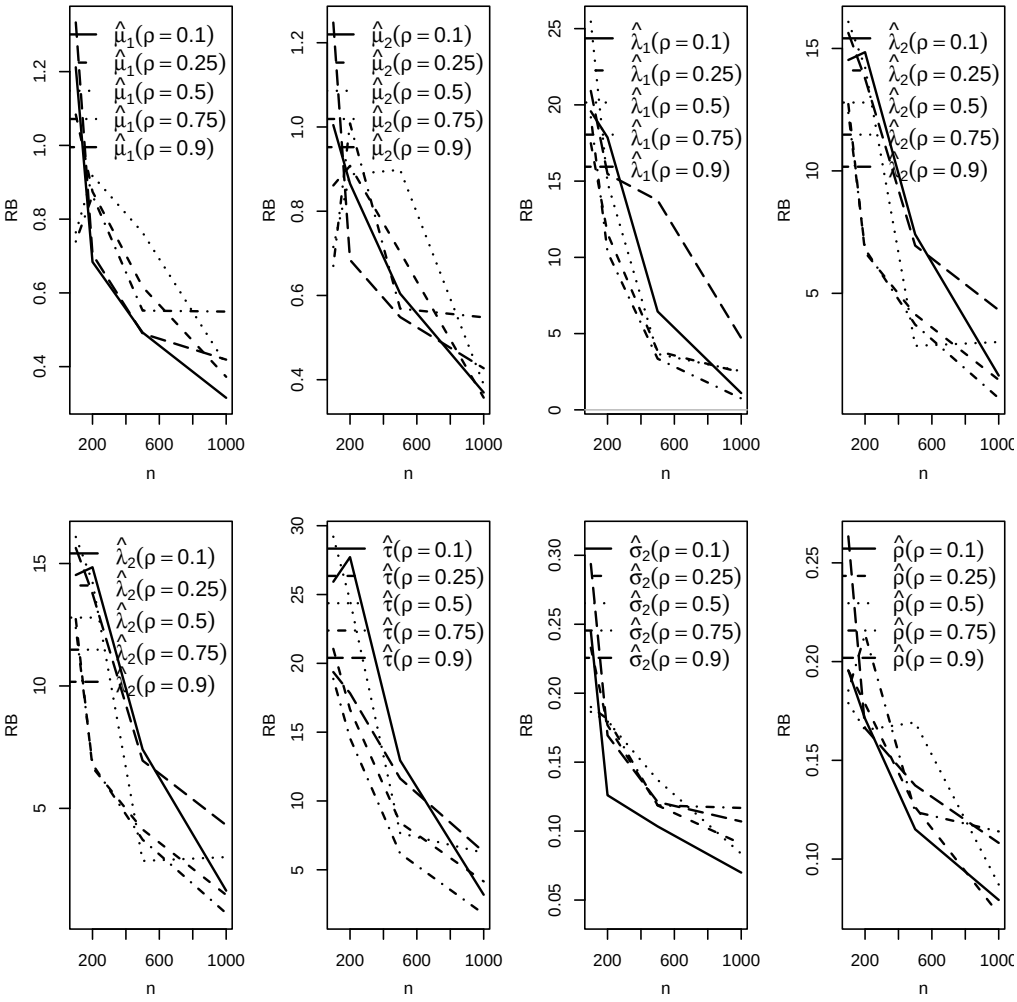


Figure 4.3: Relative bias for $G_i(x) = \log(x)$.

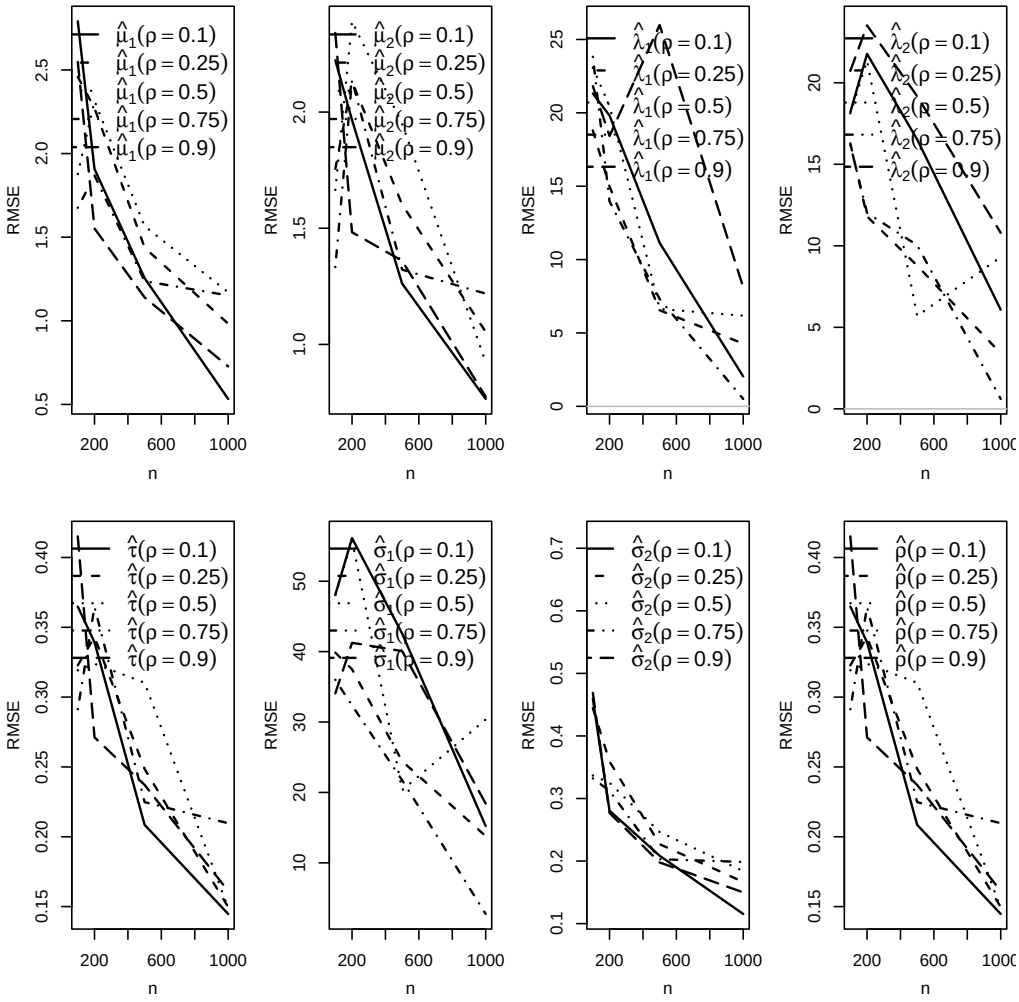


Figure 4.4: Root mean squared error for $G_i(x) = \log(x)$.

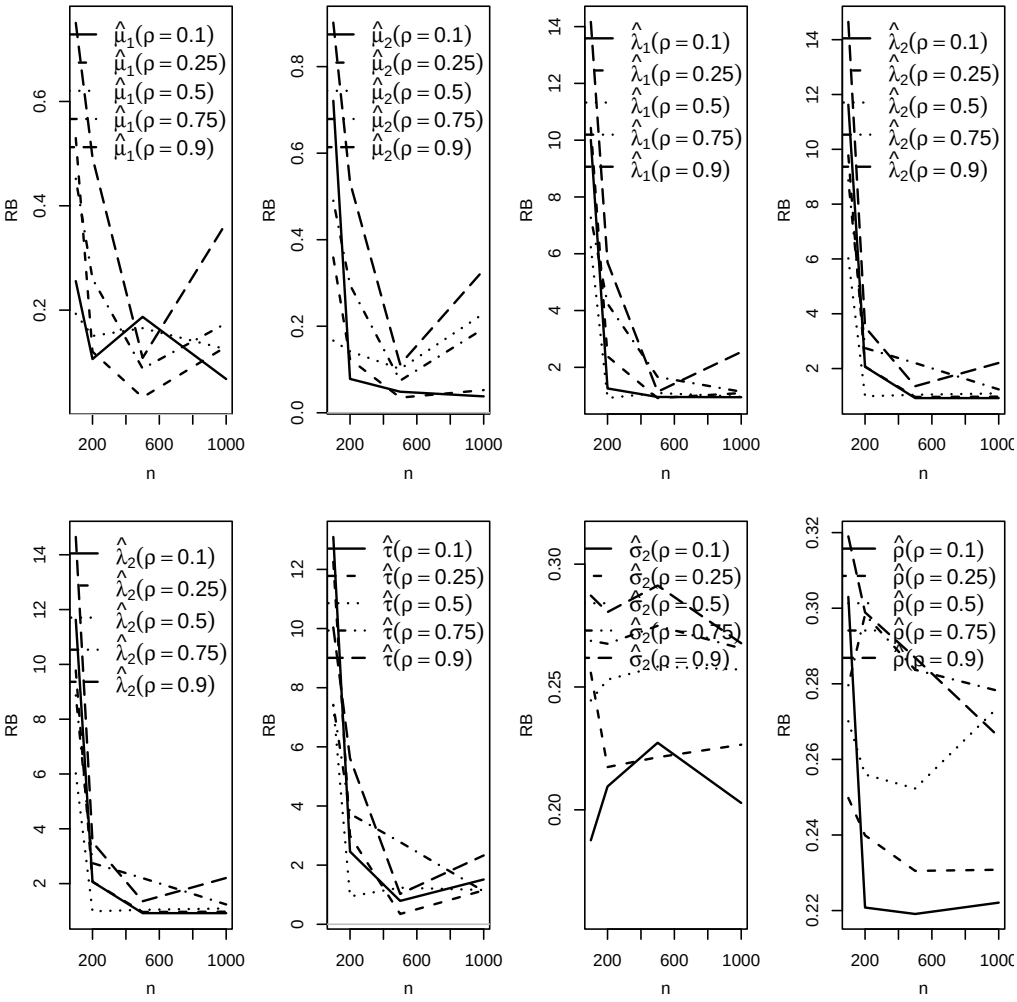


Figure 4.5: Relative bias for $G_i(x) = \cosh^{-1}(x + 1)$.

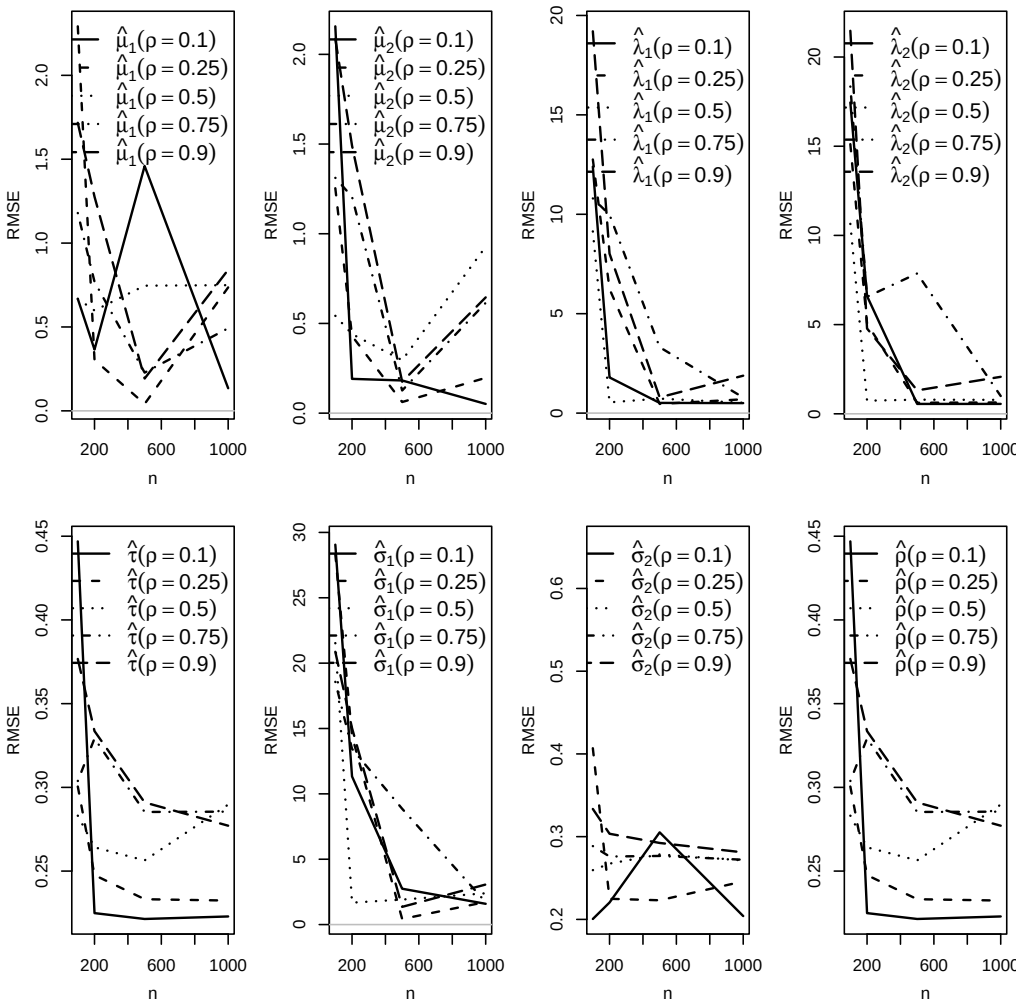


Figure 4.6: Root mean squared error for $G_i(x) = \cosh^{-1}(x + 1)$.

