University of Brasília
Institute of Exact Sciences
Department of Statistics

Master's Dissertation

Bivariate Log-symmetric Models: Theoretical Properties and Parameter Estimation

by

Ana Lívia Protázio Sá

Brasília, January 6, 2023

Bivariate Log-symmetric Models: Theorical Properties and Parameter Estimation

by

Ana Lívia Protázio Sá

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.

Supervisor: Prof. Dr. Roberto Vila Gabriel

Brasília, January 6, 2023

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. Helton Saulo Bezerra dos Santos

Co-supervisor, EST/UnB

Prof. Dr. José Augusto Fiorucci

EST/UnB

Prof. Dr. Jeremias da Silva Leão

DE/UFAM

To my family and friends.

Acknowledgments

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

Resumo

MODELOS LOG-SIMÉTRICOS BIVARIADOS: PROPRIEDADES TEÓRICAS E ESTIMAÇÃO DE PARÂMETROS

A distribuição gaussiana bivariada tem sido a base da probabilidade e da estatística por muitos anos. No entanto, esta distribuição enfrenta alguns problemas, principalmente devido ao fato de que muitos fenômenos do mundo real geram dados que seguem distribuições assimétricas. Modelos log-simétricos bidimensionais possuem propriedades atrativas e podem ser considerados boas alternativas para resolver este problema, pois possuem propriedades estatísticas que podem torná-las preferíeis a distribuição guassiana. Nesta dissertação, propomos novas caracterizações de distribuições log-simétricas bivariadas e suas aplicações. Esta dissertação visa desenvolver importantes contribuições para a estatística probabilística, teórica e aplicada devido à flexibilidade e propriedades interessantes dos modelos descritos. Teoricamente, uma distribuição é log-simétrica quando a variável aleatória correspondente e sua recíproca têm a mesma distribuição (ver Jones 2008). Uma caracterização de distribuições desse tipo pode ser construída tomando a função logaritmo de uma variável aleatória simétrica. Portanto, distribuições log-simétricas são usadas para descrever o comportamento de dados estritamente positivos. A classe desse tipo de distribuição é bastante ampla e inclui grande parte das distribuições bimodais e aquelas com caudas mais leves ou mais pesadas que a distribuição log-normal; ver, por exemplo, Vanegas e Paula (2016). Alguns exemplos de distribuições log-simétricas são:

log-normal, log-Student-t, log-logistic, log-Laplace, log-Cauchy, log-power-exponencial, log-slash, harmonic law, Birnbaum-Saunders, e Birnbaum-Saunders-t; ver, por exemplo, Crow e Shimizu (1988), Birnbaum e Saunders (1969), Rieck e Nedelman (1991), Johnson et al. (1994), 1995, Díaz-García e Leiva (2005), Marshall e Olkin (2007), Jones (2008) e Vanegas e Paula (2016). Estudamos as principais propriedades estatísticas dos modelos, no capítulo 1 apresentemos o modelo log-simétrico bivariado (BLS) proposto, ademais neste capítulo as principais propriedades matemáticas, como representação estocástica,função quantil, distribuição condicional, distância Mahalanobis, independência, momentos, função de correlação, entre outras propriedades do modelo BLS são discutidas. No capítulo 2, propomos o método de máxima verossimilhança para a estimação dos parâmetros das distribuições propostas. No capítulo 3, realizamos a simulação de Monte Carlo para avaliar a performance dos estimadores de máxima verossimilhança, utilizando o viés e o Erro Quadrático Médio, considerando vários cenários para diferentes distribuições, o que mostrou bons resultados com valores próximos a zero. No Capítulo 4, realizamos a aplicação a um conjunto de dados reais refentes a *fatigue*, os dados são baseados no artigo de Marchant et al. (2015), no qual ele propôs um modelo de regressão multivariado Birnbaum-Saunders, realizamos a estimação dos parâmetros utilizando o método de Máxima verossimilhança e usamos as seguintes váriaveis *Von Mises stress* (T_1) e *die limetime* (T_2), para a estimação dos parâmetros extras utilizamos estimação perfilada, além disso computamos os valores de critério de informação de Akaike (AIC) e Bayesiano (BIC), para utilizarmos como critério de seleção de modelo. Os resultados são vistos como favoráveis ao modelo log-Laplace

Palavras-Chaves: Modelos Log-Simétricos Bivariados, simulação de Monte Carlo, método de máxima verossimilhança, software R.

Abstract

The bivariate Gaussian distribution has been the basis of probability and statistics for many years. Nonetheless, this distribution faces some problems, mainly due to the fact that many real-world phenomena generate data that follow asymmetric distributions. Bidimensional log-symmetric models have attractive properties and can be considered as good alternatives to solve this problem. In this dissertation, we propose new characterizations of bivariate log-symmetric distributions and their applications. This dissertation aims to develop important contributions to probability, theoretical and applied statistics due to the flexibility and interesting properties of the outlined models. We implemented maximum likelihood estimation for the parameters of the distributions. A Monte Carlo simulation study was performed to evaluate the performance of the parameter estimation. Finally, we applied the proposed methodology to a real data set.

Keywords: Bivariate Log-symmetric Models, Monte Carlo simulation, maximum likelihood method, R software.

Contents

List of Tables

List of Figures

Introduction

Theoretically, a distribution is log-symmetric when the corresponding random variable and its reciprocal have the same distribution (see Jones 2008). A characterization of distributions of this type can be constructed by taking the logarithm function of a symmetric random variable. Therefore, log-symetric distributions are used to describe the behavior of strictly positive data. The class of this type of distribution is quite broad and includes a large portion of bimodal distributions and those with lighter or heavier tails than the log-normal distribution; see e.g., Vanegas and Paula (2016). Some examples of log-symmetric distributions are: log-normal, log-Student-t, log-logistic, log-Laplace, log-Cauchy, log-power-exponential, log-slash, harmonic law, Birnbaum-Saunders, and Birnbaum- Saunders-t; see e.g., Crow and Shimizu (1988), Birnbaum and Saunders (1969), Rieck and Nedelman (1991), Johnson et al. (1994), 1995, Díaz-García and Leiva (2005), Marshall and Olkin (2007), Jones (2008), and Vanegas and Paula (2016).

Another important feature of the log-symmetric class is that they are closed under scale change and under reciprocity, according to Puig (2008), which are very desirable properties for distributions that are used to describe strictly positive data, and log-symmetric models allow you to model the median or the asymmetry (relative dispersion).

Furthermore, the log-symmetric class has statistical properties that might make it preferable to the alternative distribution. For example, the two parameters of the log-symmetric distribution are orthogonal and they can be interpreted directly as median and skewness (or relative dispersion, taking into account two parameters that are interpreted as measures of position and scale, as stated by Vanegas and Paula (2016), which are, in the context of asymmetric distributions, the ones that mean the most being complete measures of location and shape, respectively. Several studies have been carried out including the log-symmetric distributions proposed in Vanegas and Paula (2016).

Within this context, the main objective of the current work is to extend in a natural way the definition of univariate log-symmetric distributions to the bivariate case, to study its main statistical properties, to propose the maximum likelihood method for the estimation parameters and to show an application to real data. The remainder of this work is organized as follows.

In Chapter 1 the bivariate log-symmetric (BLS) model is proposed. Moreover, in this section, the main mathematical properties, as stochastic representation, quantile function, conditional distribution, Mahalanobis distance, independence, moments, correlation function among others; of the BLS model are discussed. In Chapter 2, the maximum likelihood estimator for the bivariate log-symmetric models is proposed. In Chapter 3 we performed Monte Carlo simulation to evaluate the performance of the maximum likelihood estimators. In Chapter 4 we apply the BLS models to a data set. Finally, we close this chapter by presenting some concluding remarks.

Chapter 1

Bivariate Log-Symmetric Model

A continuous random vector $\boldsymbol{T} = \begin{pmatrix} T_1 \\ T_2 \end{pmatrix}$ follows a bivariate log-symmetric (BLS) distribution if its joint probability density function (PDF) is given by

$$f_{T_1,T_2}(t_1, t_2; \boldsymbol{\theta}) = \frac{1}{t_1 t_2 \sigma_1 \sigma_2 \sqrt{1-\rho^2} Z_{g_c}} \, g_c\left(\frac{\widetilde{t_1}^2 - 2\rho \widetilde{t_1}\widetilde{t_2} + \widetilde{t_2}^2}{1-\rho^2}\right), \quad t_1, t_2 > 0, \tag{1.1}$$

where

$$\widetilde{t_i} = \log\left[\left(\frac{t_i}{\eta_i}\right)^{1/\sigma_i}\right], \ \eta_i = \exp(\mu_i), \ i = 1, 2,$$

where $\boldsymbol{\theta} = (\eta_1, \eta_2, \sigma_1, \sigma_2, \rho)$ is the parameter vector with $\mu_i \in \mathbb{R}$, $\sigma_i > 0$, $i = 1, 2$; $\rho \in (-1, 1)$; $Z_{g_c} > 0$ is the partition function, that is,

$$Z_{g_c} = \int_0^\infty \int_0^\infty \frac{1}{t_1 t_2 \sigma_1 \sigma_2 \sqrt{1-\rho^2}} \, g_c\left(\frac{\widetilde{t_1}^2 - 2\rho \widetilde{t_1}\widetilde{t_2} + \widetilde{t_2}^2}{1-\rho^2}\right) \mathrm{d}t_1 \mathrm{d}t_2, \tag{1.2}$$

and g_c is a scalar function referred to as the density generator; see Fang et al. (1990). We use, in this case, the notation $\boldsymbol{T} \sim \mathrm{BLS}(\boldsymbol{\theta}, g_c)$.

In this work we prove that, when it exists, the variance-covariance matrix of a random vector $\boldsymbol{T} \sim \mathrm{BLS}(\boldsymbol{\theta}, g_c)$, denoted by $K_{\boldsymbol{T}}$, is a matrix function of the following dispersion matrix (see subsections 1.1.7 and 1.1.8 below):

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

In other words, $K_{\boldsymbol{T}} = \psi(\boldsymbol{\Sigma})$ for some matrix function $\psi : \mathcal{M}_{2,2} \longmapsto \mathcal{M}_{2,2}$, where $\mathcal{M}_{2,2}$ denotes the set of all 2-by-2 real matrices.

Based on the references Saulo et al. (2017) and Vanegas and Paula (2016), Table 1.1 presents some examples of bivariate log-symmetric distributions.