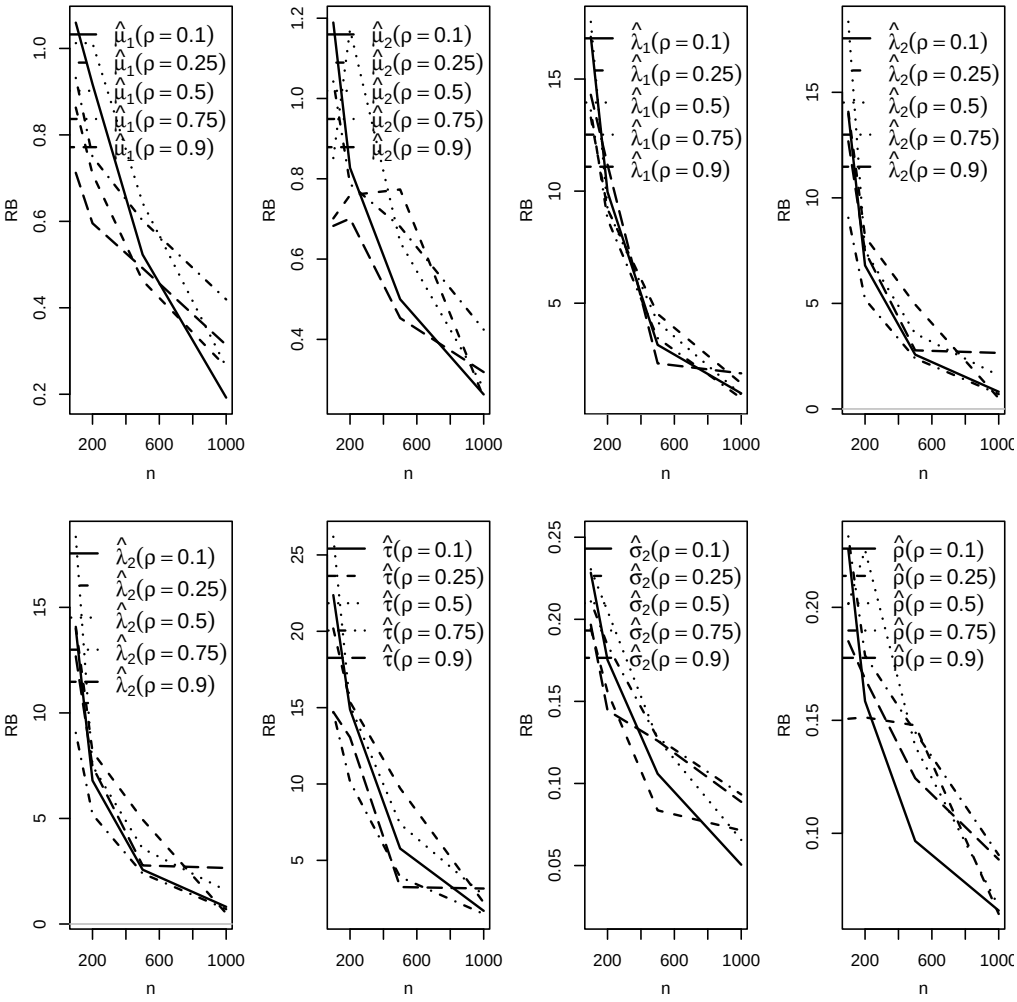


Figure 4.7: Relative bias for $G_i(x) = \log(\log(x + 1))$.

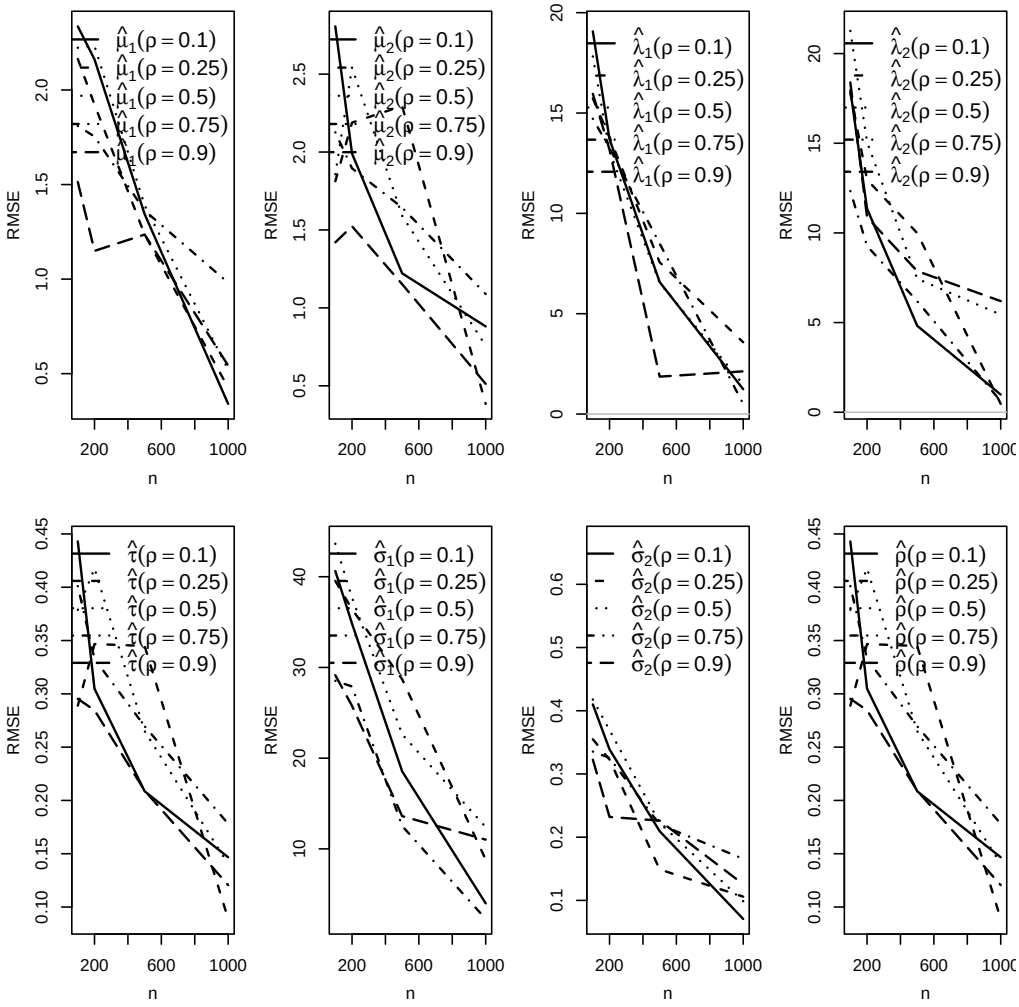


Figure 4.8: Root mean squared error for $G_i(x) = \log(\log(x + 1))$.

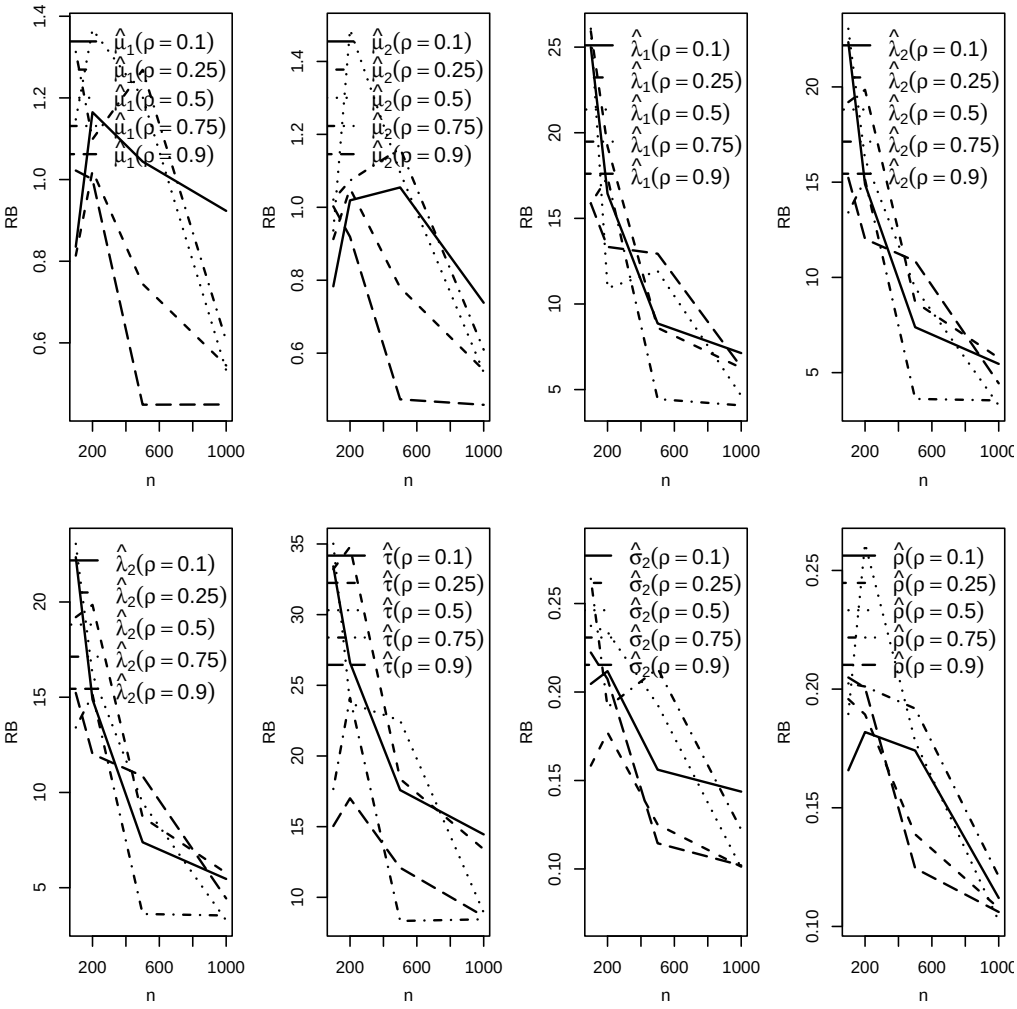


Figure 4.9: Bias for $G_i(x) = x - \frac{1}{x}$.

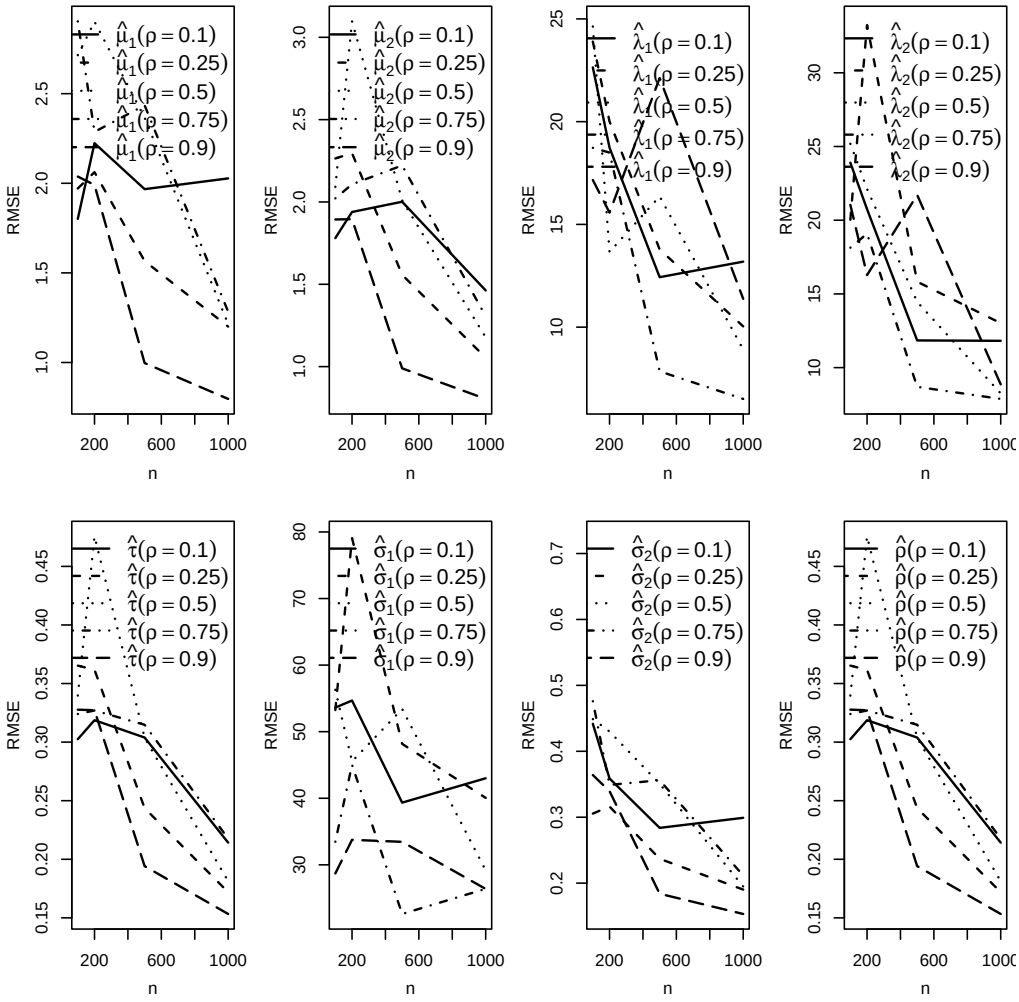


Figure 4.10: Root mean squared error for $G_i(x) = x - \frac{1}{x}$.

Chapter 5

Applications

In this section, we apply the proposed models to real data, utilizing a range of metrics to evaluate the quality of the model fits. To assess how well the data fits with the proposed distributions, we employ well-established statistical tests, including the Kolmogorov-Smirnov, Anderson-Darling, and Cramér-von Mises tests. The aim is to identify which distributions offer the best fit for the data

"*Without data, you're just another person with an opinion.*"

- W. Edwards Deming

5.1 Application to real data

5.1.1 Data description

In this chapter we will present a dataset extracted from (Cook and Weisberg, 2009), which contains information about athletes from the Australian Institute of Sport (AIS). The available data include the athlete's sex (Sex), sport (Sport), red blood cell count (RCC), white blood cell count (WCC), haematocrit (Hc), haemoglobin concentration (Hg), blood iron (Fe), body mass index (BMI), sum of skin folds (SSF), body fat percentage (Bfat), lean body mass (LBM), height (Ht) and weight (Wt).

Initially, height and weight information will be analysed without distinction of sex, but later these variables will be separated by sex. For ease of understanding, height will be referred to by the acronym Ht and weight by Wt. Where there is a distinction by sex, we will use the acronyms Ht-M and Wt-M for males and Ht-F and Wt-F for females. The analysis is bivariate. Descriptive statistics and data summaries are presented below.

Table 5.1: Summary statistics covering both sexes.

Variables	n	Minimum	Median	Mean	Maximum	SD	CV	CS	CK
Ht	202.00	148.90	179.70	180.10	209.40	9.73	5.40	-0.20	3.53
Wt	202.00	37.80	74.40	75.01	123.20	13.93	18.57	0.24	3.39

Table 5.2: Summary statistics for males.

Variables	n	Minimum	Median	Mean	Maximum	SD	CV	CS	CK
Ht-M	102.00	165.30	185.55	185.51	209.40	7.90	4.26	0.07	3.00
Wt-M	102.00	53.80	83.00	82.52	123.20	12.41	15.03	0.39	3.41

Table 5.3: Summary statistics for females.

Variables	n	Minimum	Median	Mean	Maximum	SD	CV	CS	CK
Ht-F	100.00	148.90	175.00	174.59	195.90	8.24	4.72	-0.56	4.20
Wt-F	100.00	37.80	68.05	67.34	96.30	10.92	16.21	-0.17	3.13

Then, the scatter and contour plots of each of the previously mentioned data can be analyzed.