Table 1.1: Partition functions (Z_{g_c}) and density generators (g_c) for some distributions.

Distribution	Z_{g_c}	g_c	Parameter
Bivariate Log-normal	2π	$\exp(-x/2)$	$-$
Bivariate Log-Kotz type	$\frac{\pi\Gamma(\zeta/\delta)}{\delta\lambda^{\zeta/\delta}}$	$x^{\zeta-1}\exp(-\lambda x^{\delta})$	$\delta > 0, \lambda > 0, \zeta > 0$
Bivariate Log-contaminated normal	$2\pi(\frac{1}{\sqrt{\vartheta_2}} + \frac{1}{\vartheta_1} - 1)$	$\sqrt{\vartheta_2}\exp(-\frac{1}{2}\vartheta_2 x) + \frac{(1-\vartheta_1)}{\vartheta_1}\exp(-\frac{1}{2}x)$	$0 < \vartheta_1, \vartheta_2 < 1$
Bivariate Log-Student-t	$\frac{\Gamma(\nu/2)\nu\pi}{\Gamma((\nu+2)/2)}$	$(1 + \frac{x}{\nu})^{-(\nu+2)/2}$	$\nu > 0$
Bivariate Log-Pearson Type VII	$\frac{\Gamma(\xi-1)\theta\pi}{\Gamma(\xi)}$	$(1 + \frac{x}{\theta})^{-\xi}$	$\xi > 1$, $\theta > 0$
Bivariate Log-hyperbolic	$\frac{2\pi(\nu+1)\exp(-\nu)}{\nu^2}$	$\exp(-\nu\sqrt{1+x})$	$\nu > 0$
Bivariate Log-Laplace	π	$K_0(\sqrt{2x})$	$-$
Bivariate Log-slash	$\frac{\pi}{\nu-1}2^{\frac{3-\nu}{2}}$	$x^{-\frac{\nu+1}{2}}\gamma(\frac{\nu+1}{2}, \frac{x}{2})$	$\nu > 1$
Bivariate Log-power-exponential	$2^{\xi+1}(1+\xi)\Gamma(1+\xi)\pi$	$\exp\left(-\frac{1}{2}x^{1/(1+\xi)}\right)$	$-1 < \xi \leqslant 1$
Bivariate Log-Logistic	$\pi/2$	$\frac{\exp(-x)}{(1+\exp(-x))^2}$	$-$

Here, in the Table 1.1, $\Gamma(t) = \int_0^\infty x^{t-1}\exp(-x)\,\mathrm{d}x$, $t > 0$, is the gamma function, $K_0(u) = \int_0^\infty t^{-1}\exp(-t - \frac{u^2}{4t})\,\mathrm{d}t/2$, $u > 0$, is the Bessel function of the third kind (for more details on the main properties of K_0, see appendix of Kotz et al. 2001); and $\gamma(s, x) = \int_0^x t^{s-1}\exp(-t)\,\mathrm{d}t$ is the lower incomplete gamma function.

In Figures 1.1-1.9 some graphical sketches of the BLS PDF from Table 1.1 are presented.

By using (1.1) it is clear that the random vector $\boldsymbol{X} = \begin{pmatrix} X_1 \\ X_2 \end{pmatrix}$, with $X_i = \log(T_i)$, $i = 1, 2$, has a bivariate elliptically symmetric (BES) distribution; see p. 592 in Balakrishnan and Lai (2009). In other words, the PDF of $\boldsymbol{X}$ is as follows

$$f_{X_1,X_2}(x_1, x_2; \boldsymbol{\theta}_*) = \frac{1}{\sigma_1\sigma_2\sqrt{1-\rho^2}Z_{g_c}}\, g_c\left(\frac{\widetilde{x_1}^2 - 2\rho\widetilde{x_1}\widetilde{x_2} + \widetilde{x_2}^2}{1-\rho^2}\right), \quad -\infty < x_1, x_2 < \infty, \tag{1.3}$$

$$\widetilde{x_i} = \frac{x_i - \mu_i}{\sigma_i},\ i = 1, 2,$$

where $\boldsymbol{\theta}_* = (\mu_1, \mu_2, \sigma_1, \sigma_2, \rho)$ is the parameter vector and Z_{g_c} is the partition function stated in (1.2). In this case, the notation $\boldsymbol{X} \sim \mathrm{BES}(\boldsymbol{\theta}_*, g_c)$ is used.

A simple standard calculation shows that the joint cumulative distribution function (CDF) of $\boldsymbol{T} \sim \mathrm{BLS}(\boldsymbol{\theta}, g_c)$, denoted by $F_{T_1,T_2}(t_1, t_2; \boldsymbol{\theta})$, is expressed as

$$F_{T_1,T_2}(t_1, t_2; \boldsymbol{\theta}) = F_{X_1,X_2}\Big(\log(t_1), \log(t_2); \boldsymbol{\theta}_*\Big),$$

with $F_{X_1,X_2}(x_1, x_2; \boldsymbol{\theta}_*)$ the CDF of $\boldsymbol{X} \sim \mathrm{BES}(\boldsymbol{\theta}_*, g_c)$. Except in the case of bivariate normal, there is no single closed-form expression for the CDF of $\boldsymbol{X}$.