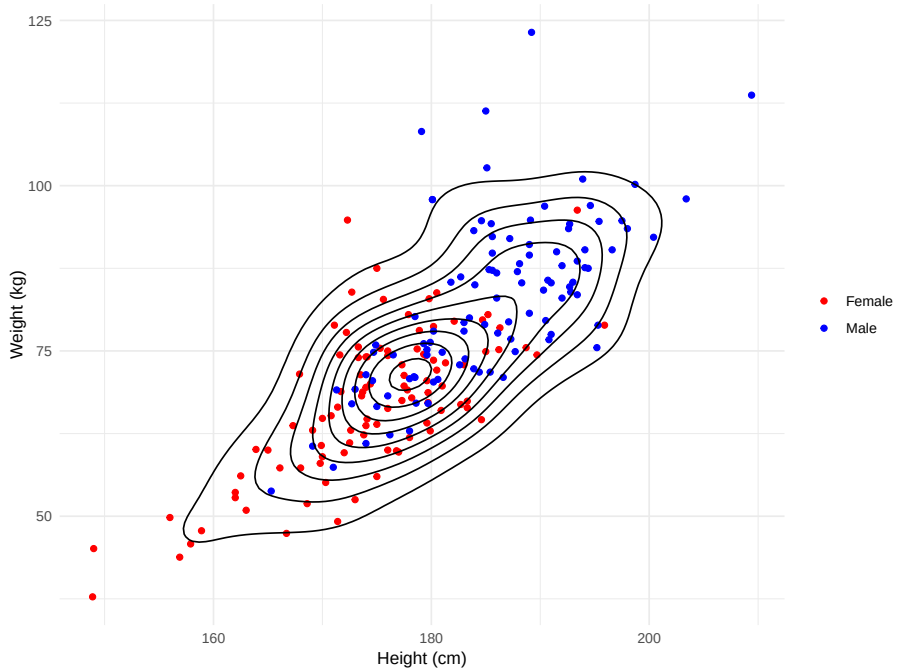


Figure 5.1: Scatterplot with contour lines illustrating the relationship between height and weight of athletes from the AIS database.

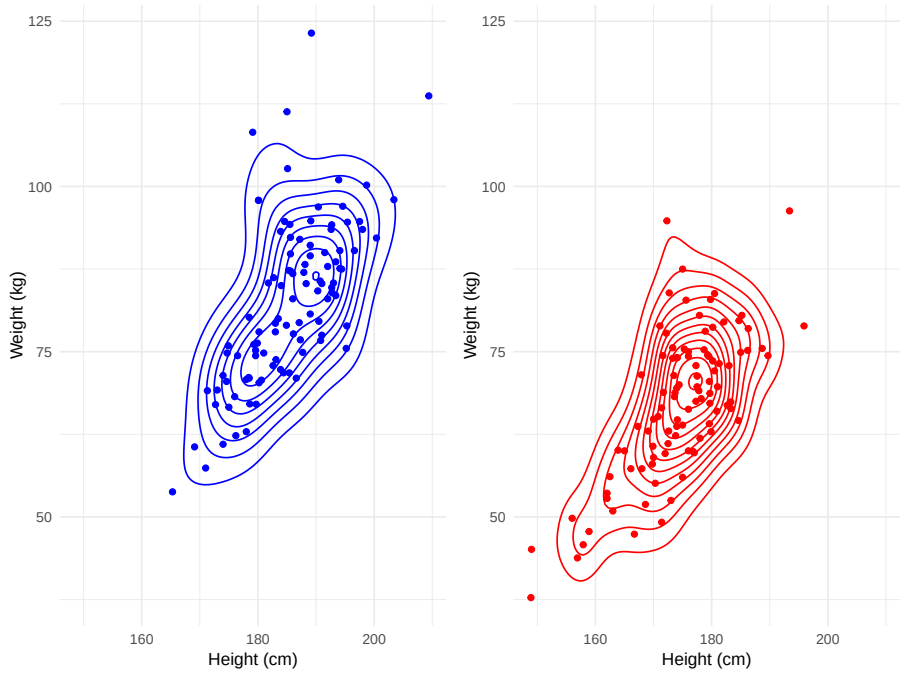


Figure 5.2: Scatterplot with contour lines showing the relationship between height and weight of athletes in the AIS database, with male athletes on the left and female athletes on the right.

Note that the graphs have distorted elliptical shapes, characteristics that are of particular interest for modelling the data in question.

Below are the boxplots for the three characteristics (both sexes, male and female), first for the weights and then for the heights.

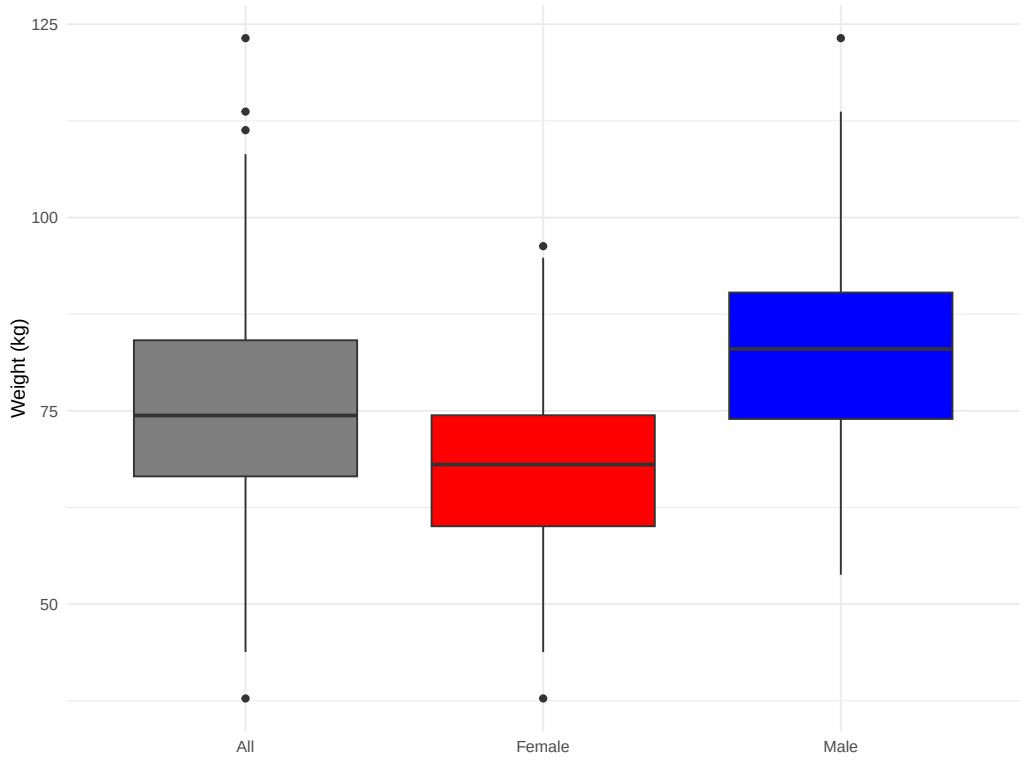


Figure 5.3: Boxplot of the athletes' weight, separated by sex (male and female) and with an overview of the total.

It can be concluded that there are significant differences in weight between the sexes, with men on average being heavier than women, as shown by the higher median weight in the male group. This difference can be seen in the fact that most of the male weights are concentrated between 70 kg and 90 kg, while the female weights are between 60 kg and 75 kg, reinforcing the trend towards greater body mass in men.

Furthermore, the presence of outliers in both groups, especially in the highest values, indicates the existence of individuals with weights significantly outside the expected standard, both

above and below the main range. These outliers contribute to the formation of heavy tails in the distributions, highlighting the variability and dispersion of the weights.

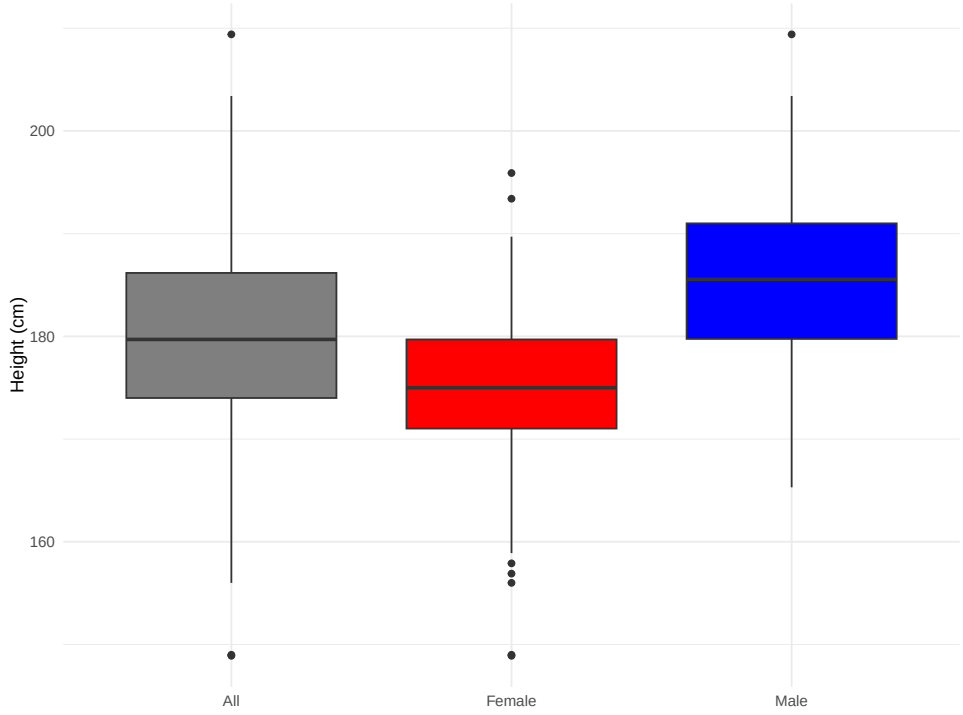


Figure 5.4: Boxplot of the athletes' height, separated by sex (male and female) and with an overview of the total.

It can be concluded that there are significant differences in height between the sexes, with men on average being taller than women. The median height is higher in the male group, close to 185 cm, while in the female group the median is around 170 cm, showing a clear difference between the distributions.

Men show a wider dispersion of heights, with most heights between 175 cm and 195 cm, while the female group has a narrower concentration, with heights predominantly between 165 cm and 175 cm.

Furthermore, the presence of outliers in the three groups, with values significantly above 200 cm or below 160 cm, suggests the existence of individuals with heights outside the expected standard, resulting in distributions with heavy tails.

5.1.2 Fit I

In this first data model using the EGSE distribution, the fit was performed using the log-q function. This function was developed based on the ideas presented by Constantino Tsallis in his seminal article (Tsallis, 1988), a fundamental reference for the generalization of logarithmic and exponential functions in the context of non-extensive statistics. In this work, Tsallis introduces the entropy S_q, a generalized function parameterized by q, which extends the applications of traditional functions.

The main reference for understanding this function in the context of this study was found in (Yamano, 2002). The extended logarithmic function, or skewed as it is commonly called, allows for more refined control of growth behavior. Consequently, this concept can also be applied to the exponential case with its inverse function. The log-q distribution is defined as

$$\ln_q x := \frac{x^{1-q} - 1}{1-q}. \tag{5.1.1}$$

Therefore, its inverse is given by:

$$\begin{aligned}
&\frac{x^{1-q} - 1}{1-q} = y, \\
&x^{1-q} - 1 = y(1-q), \\
&x^{1-q} - 1 = y - qx, \\
&x^{1-q} = y - qx + 1, \\
&x = [y - qx + 1]^{\frac{1}{1-q}},
\end{aligned}$$

where

$$e_q^x := \begin{cases} [1 + (1-q)x]^{\frac{1}{1-q}}, & \text{if } 1 + (1-q)x \geq 0, \\ 0, & \text{otherwise.} \end{cases}$$

Remembering that the value of the expression within the root must not be negative, which justifies the condition established previously.

The derivative of Equation 5.1.1 can be expressed as follows:

$$\frac{\partial \ln_q(x)}{\partial x} = \frac{(1-q)\,(x^{1-q-1})}{(1-q)} = x^{-q} = \frac{1}{x^q}$$

Note that the final expression presented above is similar to the derivative of $\ln(x)$, differing only by the presence of the exponent q in the denominator.

These three expressions are of fundamental importance for the distribution, as the function and its derivative are essential in calculating the density and the accumulated function, while the inverse function is crucial for generating data from the stochastic representation.

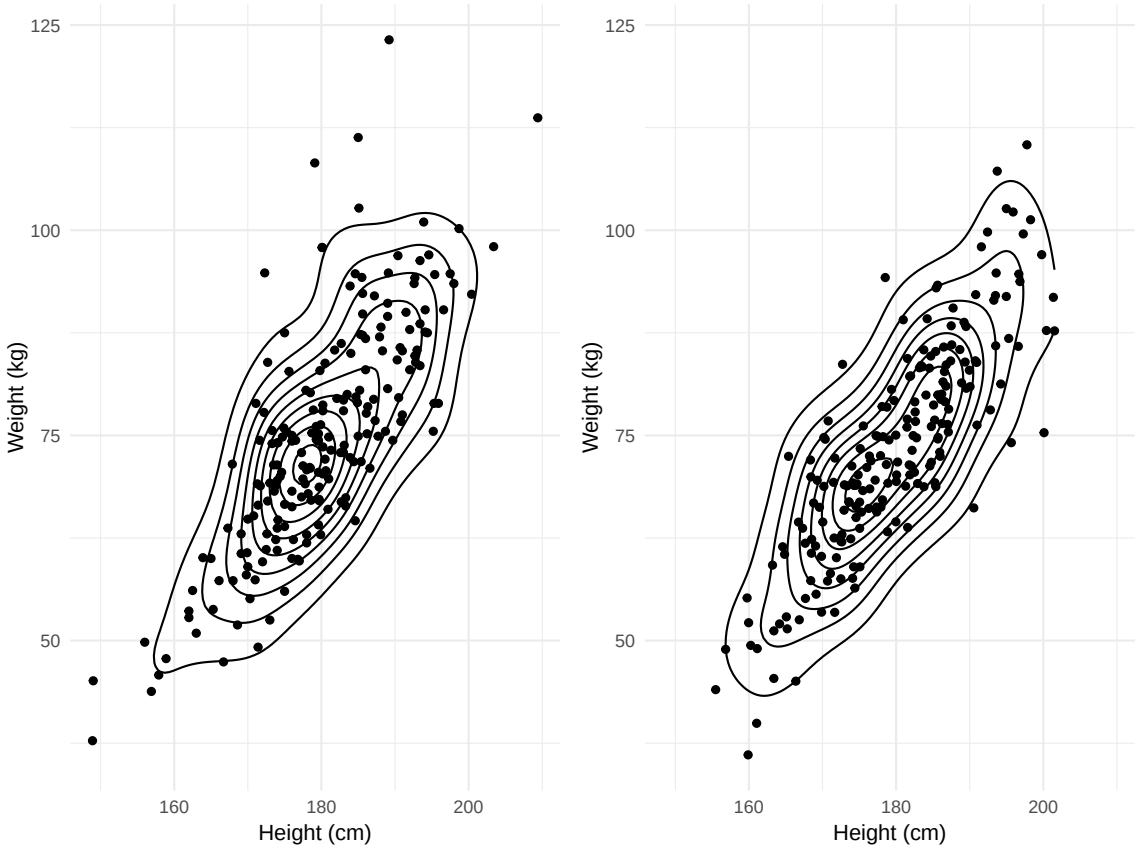
Using AIS data considering both sexes and applying the $\log -q$ function with variation in q values, the degrees of freedom were estimated in the context of the Extended G-skew distribution t-Student-. The following table presents the results of the statistical adherence tests, highlighting the metrics KS test, AD test and the log-MLE value for each variation of q:

Table 5.4: Resultados dos testes para o modelo EGSE$_2$-t.

q	df	KS.test	AD.test	log-MLE
0.75	34	0.070	0.132	1452.744
0.80	35	0.041	0.088	1452.398
0.85	22	0.054	0.108	1452.771
0.90	69	0.031	0.084	1455.953
0.95	14	0.070	0.125	1453.524
1.00	15	0.041	0.094	1454.505
1.05	13	0.041	0.087	1457.627
1.10	33	0.054	0.114	1459.853
1.15	11	0.115	0.174	1456.742
1.20	11	0.054	0.066	1457.416
1.25	8	0.796	0.412	1456.451

Taking into account the two tests carried out, the most appropriate model among those simulated in Extended G-skew-tStudent is the one with $q = 1.25$ and 8 degrees of freedom. Below are scatter plots illustrating the model fit: on the left, the actual values; on the right, the values generated with the parameters estimated by the stochastic representation.