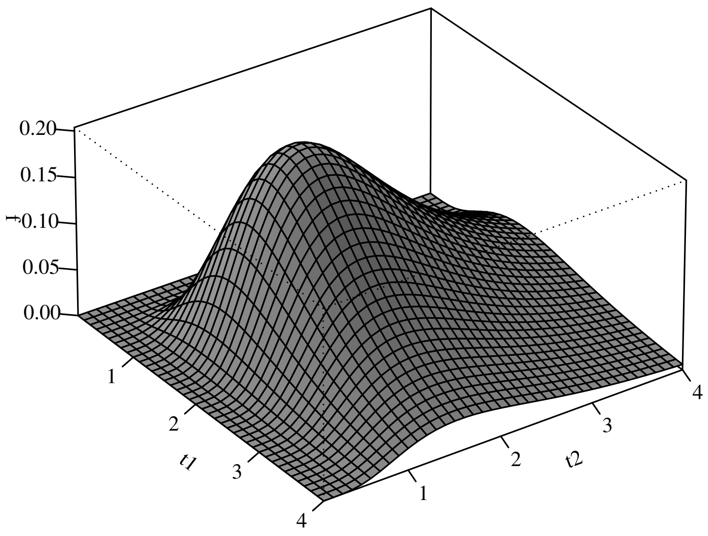

Figure 1.1: Bivariate log-normal PDF plot with parameters $\boldsymbol{\theta}_* = (2, 2, 0.5, 0.5, 0)$.

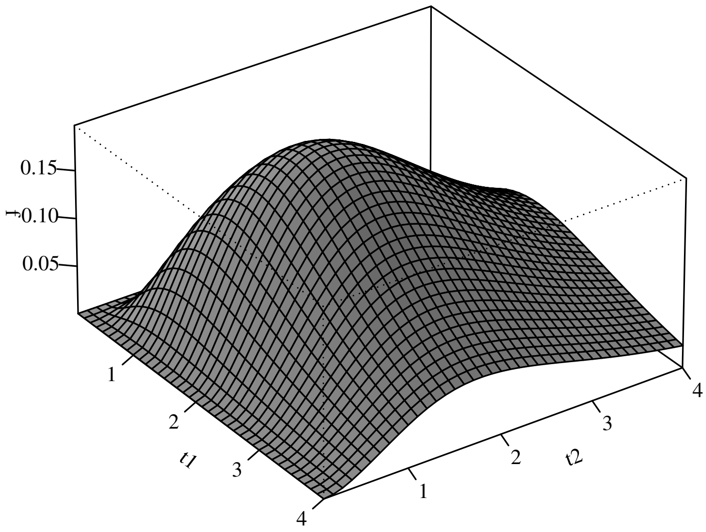

Figure 1.2: Bivariate log-contaminated Normal PDF plot with parameters $\boldsymbol{\theta}_* = (2, 2, 0.5, 0.5, 0)$ and $\vartheta_1 = 0.9, \vartheta_2 = 0.3$.

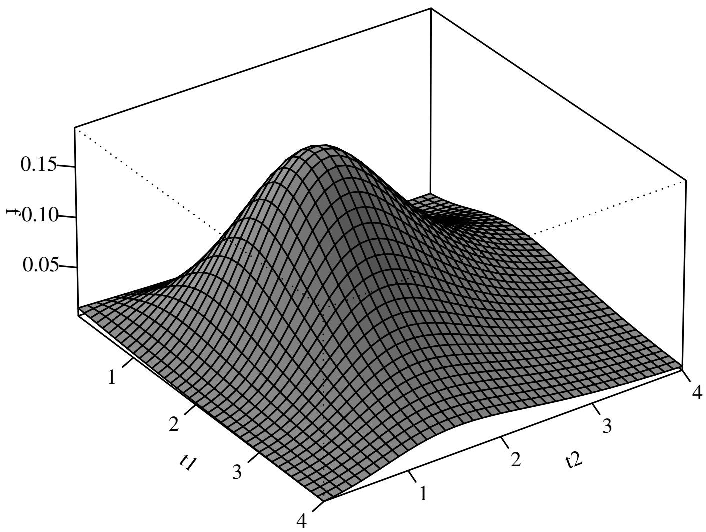

Figure 1.3: Bivariate log-Student-t PDF plot with parameters $\theta_* = (2, 2, 0.5, 0.5, 0)$ and $\nu = 3$.

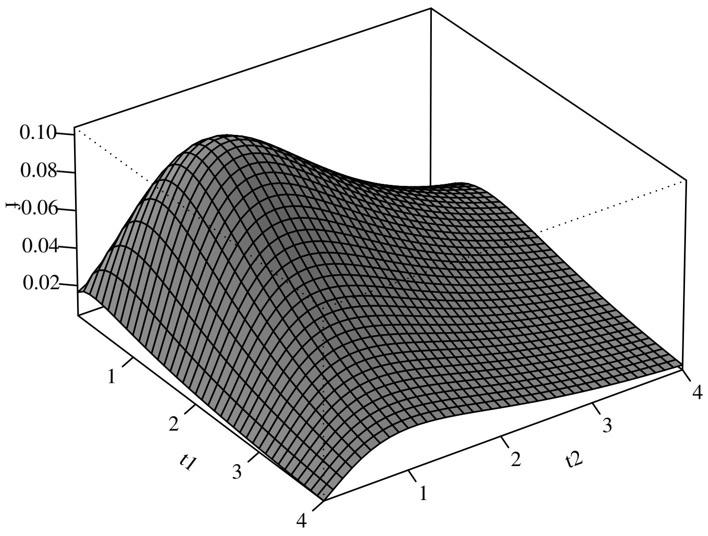

Figure 1.4: Bivariate log-Pearson type VII PDF plot with parameters $\theta_* = (2, 2, 0.5, 0.5, 0)$ and $\xi = 5$, $\theta = 22$.