Figure 5.5: On the left are the real data, and on the right are the simulated data with the estimated parameters.



Below, the estimated parameters are presented along with their respective estimation standard errors, indicated in parentheses.

Table 5.5: Estimation of parameters with standard errors indicated in parentheses.

$\hat{\mu}_1$	$\hat{\mu}_2$	$\hat{\lambda}_1$	$\hat{\lambda}_2$	$\hat{\sigma}_1$	$\hat{\sigma}_2$	$\hat{\rho}$	$\hat{\nu}$
2.688	2.918	-18.437	-4.347	0.080	0.017	0.869	8
(0.012)	(0.003)	(6.066)	(29.651)	(0.009)	(0.002)	(0.037)	-

To estimate the model parameters, we chose to set the parameter τ to zero, with the aim of simplifying the model and reducing computational costs. This choice is particularly strategic when the inclusion of τ does not significantly improve the fit or when its presence increases complexity, increasing the risk of overfitting and making the estimation process more laborious.

The parameter τ is valued for its ability to add flexibility to the model and can be adjusted according to the needs of the application. Setting it to zero is beneficial in situations where its extension function is unnecessary, resulting in a simpler and more efficient model. When the extra flexibility provided by τ is needed, it can be reintroduced and adjusted according to the context, maintaining the balance between accuracy and efficiency.

In summary, τ is an important component of the model that provides versatility in application. It can be estimated to capture additional details of the data, or fixed to maintain simplicity, adapted according to the specific objectives of the analysis.

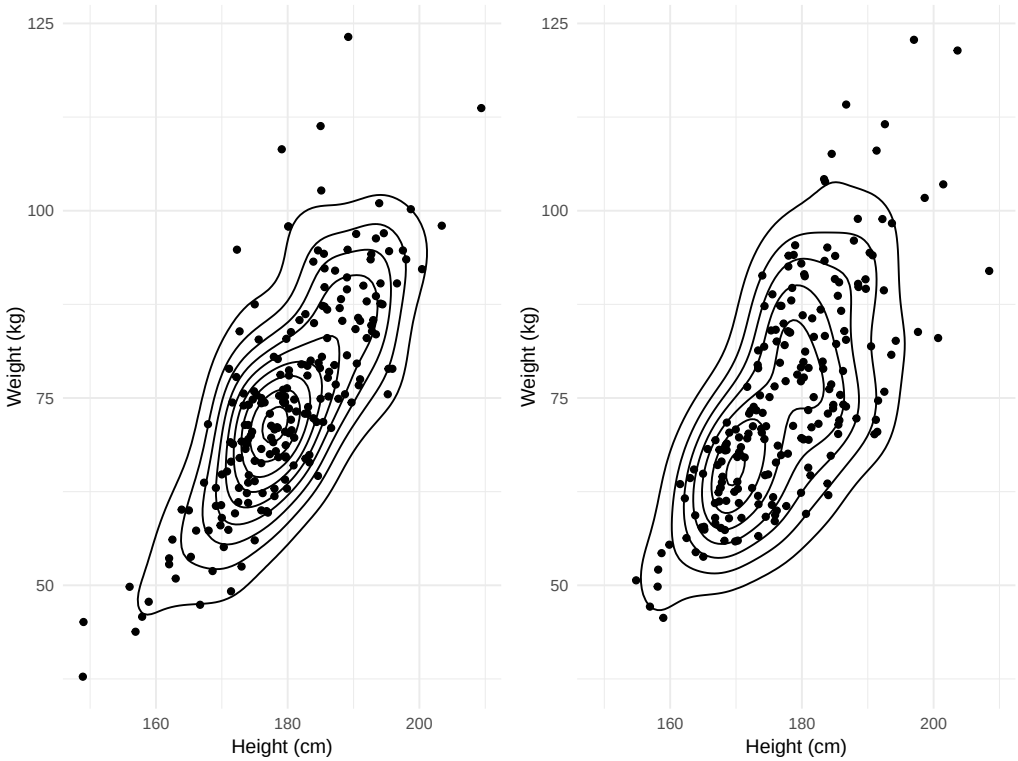
It is now applied in the case of the Extended G-skew-normal:

Table 5.6: Test results for the EGSE_2-Normal model.

q	KS.test	AD.test	log-MLE
0.75	0.031	0.061	1461.284
0.80	0.031	0.061	1460.881
0.85	0.023	0.057	1461.131
0.90	0.023	0.059	1460.769
0.95	0.007	0.039	1465.461
1.00	0.023	0.065	1460.380
1.05	0.017	0.061	1460.917
1.10	0.115	0.075	1460.819
1.15	0.017	0.062	1460.669
1.20	0.000	0.000	1616.649
1.25	0.000	0.000	1723.706

It is possible to infer that the best model was the one with $q = 1.10$, where both tests present a significance level greater than 0.05. Below, scatter plots illustrating the model fit are displayed: on the left, the observed values; on the right, the values generated with the parameters estimated by the stochastic representation.

Figure 5.6: On the left are the real data, and on the right are the simulated data with the estimated parameters.



The estimated parameters, together with their respective estimation standard errors (in parentheses), are presented below.

Table 5.7: Estimation of parameters with standard errors indicated in parentheses.

$\hat{\mu}_1$	$\hat{\mu}_2$	$\hat{\lambda}_1$	$\hat{\lambda}_2$	$\hat{\sigma}_1$	$\hat{\sigma}_2$	$\hat{\rho}$
4.017	3.486	160.507	-29.452	0.046	0.123	0.566
(0.003)	(0.023)	(30.379)	(5.307)	(0.003)	(0.006)	(0.123)

Between the two distributions evaluated, the Extended G-skew-t Student model stood out as the best fit, possibly due to the inclusion of an additional parameter, the degree of freedom, which gives the model greater flexibility. However, it is important to stress that the results of the goodness of fit tests (Kolmogorov-Smirnov and Anderson-Darling) are not definitive, but only serve as a guideline for the selection of the most appropriate model. In several simulations,

in particular for Extended G-skew-tStudent, a high degree of similarity between the simulated and real values was observed graphically, providing visual support for the choice.

Despite the importance of goodness of fit tests, other validation metrics such as information criteria should also be considered for a more robust evaluation, such as the Akaike Information Criterion (AIC) (Akaike, 1998) and the Bayesian Information Criterion (BIC) (Schwarz, 1978). These metrics allow the goodness of fit between models to be compared, taking into account both fit and complexity, providing a more comprehensive basis for selecting the optimal model.

No specific applications were performed for the Extended G-skew-Cauchy model, as we believe that this case is assumed when the degree of freedom is equal to 1. However, as this value was not found in any of the estimates obtained, it was not possible to identify the presence of the Extended G-skew-Cauchy model in the results.

It can be considered that the best fit between the two families presented was that of the t-student distribution, according to the Kolmogorov-Smirnov (KS) test, which showed a significantly superior performance when comparing the two best models in each case.

To consolidate the model fit, in this and subsequent cases we will use the RQ residuals (Randomized Quantile Residuals), defined by (Dunn and Smyth, 1996) as $R_i^{RQ} = \Phi^{-1}[F(y_i|\theta)]$. In this formula, $F(y_i|\theta)$ is the cumulative distribution function (CDF) of the model fitted to each observation y_i, and Φ^{-1} is the function quantile of the standard normal distribution $N(0,1)$. When F the model is correctly specified, the RQ residuals follow a standard normal distribution. The QQ plot of the RQ residuals is used to assess the compliance of the residuals with normality. In this plot, the quantiles of the residuals are compared with the theoretical quantiles of the standard normal distribution. The proximity of the points to the reference line indicates that the residuals follow a normal distribution, indicating a good fit of the model. Substantial deviations from the line indicate possible problems, such as poor model specification or the presence of outliers. To improve the analysis, a 95% confidence envelope is included, which helps to visualise the expected variability of the residuals and facilitates the identification of significant deviations.

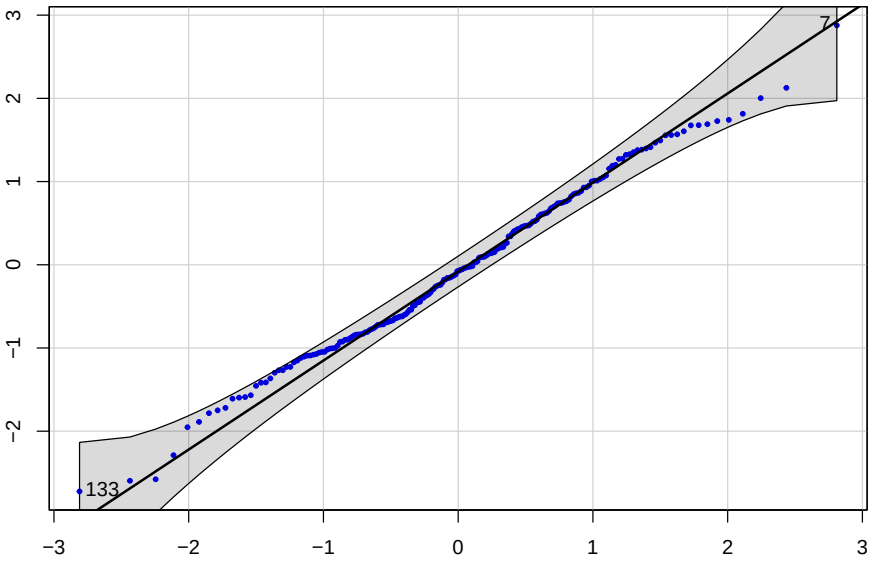


Figure 5.7: Q-Q plot of RQ residuals - t-Student Family.

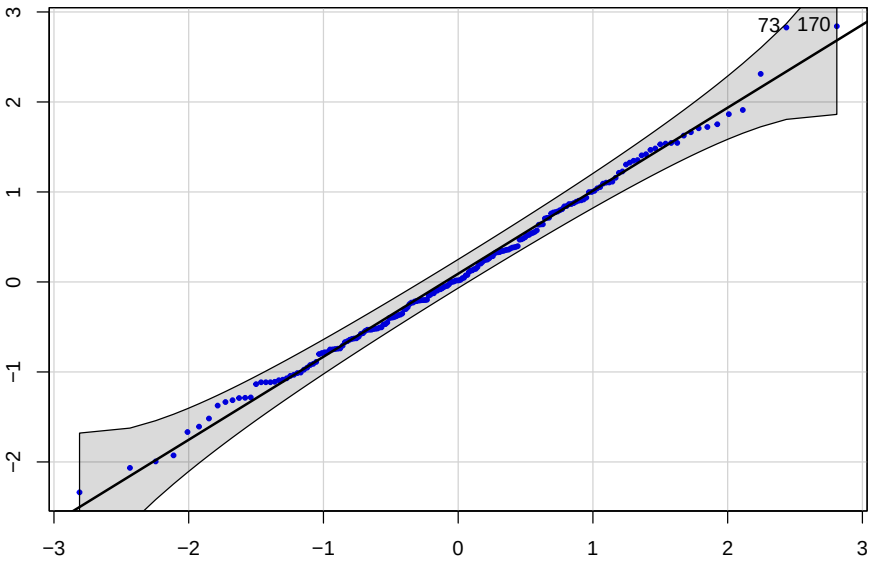


Figure 5.8: Q-Q plot of the RQ residuals - Normal Family.

5.1.3 Fit II

To continue the analysis of the male athletes' data, we focus on modeling based on the EGSE$_2$-Student-t distribution. This approach uses a specific function, described by the equation $\frac{1}{\alpha}\left(\sqrt{\frac{x}{\beta}} - \sqrt{\frac{\beta}{x}}\right)$, which comes from the Birnbaum-Saunders distribution, as discussed in (Leiva, 2015). The choice of this function is justified by its ability to capture complex behaviors in the data, being particularly useful for modeling asymmetric and heavy-tailed distributions, characteristics frequently observed in sports performance datasets.

To demonstrate the flexibility of the proposed distribution, we used widely recognized constants in the field of mathematics, such as the value of pi ($\pi = 3.1415$) and Euler's constant ($e = 2.71828$). These constants were chosen not only for their mathematical relevance, but also for their applicability in different modeling contexts, allowing a robust and intuitive parameterization of the distribution.

The following table presents the results of the tests for the t-Student variant of the model, which demonstrate the quality of the fit to the athlete data.

Table 5.8: Test results for the EGSE$_2$-t model.

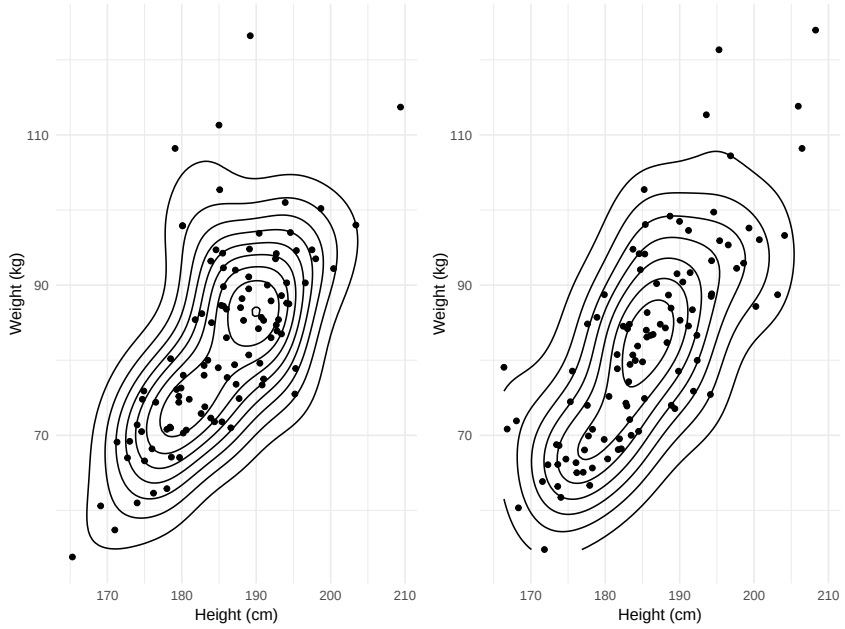
α	β	**df**	**KS.test**	**AD.test**	**log-MLE**
π	e	14	0.82	0.54	714.536

The results obtained, as presented in Table 5.19, indicate a robust fit of the t-Student model to the analyzed data. The value of α was parameterized as π, while β was fixed as e, reflecting the strategic use of these constants. The degrees of freedom parameter was estimated at 14, which suggests a distribution with slightly heavier tails than the standard normal.

The goodness-of-fit tests, including the Kolmogorov-Smirnov test and the Anderson-Darling test, presented values of 0.82 and 0.54, respectively. These values are significantly above the 5% significance level, demonstrating that the model fits the data well and that there is no evidence of significant deviation between the theoretical distribution and that observed in the athletes' data.

These results highlight the ability of the EGSE$_2$-Student-t distribution to capture the nuances present in the data and to provide an appropriate statistical representation of the athletes' performance.

Figure 5.9: Comparison between real (left) and simulated (right) data with estimated parameters.



The best value estimate for the parameter ν was 14. The graph confirms that the simulated data fits the real data well, demonstrating the quality of the model.

Table 5.9: Estimation of parameters with standard errors indicated in parentheses.

$\hat{\mu}_1$	$\hat{\mu}_2$	$\hat{\lambda}_1$	$\hat{\lambda}_2$	$\hat{\sigma}_1$	$\hat{\sigma}_2$	$\hat{\rho}$	$\hat{\nu}$
1.592	2.595	24.616	-43.254	0.155	0.053	0.497	14
(0.024)	(0.011)	(6.470)	(12.129)	(0.018)	(0.004)	(0.121)	-

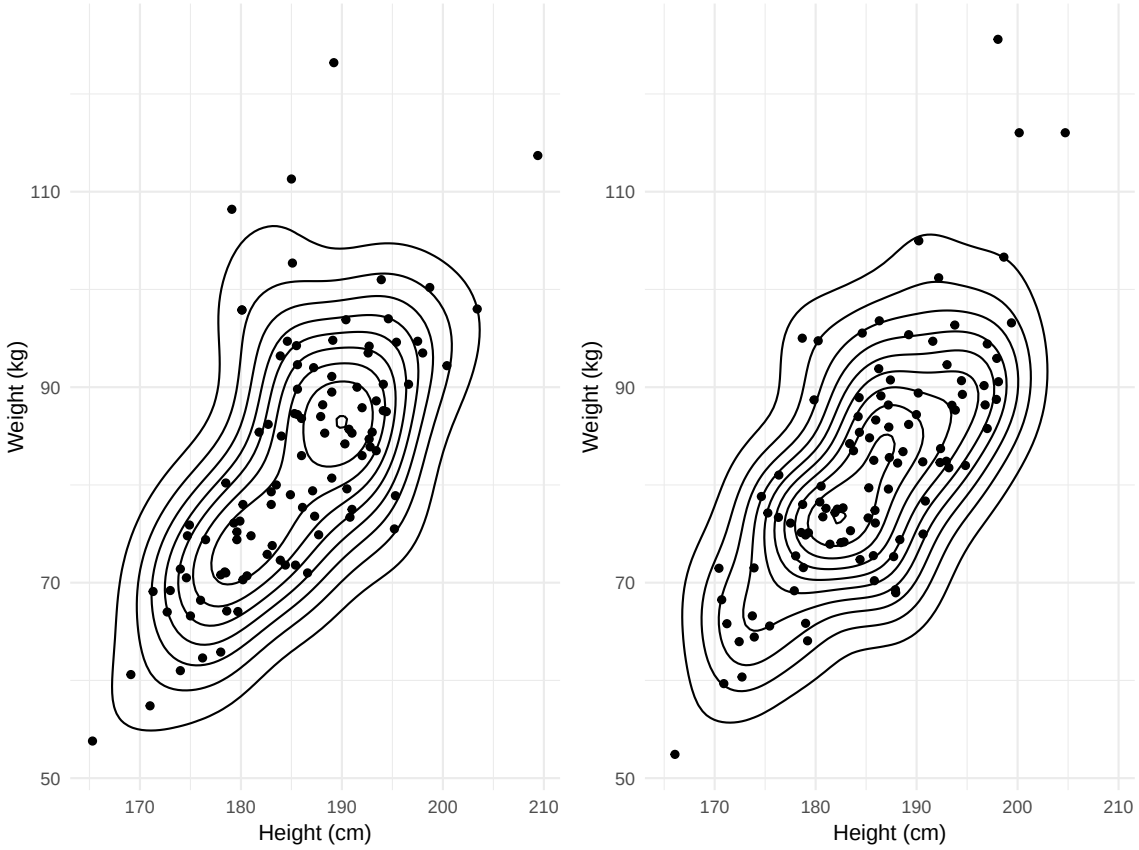
We then analyze the fit using the EGSE$_2$-Normal distribution.

Table 5.10: Test results for the EGSE$_2$-Normal model.

α	β	**KS.test**	**AD.test**	**log-MLE**
π	e	0.48	0.20	713.505

The results indicate that the fit to the Normal distribution was also satisfactory, with both tests above the significance level.

Figure 5.10: Comparison between real (left) and simulated (right) data with estimated parameters.



The graph confirms that the simulated data fit the real data well. The table below shows the estimated parameters.

Table 5.11: Estimation of parameters with standard errors indicated in parentheses.

$\hat{\mu}_1$	$\hat{\mu}_2$	$\hat{\lambda}_1$	$\hat{\lambda}_2$	$\hat{\sigma}_1$	$\hat{\sigma}_2$	$\hat{\rho}$
1.583	2.593	24.122	-41.144	0.172	0.056	0.507
(0.023)	(0.011)	(6.196)	(10.931)	(0.019)	(0.003)	(0.117)

The Q-Q plots of the fits reveal that the residuals follow the proposed distributions well.

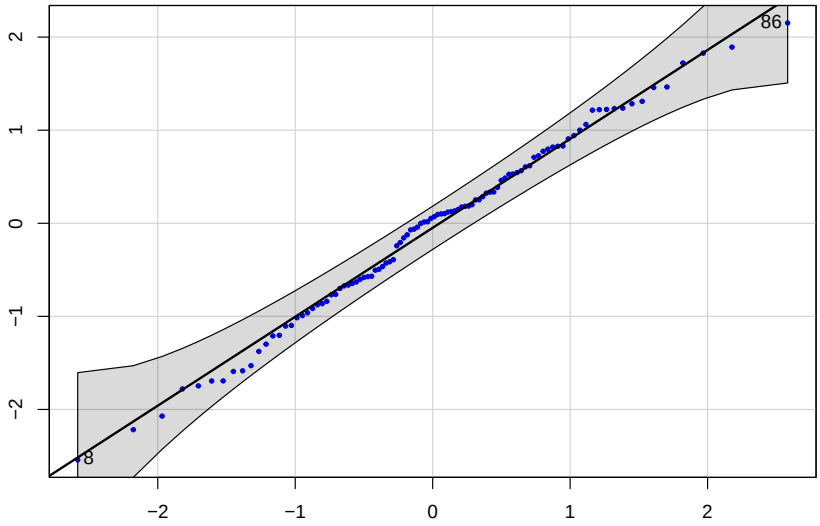


Figure 5.11: Q-Q plot of RQ residuals - t-Student Family.

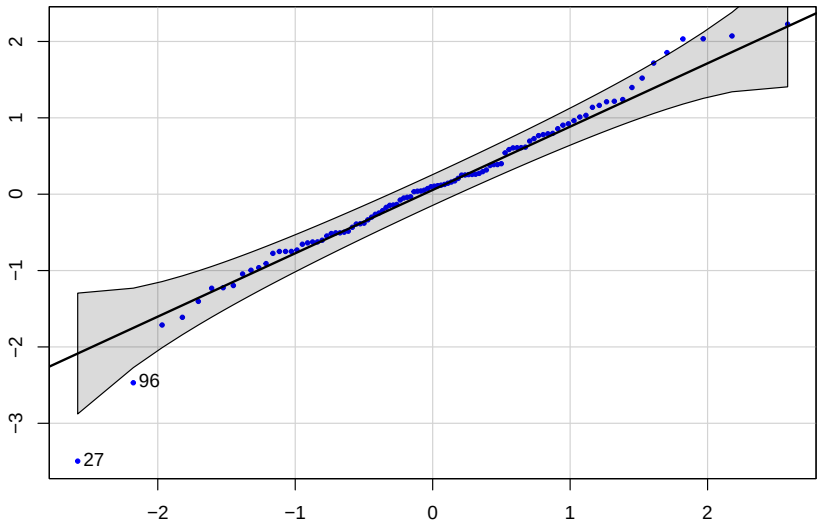


Figure 5.12: Q-Q plot of RQ residuals - Normal Family.

Comparing the models, the Extended G-skew-Student-t showed the best overall perfor-

mance, considering both the tests and the graphs. However, the G-skew-Student-Normal model also showed a good fit, being a viable alternative for modeling.

5.1.4 Fit III

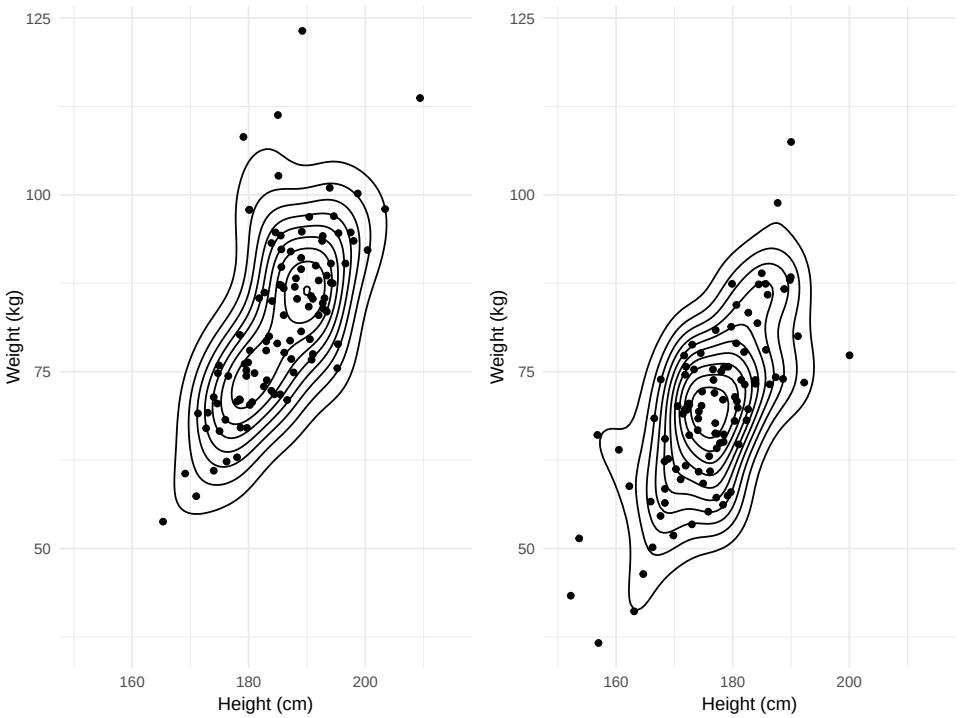
Now, we consider the fit to the female data using the function $\cosh^{-1}(x+1)$ within the G-skew-Student-t distribution.

Table 5.12: Test results for the EGSE_2-t model.

df	KS.test	AD.test	log-MLE
6	0.48	0.21	692.415

The test results were similar to those obtained for the male data, all above the 5% significance level, indicating that the fit is adequate. The following figure compares the real and simulated data.

Figure 5.13: Comparison between real (left) and simulated (right) data with estimated parameters.



The standard errors of the estimated parameters remained high, reflecting the difficulty of identifiability mentioned above.

Table 5.13: Estimation of parameters with standard errors indicated in parentheses.

$\hat{\mu}_1$	$\hat{\mu}_2$	$\hat{\lambda}_1$	$\hat{\lambda}_2$	$\hat{\sigma}_1$	$\hat{\sigma}_2$	$\hat{\rho}$	$\hat{\nu}$
5.062	5.902	-9.287	-20.973	0.198	0.054	0.848	6
(0.025)	(0.006)	(4.226)	(10.602)	(0.025)	(0.006)	(0.041)	-

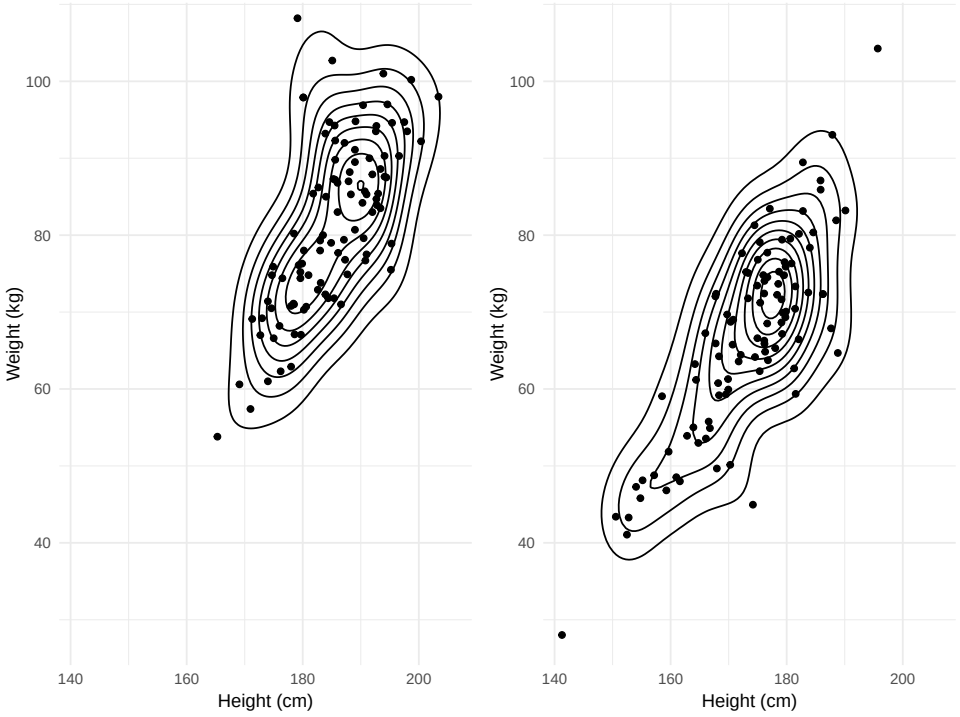
For the G-skew-Student-Normal case, the results are presented below.