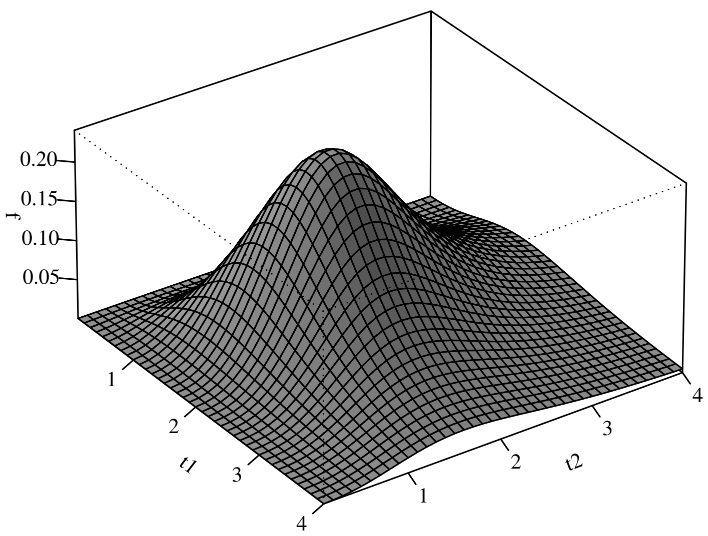

Figure 1.5: Bivariate log-hyperbolic PDF plot with parameters $\theta_* = (2, 2, 0.5, 0.5, 0)$ and $\nu = 2$.

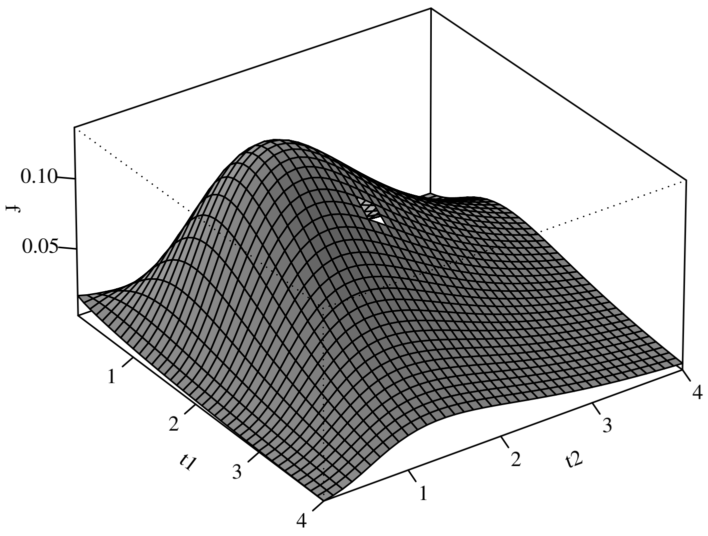

Figure 1.6: Bivariate log-slash PDF plot with parameters $\theta_* = (2, 2, 0.5, 0.5, 0)$ and $\nu = 4$.

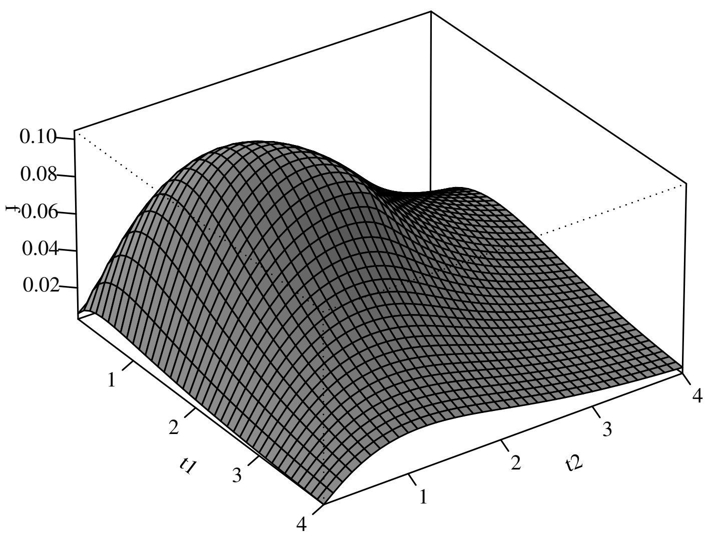

Figure 1.7: Bivariate log-power-exponential PDF plot with parameters $\theta_* = (2, 2, 0.5, 0.5, 0)$ and $\xi = 0.5$.