Table 5.14: Test results for the EGSE_2-Normal model.

KS.test	**AD.test**	**log-MLE**
0.28	0.20	690.239

Although the KS test has decreased, the fit is still considered good as shown graphically.

Figure 5.14: Comparison between real (left) and simulated (right) data with estimated parameters.



Below, the estimated parameters are presented.

Table 5.15: Estimation of parameters with standard errors indicated in parentheses.

$\hat{\mu}_1$	$\hat{\mu}_2$	$\hat{\lambda}_1$	$\hat{\lambda}_2$	$\hat{\sigma}_1$	$\hat{\sigma}_2$	$\hat{\rho}$
5.082	5.907	-7.980	-18.285	0.244	0.067	0.877
(0.022)	(0.006)	(2.892)	(7.514)	(0.024)	(0.006)	(0.027)

The Q-Q plots confirm the good fit of the residuals to the proposed models.

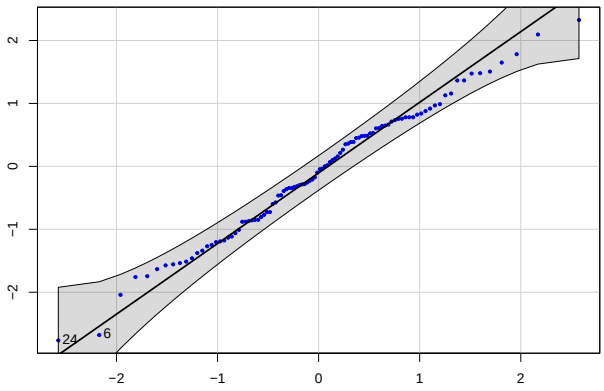


Figure 5.15: Q-Q plot of RQ residuals - t-Student Family.

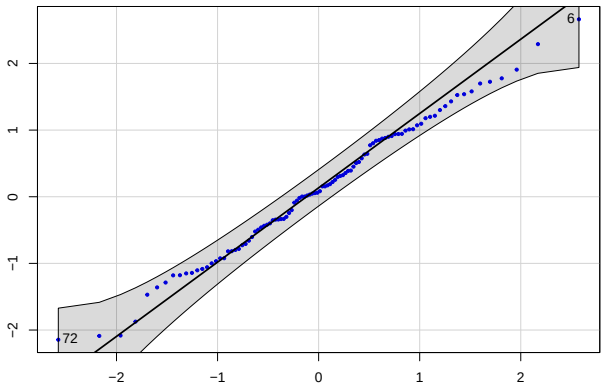


Figure 5.16: Q-Q plot of RQ residuals - Normal Family.

Based on the analysis of the fits and the test results, we conclude that both the G-skew-Student-t and the G-skew-Student-Normal models offer good performance in modeling the data, with the former showing slightly superior results.

5.1.5 Fit IV

In this excerpt, we will use the data presented by (Becker, Chambers, and Wilks, 1988), which cover information from the 50 states of the United States. Among the variables available in the database, we will analyze life expectancy in years (1969–1971) and the rate of homicides and negligent deaths per 100,000 inhabitants (1976) for each state, totaling 50 bivariate samples.

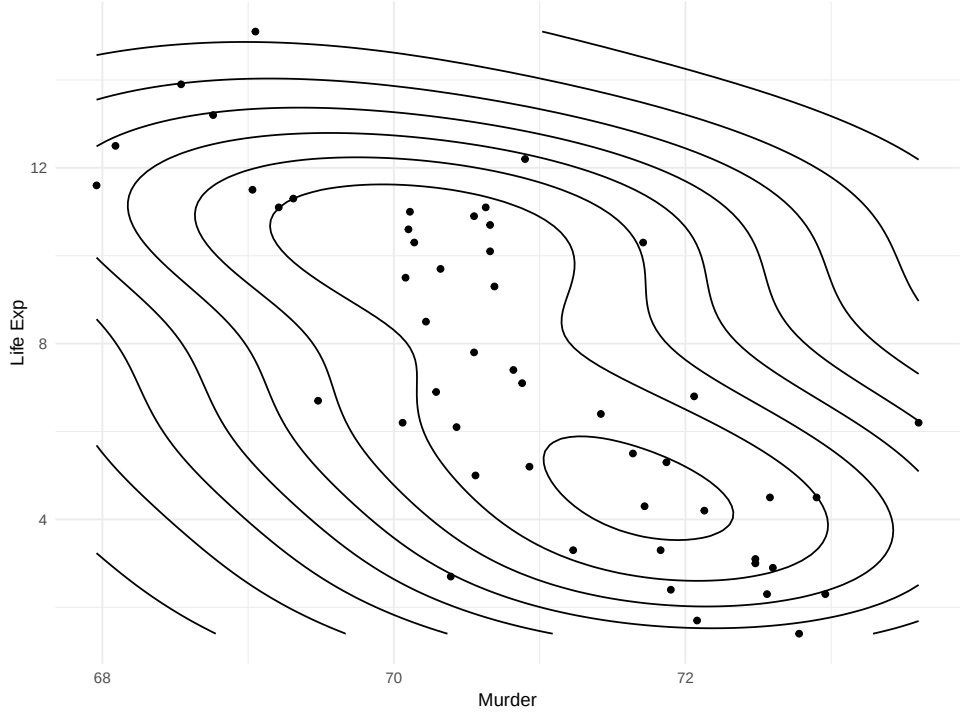


Figure 5.17: Scatterplot with contour lines illustrating the relationship between life expectancy and the rate of homicide or negligent deaths.

Below, we present a descriptive table that provides an overview of the main information of the dataset in question.

Table 5.16: Summary statistics.

Variables	n	Minimum	Median	Mean	Maximum	SD	CV	CS	CK
Life Exp	50.00	67.96	70.68	70.88	73.60	1.34	1.89	-0.16	-0.57
Murder	50.00	1.40	6.85	7.38	15.10	3.69	50.03	0.13	-1.14

The fits will now be evaluated using the Cramér–von Mises criterion (also known as the Cramér–von Mises test), originally developed by Harald Cramér and Richard Edler von Mises, later adapted for the comparison between two samples, as described in (Anderson, 1962). These fits will be applied to the distributions of the functions $\log(\log(x+1))$ and $x-\frac{1}{x}$, both monotonically increasing, invertible and differentiable. Below, we present a table containing the fitted data in the context of the Extended G-skew-normal model, showing the p-value of the test and the log-MLE.

Table 5.17: Test results for the EGSE_2-Normal model

$G(x)$	**Cramér–von Mises criterion**	**log-MLE**
$\log(\log(x+1))$	0.62	202.999
$x-\frac{1}{x}$	0.83	199.942

It is confirmed that the adjustments in both distributions are good, since the p-value is above the defined significance level, which is 5% or 0.05, the same used in the other tests.

Table 5.18: Estimation of parameters with standard errors indicated in parentheses.

$G(x)$	$\hat{\mu}_1$	$\hat{\mu}_2$	$\hat{\lambda}_1$	$\hat{\lambda}_2$	$\hat{\sigma}_1$	$\hat{\sigma}_2$	$\hat{\rho}$
$\log(\log(x+1))$	1.453	0.628	8.939	0.658	0.004	0.281	-0.719
-	(0.004)	(0.117)	(507.986)	(2.439)	(0.0002)	(0.0313)	(0.018)
$x-\frac{1}{x}$	68.649	14.185	59.616	-24.808	2.359	7.330	-0.937
-	(0.148)	(0.359)	(81.842)	(34.835)	(0.243)	(0.745)	(0.016)

Furthermore, it is possible to observe that the standard errors in the asymmetry parameters remain high, while in the other parameters the errors are well behaved. The value of ρ shows

a strong and negative correlation. Below are the Q-Q plot graphs of the residuals, which were also used in the previous adjustments.

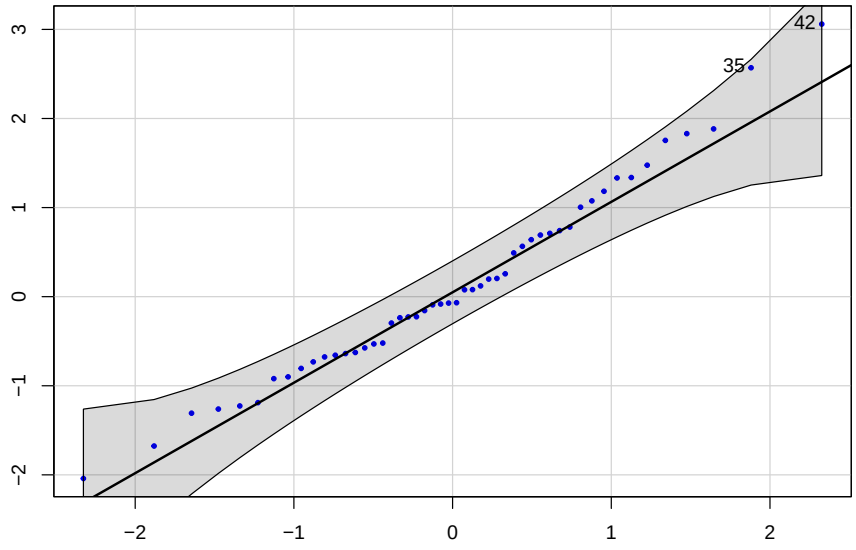


Figure 5.18: Q-Q plot of RQ residuals - Normal Family - $\log(\log(x+1))$.

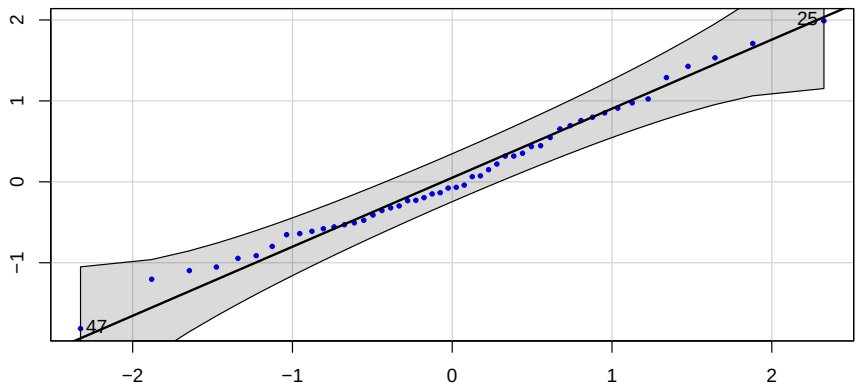


Figure 5.19: Q-Q plot of RQ residuals - Normal Family - $x - \frac{1}{x}$.

Similarly, the Extended G-skew-Student-t distribution was used with the same functions and in the same data set.

Table 5.19: Test results for the EGSE$_2$-t model.

$G(x)$	ν	**Cramér–von Mises criterion**	**log-MLE**
$\log(\log(x+1))$	15	0.69	201.434
$x-\frac{1}{x}$	24	0.54	205.857

The results continue to show good p-values, i.e., above 0.05, indicating that there is no evidence to reject the hypothesis that the data are not significantly different from each other. This suggests that the models used capture the characteristics of the data well, maintaining the consistency of the adjustments. Below, we present the parameter estimates for the applied transformations, with the respective standard errors in parentheses.

Table 5.20: Estimation of parameters with standard errors indicated in parentheses.

$G(x)$	$\hat{\mu}_1$	$\hat{\mu}_2$	$\hat{\lambda}_1$	$\hat{\lambda}_2$	$\hat{\sigma}_1$	$\hat{\sigma}_2$	$\hat{\rho}$	ν
$\log(\log(x+1))$	1.448	1.003	20.382	-32.538	-0.005	0.431	0.882	15
-	(0.0002)	(0.033)	(314.478)	(11.876)	(0.0004)	(0.040)	(0.028)	-
$x-\frac{1}{x}$	68.649	14.185	59.616	-24.808	2.359	7.333	-0.937	24
-	(0.148)	(0.359)	(81.842)	(34.835)	(0.243)	(0.745)	(0.016)	-

The estimated parameters are quite similar to those previously obtained with the Extended G-skew-normal, indicating consistency in the results between the different modeling approaches. This similarity suggests that the methods employed are robust, even when considering different functions. The Cramér–von Mises test presented slightly better results with the G-skew-Student-t, suggesting a marginally better fit for this distribution. This behavior can be attributed to the small number of samples, a situation in which the G-skew-Student-t is theoretically more effective.

In addition, the analysis of the residuals reinforces the quality of the fit, with the residual

values remaining within the 95% confidence envelope, demonstrating that the model captures the variability of the data well. This good fit of the residuals indicates that the model not only adapts well to the main data, but also that the observed deviations are consistent with what was expected, evidencing an adequate conformity to theoretical expectations.

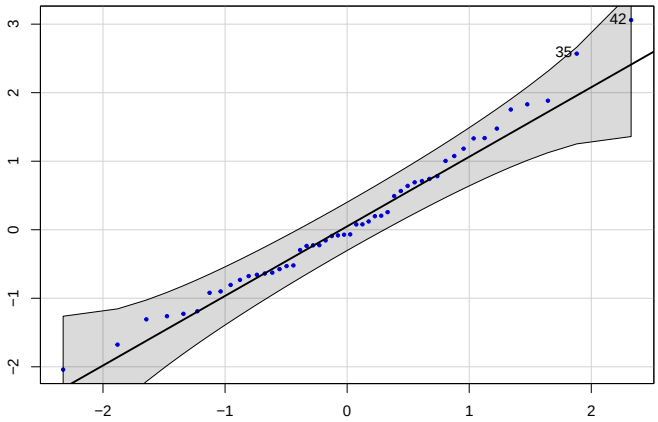


Figure 5.20: Q-Q plot of RQ residuals - t-Student Family - $\log(\log(x+1))$.

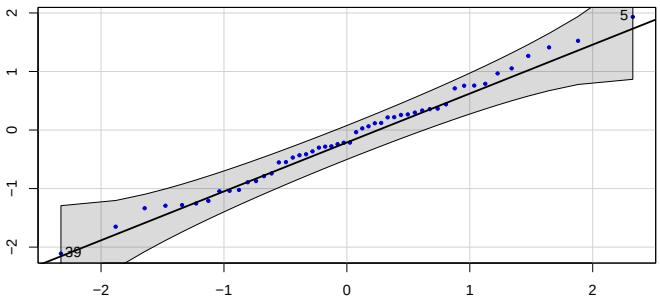


Figure 5.21: Q-Q plot of RQ residuals - t-Student Family - $x - \frac{1}{x}$.