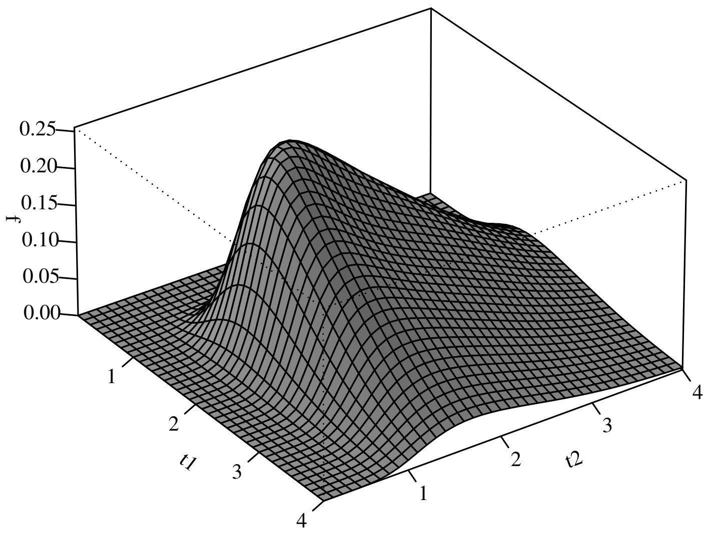

Figure 1.8: Bivariate log-logistic PDF plot with parameters $\boldsymbol{\theta}_* = (2, 2, 0.5, 0.5, 0)$.

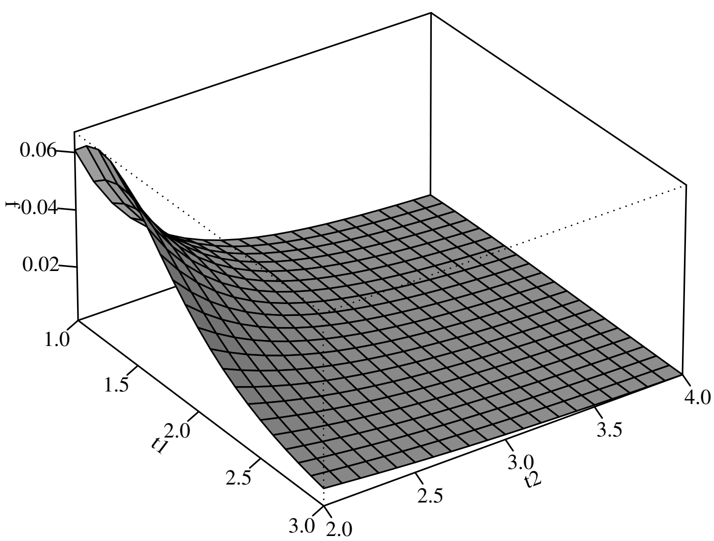

Figure 1.9: Bivariate log-Laplace PDF plot with parameters $\boldsymbol{\theta}_* = (2, 2, 0.5, 0.5, 0)$.

1.1 Some basic properties of model

In this section, some mathematical properties of proposed bivariate log-symmetric distribution are discussed.

1.1.1 Characterization of the partition function Z_{g_c}

Proposition 1.1.1. *The partition function Z_{g_c} (1.2) is independent of the parameter vector $\boldsymbol{\theta}$. More precisely,*

$$Z_{g_c} = \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} g_c\Big({z_1}^2 + {z_2}^2\Big)\,\mathrm{d}z_1\mathrm{d}z_2 = \pi\int_0^{\infty} g_c(u)\,\mathrm{d}u.$$

Proof. The proof of the first identity follows by considering in (1.2) the following change of variables (Jacobian Method):

$$z_1 = \widetilde{t}_1, \quad z_2 = \frac{\widetilde{t}_2 - \rho\widetilde{t}_1}{\sqrt{1-\rho^2}},$$

where $\widetilde{t}_i$, $i = 1, 2$, are defined in (1.1).

On the other hand, by using integration in polar coordinates: $z_1 = r\cos(\theta)$, $z_2 = r\sin(\theta)$, with $r \geqslant 0$ and $0 \leqslant \theta \leqslant 2\pi$ in the identity $Z_{g_c} = \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} g_c\Big({z_1}^2 + {z_2}^2\Big)\,\mathrm{d}z_1\mathrm{d}z_2$, we have

$$Z_{g_c} = \int_0^{2\pi}\int_0^{\infty} g_c(r^2)r\,\mathrm{d}r\mathrm{d}\theta.$$

Hence, by take the change of variables $u = r^2$, $\mathrm{d}u = 2r\mathrm{d}r$, the proof of the second identity follows. □

1.1.2 Stochastic representation

Proposition 1.1.2. *The random vector* $\boldsymbol{T} = (T_1, T_2)$ *has a BLS distribution if*

$$
\begin{aligned}
T_1 &= \eta_1 \exp(\sigma_1 Z_1),\\
T_2 &= \eta_2 \exp\left(\sigma_2 \rho Z_1 + \sigma_2 \sqrt{1-\rho^2} Z_2\right),
\end{aligned}
$$

where $Z_1 = RDU_1$ *and* $Z_2 = R\sqrt{1-D^2}U_2$*;* U_1*,* U_2*,*R*, and* D *are mutually independent random variables,* $\rho \in (-1, 1)$*,* $\eta_i = \exp(\mu_i)$*, and* $\mathbb{P}(U_i = -1) = \mathbb{P}(U_i = 1) = 1/2$*,* $i = 1, 2$*. The random variable* D *is positive and has PDF*

$$
f_D(d) = \frac{2}{\pi\sqrt{1-d^2}}, \quad d \in (0, 1).
$$

Furthermore, the positive random variable R *is called the generator of the elliptical random vector* $\boldsymbol{X} = (X_1, X_2)$*. In other words,* R *has PDF given by*

$$
f_R(r) = \frac{2r g_c(r^2)}{\int_0^\infty g_c(u)\,\mathrm{d}u}, \quad r > 0.
$$