The use of the Cramér–von Mises (CvM) test and its satisfactory results reinforce the effectiveness of this approach for modeling asymmetric or distorted data. The choice of this test in the final example aimed to demonstrate this capability, highlighting that the CvM presents a ro-

bust performance across the data distribution, being less sensitive to small samples and effective in complex distributions. In contrast, the Kolmogorov–Smirnov (KS) test is more intuitive and easy to interpret, but presents unequal sensitivity across the distribution. The Anderson–Darling (AD) test is particularly efficient in detecting differences in the tails, but may be less consistent across the rest of the distribution. In addition, the low p-values observed in the KS and AD tests are often due to the presence of ties, compromising the results, which is not the case with the CvM, demonstrating its superiority in these contexts.

Chapter 6

Conclusion

Based on the results of this work, it was possible to verify that the distributions studied demonstrate great versatility in different contexts. As evidenced in the application, these distributions, despite their complexity, are rich in details that allow you to configure different aspects of data with their particularities. They have the characteristic of being modeled based on differentiable, invertible and monotonically increasing functions, inserted in a class of asymmetric distributions, which has been widely explored by several authors. This approach enabled the development of distributions with the capacity to model different sets of positive data in different ways, expanding the scope of study of normal distributions.

This study marks the beginning of a new class of distributions, paving the way for the construction of new metrics and probabilistic and statistical concepts. The importance of asymmetric data in our world is highlighted, since most real data presents some degree of asymmetry, which is rarely considered in the multivariate context. The proposed approach differentiates itself by dealing with such asymmetries in a practical and efficient way.

During the simulations carried out in the study phase, a tendency towards reduction of the two metrics used — relative bias and mean squared error — was observed in most cases of parameter estimation. This trend suggests a very effective estimation, although the model presents a characteristic of non-identifiability. This aspect is important in certain applications, which

practically suggests that different sets of parameters can result in the same maximum likelihood function, significantly impacting parametric estimation processes.

Additionally, it was identified that, in some mathematical contexts, such as marginal moments, it is not possible to guarantee the finiteness of these moments due to the domain of the functions involved in density. Another relevant observation is that the product present in the density function does not directly affect the parameter estimation process, but can influence the scale of the density function. An explicit characterization for the Mahalanobis distance was presented, and the emphasis given to heavy-tailed distributions at the beginning of this study makes it possible to classify them in a new category, in addition to the traditional asymmetric distributions.

Finally, it is essential to recognize the contribution of renowned authors who were pioneers in this field, such as Adelchi Azzalini, Marc Genton, Arnold Balakrishna, Yulia Machekha, Samuel Kotz, Kai-Tai Fang, Marlene Branco, Eustáquio A. Valle, Sangyeol Lee, Dimitris Karlis and Guangyuan Yu. The work developed dialogues with the ideas of these scholars, advancing the understanding and application of new distributions for statistical modeling.

Appendix A

Cumulative Distribution Function

To find the best fit of the proposed distributions, it was necessary to calculate the cumulative distribution function (CDF) of the multivariate distributions. In general, in the multivariate case, the CDF is given by:

$$F_{\mathbf{Y}}(\mathbf{y}) = \int_0^{y_1} \int_0^{y_2} \dots \int_0^{y_n} f_{\mathbf{Y}}(u_1, u_2, \dots, u_n)\, du_n \dots du_2 du_1.$$

Note that the integrals start at 0, since the support of the distributions considered is the range from zero to infinity. For the distributions developed in this work, we have:

- Extended G-skew-Student-t:

$$F_{(\boldsymbol{\mu},\boldsymbol{\Sigma},\boldsymbol{\lambda},\tau,g^{(n)},\nu)}(\mathbf{y}) = \int_0^{y_1} \int_0^{y_2} \dots \int_0^{y_n} t_n(y_G;\mu,\Sigma,\nu) \frac{F_{\nu+1}\left(\lambda^\top(y_G-\mu) + \sqrt{\frac{\nu+1}{\nu+(y_G-\mu)^\top\Sigma^{-1}(y_G-\mu)}}\right)}{F_\nu\left(\sqrt{\frac{\nu+(y_G-\mu)^\top\Sigma^{-1}(y_G-\mu)}{1+\lambda^\top\Sigma\lambda}}\right)} \times \prod_{i=1}^{n} G_i'(y_i)\, du_n \dots du_2 du_1.$$

- Extended G-skew-Cauchy:

$$F_{(\boldsymbol{\mu},\boldsymbol{\Sigma},\boldsymbol{\lambda},\tau,g^{(n)})}(\mathbf{y}) = \int_0^{y_1}\int_0^{y_2}\cdots\int_0^{y_n} c_n(y_G;\mu,\Sigma)\frac{F_2\left(\left[\lambda^\top(y_G-\mu)+\tau\right]\sqrt{\frac{2}{1+q(y_G)}}\right)}{F_1\left(\frac{\tau}{\sqrt{1+\lambda^\top\Sigma\lambda}}\right)} \times \prod_{i=1}^{n} G_i'(y_i)\,du_n\ldots du_2du_1.$$

- Extended G-skew-normal:

$$F_{(\boldsymbol{\mu},\boldsymbol{\Sigma},\boldsymbol{\lambda},\tau,g^{(n)})}(\mathbf{y}) = \int_0^{y_1}\int_0^{y_2}\cdots\int_0^{y_n} \phi_n(y_G;\mu,\Sigma)\frac{\Phi\left(\frac{\lambda^\top(y_G-\mu)+\tau}{\sqrt{1+\lambda^\top\Sigma\lambda}}\right)}{\Phi\left(\frac{\tau}{\sqrt{1+\lambda^\top\Sigma\lambda}}\right)}\prod_{i=1}^{n} G_i'(y_i)\,du_n\ldots du_2du_1.$$

As the calculation of these integrals is analytically unfeasible, there are no known closed expressions for the cumulative distribution function (CDF) of these distributions. Due to the complexity of the multivariate integrals involved, numerical methods were used to evaluate these functions.

To overcome computational difficulties, adaptive integration techniques have been explored, such as those presented by (Piessens et al., 1983), which dynamically adjust the integration process to achieve the desired accuracy, efficiently dealing with function variations and complexities. Algorithms such as Vegas, Suave, Divonne and Cuhre, described by (Hahn, 2005), use these methods to refine estimates based on the local characteristics of the integrand, being especially effective when dealing with multivariate distributions with complex behaviors.

The process begins with dividing the integration interval into smaller subintervals, allowing simultaneous evaluation of the function in different regions. The function is evaluated at strategic points within these subintervals, capturing the overall shape of the integrand. With these initial evaluations, estimates of the integral are obtained by methods such as the trapezoid or Simpson's rule, serving as a basis for subsequent refinements.

If the desired precision is not achieved, the algorithm subdivides the intervals with the greatest errors again, introducing new evaluation points to improve estimates in problem areas. Techniques such as adaptive subdivision and Monte Carlo sampling with variance reduction are applied iteratively until the difference between successive estimates is below a tolerance limit, ensuring the accuracy of the integral calculation.

Appendix B

Monte Carlo Simulation

In this section, we present graphics for $G_i = \log_q(x)$, derived from Monte Carlo simulations using sample sizes ranging from 200 to 1000, with 100 replicas for each size. The parameter q varied across the set $\{0.75, 0.85, 0.95, 1.05, 1.15\}$. The grid of initial values considered was:

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

The metrics used were Relative Bias (RB) and Root Mean Square Error (RMSE). Overall, we observe that both metrics converge towards zero.

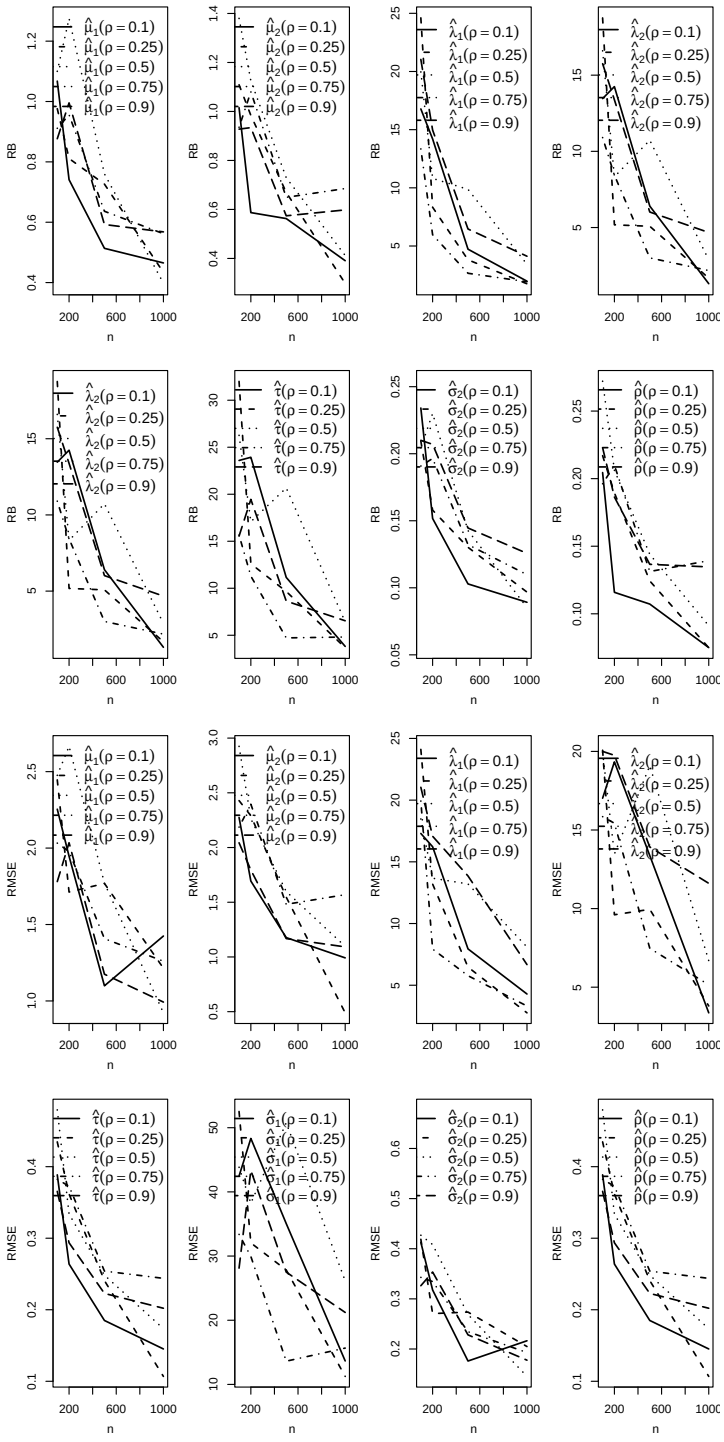


Figure B.1: Relative bias for and root mean squared error for $G_i(x) = \log_{0.75}(x)$.

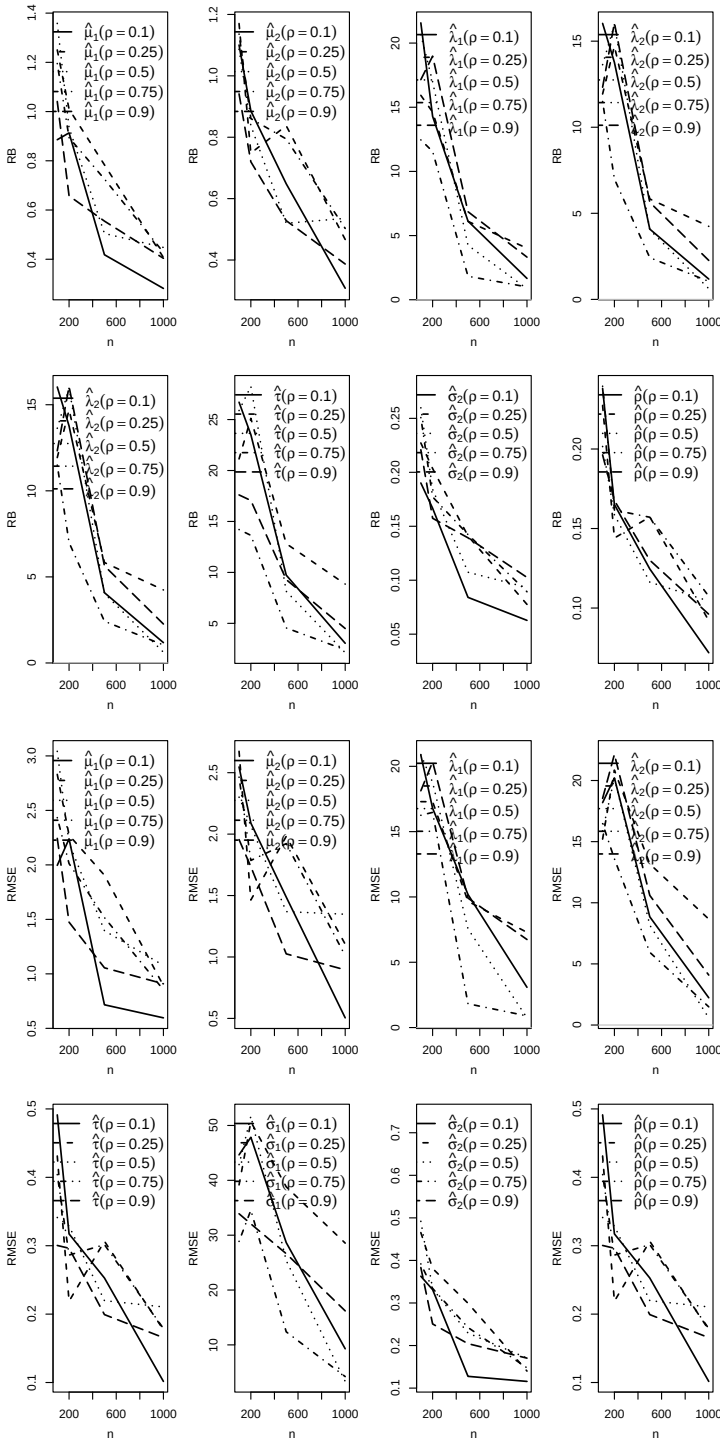


Figure B.2: Relative bias for and root mean squared error for $G_i(x) = \log_{0.85}(x)$.