Proof. It is well-known that (see Abdous et al. 2005), the vector $\boldsymbol{X}$ has a BES distribution if

$$
\begin{aligned}
X_1 &= \mu_1 + \sigma_1 Z_1,\\
X_2 &= \mu_2 + \sigma_2 \rho Z_1 + \sigma_2 \sqrt{1-\rho^2} Z_2.
\end{aligned}
\tag{1.4}
$$

Since $X_i = \log(T_i)$, $i = 1, 2$, the proof follows. □

1.1.3 Quantile function

Let $\boldsymbol{T} = (T_1, T_2) \sim \text{BLS}(\boldsymbol{\theta}, g_c)$ and $p \in (0,1)$. By using the stochastic representation of Proposition 1.1.2, we obtain

$$p = \mathbb{P}(T_1 \leqslant Q_{T_1}) = \mathbb{P}\Big(\eta_1 \exp(\sigma_1 Z_1) \leqslant Q_{T_1}\Big) = \mathbb{P}\left(Z_1 \leqslant \log\left[\left(\frac{Q_{T_1}}{\eta_1}\right)^{1/\sigma_1}\right]\right)$$

and

$$\begin{aligned} p = \mathbb{P}(T_2 \leqslant Q_{T_2}) &= \mathbb{P}\Big(\eta_2 \exp(\sigma_2 \rho Z_1 + \sigma_2\sqrt{1-\rho^2} Z_2) \leqslant Q_{T_2}\Big) \\ &= \mathbb{P}\left(\rho Z_1 + \sqrt{1-\rho^2} Z_2 \leqslant \log\left[\left(\frac{Q_{T_2}}{\eta_2}\right)^{1/\sigma_2}\right]\right). \end{aligned}$$

Hence, the p-quantile of T_1 and the p-quantile of T_2 are given by

$$\log\left[\left(\frac{Q_{T_1}}{\eta_1}\right)^{1/\sigma_1}\right] = Q_{Z_1} \quad \Longleftrightarrow \quad Q_{T_1} = \eta_1 \exp(\sigma_1 Q_{Z_1})$$

and

$$\log\left[\left(\frac{Q_{T_2}}{\eta_2}\right)^{1/\sigma_2}\right] = Q_{\rho Z_1 + \sqrt{1-\rho^2} Z_2} \quad \Longleftrightarrow \quad Q_{T_2} = \eta_2 \exp(\sigma_2 Q_{\rho Z_1 + \sqrt{1-\rho^2} Z_2}),$$

respectively.

1.1.4 Conditional distribution

Proposition 1.1.3. *The joint PDF of Z_1 and Z_2, given in Proposition 1.1.2, is given by*

$$f_{Z_1,Z_2}(z_1, z_2) = \frac{g_c(z_1^2 + z_2^2)}{\pi \int_0^\infty g_c(u)\,\mathrm{d}u}, \quad -\infty < z_1, z_2 < \infty.$$

Moreover, the marginal PDFs of Z_1 and Z_2, denoted by f_{Z_1} and f_{Z_2}, respectively, are given by

$$f_{Z_1}(z_1) = \frac{\displaystyle\int_{|z_1|}^{\infty} \frac{2g_c(w^2)}{\sqrt{1-\frac{z_1^2}{w^2}}}\, \mathrm{d}w}{\pi \int_0^{\infty} g_c(u)\, \mathrm{d}u} \quad \text{and} \quad f_{Z_2}(z_2) = \frac{\displaystyle\int_{|z_2|}^{\infty} \frac{2g_c(w^2)}{\sqrt{1-\frac{z_2^2}{w^2}}}\, \mathrm{d}w}{\pi \int_0^{\infty} g_c(u)\, \mathrm{d}u}, \quad -\infty < z_1, z_2 < \infty.$$

In particular, $f_{Z_i}(z_i \,|\, Z_j = z_j)$ $(i \neq j)$ and $f_{Z_i}(z_i)$, $i, j = 1, 2$, are even functions.

Proof. The proof of this result is technical and intermediate in the arguments of the proofs developed in this work, so we decided not to put it. For more details on the proof, see Propositions 3.1 and 3.2 of reference Saulo et al. (2022). □

Definition 1.1.1. *Let X and Y be two continuous random variables with joint PDF $f_{X,Y}$, and marginal PDFs f_X and f_Y, respectively.*

- *Let B be a Borelian subset of $\mathbb{R}$. The conditional CDF of X given $\{Y \in B\}$, denoted by $F_X(x \,|\, Y \in B)$, is defined as (for every x)*

$$F_X(x \,|\, Y \in B) = \mathbb{P}(X \leqslant x \,|\, Y \in B) = \int_{-\infty}^{x} f_X(u \,|\, Y \in B)\, \mathrm{d}u, \quad \text{if } \mathbb{P}(Y \in B) > 0, \quad (1.5)$$

  where $f_X(u \,|\, Y \in B)$ is the corresponding conditional PDF given by

$$f_X(u \,|\, Y \in B) = \frac{\int_B f_{X,Y}(u, v)\, \mathrm{d}v}{\mathbb{P}(Y \in B)}.$$

  We write $X \,|\, Y \in B$ to indicate that the random variable X follows the conditional CDF (1.5) given $\{Y \in B\}$.