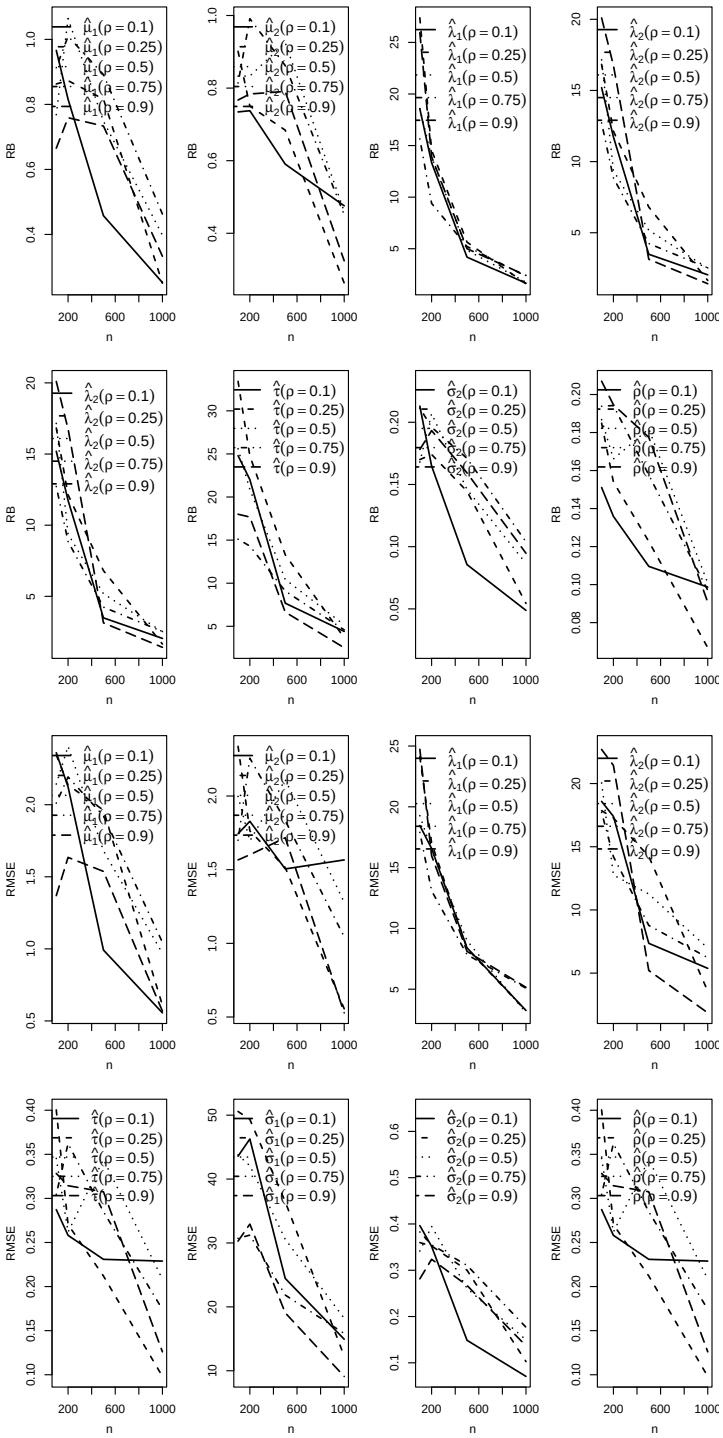


Figure B.3: Relative bias for and root mean squared error for $G_i(x) = \log_{0.95}(x)$.

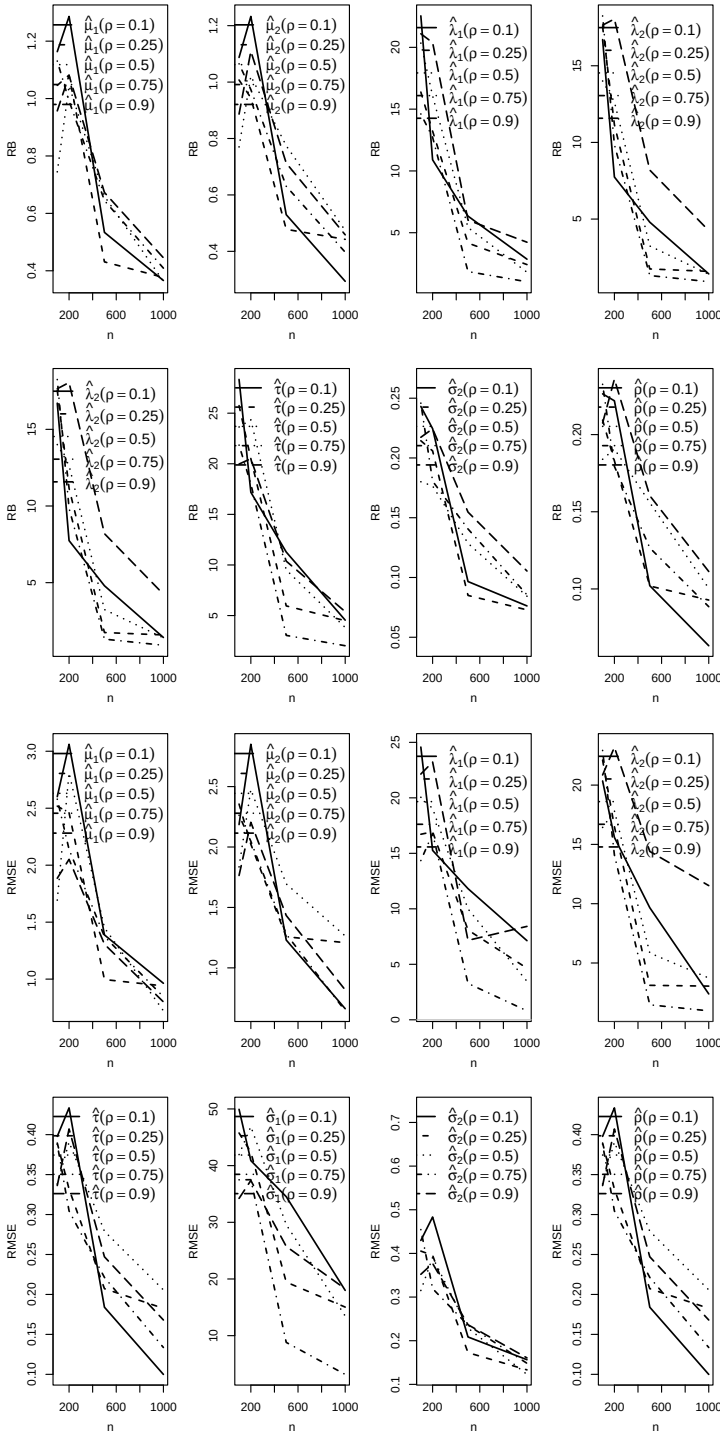


Figure B.4: Relative bias for and root mean squared error for $G_i(x) = \log_{1.05}(x)$.

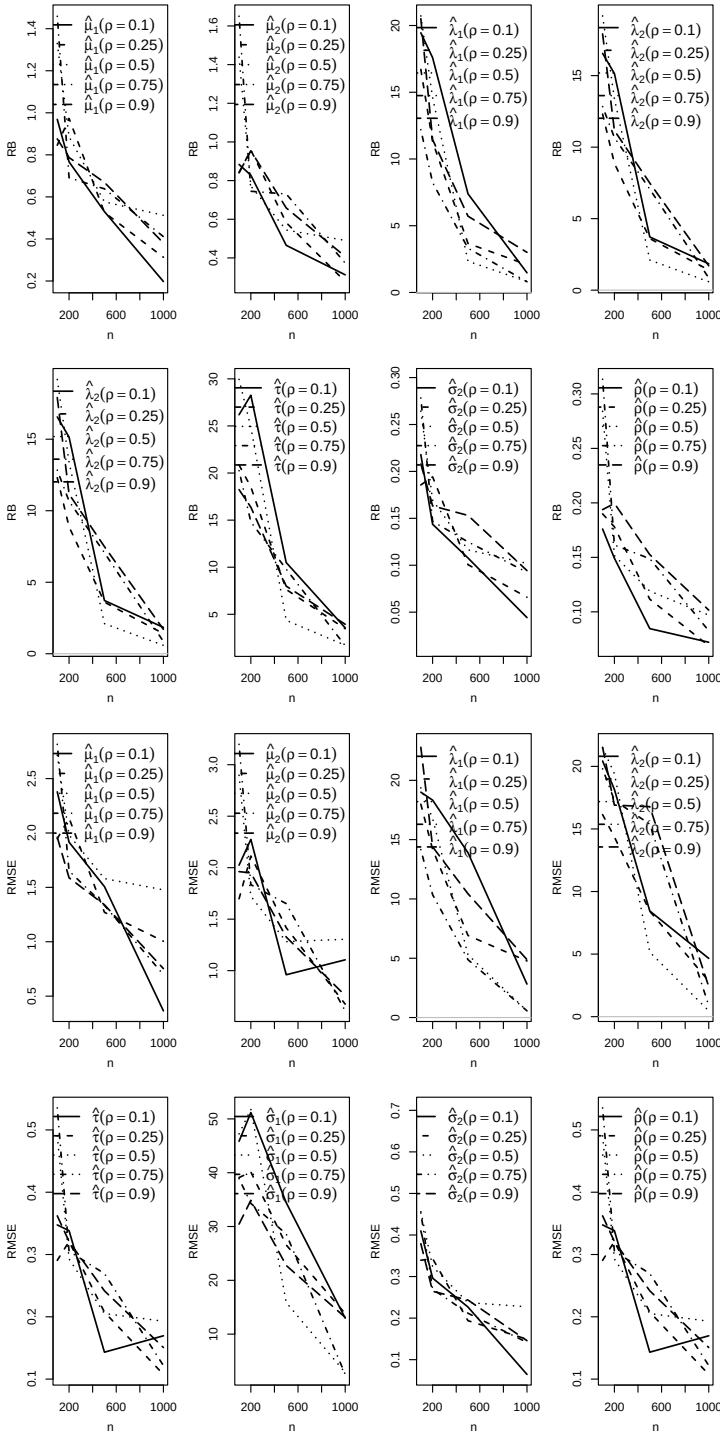


Figure B.5: Relative bias for and root mean squared error for $G_i(x) = \log_{1.15}(x)$.

References

Akaike, Hirotogu (1998). “Information theory and an extension of the maximum likelihood principle”. In: *Selected papers of hirotugu akaike*. Springer, pp. 199–213.

Anderson, Theodore W (1962). “On the distribution of the two-sample Cramer-von Mises criterion”. *The Annals of Mathematical Statistics*, pp. 1148–1159.

Anderson, Theodore W and Darling, Donald A (1952). “Asymptotic theory of certain" goodness of fit" criteria based on stochastic processes”. *The annals of mathematical statistics*, pp. 193–212.

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

Arellano-Valle, Reinaldo B and Genton, Marc G (2005). “On fundamental skew distributions”. *Journal of Multivariate Analysis* 96.1, pp. 93–116.

— (2010b). “Multivariate extended skew-t distributions and related families”. *Metron* 68, pp. 201–234.

Azzalini, Adelchi (1985). “A class of distributions which includes the normal ones”. *Scandinavian journal of statistics*, pp. 171–178.

Azzalini, Adelchi and Valle, A Dalla (1996). “The multivariate skew-normal distribution”. *Biometrika* 83.4, pp. 715–726.

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

Becker, RA, Chambers, JM, and Wilks, AR (1988). "The New S Language. Wadsworth & Brooks/Cole". *Computer Science Series, Pacific Grove, CA*.

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

Cain, Meghan K, Zhang, Zhiyong, and Yuan, Ke-Hai (2017). "Univariate and multivariate skewness and kurtosis for measuring nonnormality: Prevalence, influence and estimation". *Behavior research methods* 49, pp. 1716–1735.

Capitanio, A, Azzalini, Adelchi, and Stanghellini, Elena (2003). "Graphical models for skew-normal variates". *Scandinavian Journal of Statistics* 30.1, pp. 129–144.

Cholesky, André-Louis (2005). "Sur la résolution numérique des systèmes d'équations linéaires". *Bulletin de la Sabix. Société des amis de la Bibliothèque et de l'Histoire de l'École polytechnique* 39, pp. 81–95.

Cook, R. Dennis and Weisberg, Sanford (2009). *An Introduction to Regression Graphics*. Vol. 405. New York: John Wiley & Sons.

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

Dunn, Peter K and Smyth, Gordon K (1996). "Randomized quantile residuals". *Journal of Computational and graphical statistics* 5.3, pp. 236–244.

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.

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.

Hahn, Thomas (2005). "Cuba—a library for multidimensional numerical integration". *Computer Physics Communications* 168.2, pp. 78–95.

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.

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

Kolmogorov (1933). "Sulla determinazione empirica di una legge didistribuzione". *Giorn Dell'inst Ital Degli Att* 4, pp. 89–91.

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.

Leiva, Víctor (2015). "The Birnbaum-Saunders Distribution".

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

Nadarajah, Saralees and Kotz, Samuel (2003). "Skewed distributions generated by the normal kernel". *Statistics & probability letters* 65.3, pp. 269–277.

Nair, Jayakrishnan, Wierman, Adam, and Zwart, Bert (2024). *The Fundamentals of Heavy Tails: Properties, Emergence, and Estimation*. New Edition. Cambridge Series in Statistical and Probabilistic Mathematics 53. Cambridge University Press.

O'Hagan, A. and Leonard, Tom (Apr. 1976). "Bayes estimation subject to uncertainty about parameter constraints". *Biometrika* 63.1, pp. 201–203. ISSN: 0006-3444. DOI: 10.1093/biomet/63.1.201. eprint: https://academic.oup.com/biomet/article-pdf/63/1/201/6689378/63-1-201.pdf. URL: https://doi.org/10.1093/biomet/63.1.201.

Piessens, Robert et al. (1983). "Quadpack". *Springer Series in Computational Mathematics*.

Sahu, Sujit K, Dey, Dipak K, and Branco, Márcia D (2003). "A new class of multivariate skew distributions with applications to Bayesian regression models". *Canadian Journal of Statistics* 31.2, pp. 129–150.

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

Schwarz, Gideon (1978). "Estimating the dimension of a model". *The annals of statistics*, pp. 461–464.

Smirnov, Nikolai V (1939). "On the estimation of the discrepancy between empirical curves of distribution for two independent samples". *Bull. Math. Univ. Moscou* 2.2, pp. 3–14.

Tsallis, Constantino (1988). "Possible Generalization of Boltzmann-Gibbs Statistics". *Journal of Statistical Physics* 52.1-2, pp. 479–487. DOI: 10.1007/BF01016429.

Vaart, A. W. van der (1998). *Asymptotic Statistics*. Cambridge Series in Statistical and Probabilistic Mathematics. Cambridge University Press.

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

Vila, Roberto et al. (2023). "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.

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.

Yamano, Takuya (2002). "Some properties of q-logarithm and q-exponential functions in Tsallis statistics". *Physica A: Statistical Mechanics and its Applications* 305.3-4, pp. 486–496. DOI: 10.1016/S0378-4371(01)00548-7.

Zuo, Baishuai, Balakrishnan, Narayanaswamy, and Yin, Chuancun (2023). *An analysis of multivariate measures of skewness and kurtosis of skew-elliptical distributions*. arXiv: 2311.18176 [math.ST].