- *Let* $\varepsilon > 0$*, and suposse that* $\mathbb{P}(y-\varepsilon < Y \leqslant y+\varepsilon) > 0$*. Abusing mathematical notation, we define the conditional CDF of* X *given* $Y = y$*, denoted by* $F_X(x \,|\, Y = y)$*, as (for every* x*)*

$$F_X(x|Y=y) = \lim_{\varepsilon\to 0^+} \mathbb{P}(X \leqslant x \,|\, y-\varepsilon < Y \leqslant y+\varepsilon),$$

provided that the limit exists. If the limit exists, there is a nonnegative function $f_X(u|Y=y)$ *(called the conditional PDF) so that (for every* x*)*

$$F_X(x \,|\, Y = y) = \int_{-\infty}^{x} f_X(u \,|\, Y = y)\,\mathrm{d}u.$$

At every point (x, y) *at which* $f_{X,Y}$ *is continuous and* $f_Y(y) > 0$ *is continuous, the PDF* $f_X(u \,|\, Y = y)$ *exists and it is expressed by (see Theorem 6, p. 109, of Rohatgi and Saleh 2015)*

$$f_X(u \,|\, Y = y) = \frac{f_{X,Y}(u, y)}{f_Y(y)}.$$

For simplicity, we write $X \,|\, Y = y$*.*

Lemma 1.1.4. *If* $\boldsymbol{T} = (T_1, T_2) \sim \mathrm{BLS}(\boldsymbol{\theta}, g_c)$ *then the PDF of* $T_2|T_1 = t_1$ *is written as*

$$f_{T_2}(t_2 \,|\, T_1 = t_1) = \frac{1}{t_2\sigma_2\sqrt{1-\rho^2}}\, f_{Z_2}\left(\frac{1}{\sqrt{1-\rho^2}}\,\widetilde{t}_2 - \frac{\rho}{\sqrt{1-\rho^2}}\,\widetilde{t}_1 \,\middle|\, Z_1 = \widetilde{t}_1\right), \tag{1.6}$$

where $\widetilde{t}_i$*,* $i = 1, 2$*, are defined in* (1.1)*, and* Z_1 *and* Z_2 *are as in Proposition* 1.1.2*.*

Proof. If $T_1 = t_1$, then $Z_1 = \log\left[(t_1/\eta_1)^{1/\sigma_1}\right] = \widetilde{t}_1$. Thus, the conditional distribution of T_2 given $T_1 = t_1$ is the same as the distribution of

$$\eta_2 \exp\left(\sigma_2\rho\widetilde{t}_1 + \sigma_2\sqrt{1-\rho^2}Z_2\right) \Big|\, T_1 = t_1.$$

Consequently,

$$
\begin{aligned}
F_{T_2}(t_2 \,|\, T_1 = t_1) &= \mathbb{P}\Big(\eta_2 \exp(\sigma_2 \rho \widetilde{t}_1 + \sigma_2 \sqrt{1-\rho^2} Z_2) \leqslant t_2 \,\Big|\, T_1 = t_1\Big) \\
&= \mathbb{P}\left(Z_2 \leqslant \frac{1}{\sqrt{1-\rho^2}}\, \widetilde{t}_2 - \frac{\rho}{\sqrt{1-\rho^2}}\, \widetilde{t}_1 \,\middle|\, Z_1 = \widetilde{t}_1 \right).
\end{aligned}
$$

Then, the Formula (1.6) of the conditional PDF of T_2 given $T_1 = t_1$ follows. □

Theorem 1.1.5. *For a Borelian subset B of $(0, \infty)$, we define the following Borelian set:*

$$
B_\rho = \frac{1}{\sqrt{1-\rho^2}} \log\left[\left(\frac{B}{\eta_2}\right)^{1/\sigma_2}\right] - \frac{\rho}{\sqrt{1-\rho^2}}\, \widetilde{t}_1, \tag{1.7}
$$

where $\widetilde{t}_1$ is as in (1.1). *If $\boldsymbol{T} = (T_1, T_2) \sim \mathrm{BLS}(\boldsymbol{\theta}, g_c)$ then the PDF of $T_1 \,|\, T_2 \in B$ is written as*

$$
f_{T_1}(t_1 | T_2 \in B) = \frac{1}{t_1 \sigma_1}\, f_{Z_1}(\widetilde{t}_1)\, \frac{\int_{B_\rho} f_{Z_2}(w \,|\, Z_1 = \widetilde{t}_1)\, \mathrm{d}w}{\mathbb{P}\Big(\rho Z_1 + \sqrt{1-\rho^2} Z_2 \in B_0\Big)},
$$

with Z_1 and Z_2 as in Proposition 1.1.2.

Proof. Let B be a Borelian subset of $(0, \infty)$. Notice that

$$
f_{T_1}(t_1 \,|\, T_2 \in B) = f_{T_1}(t_1)\, \frac{\int_B f_{T_2}(t_2 \,|\, T_1 = t_1)\, \mathrm{d}t_2}{\mathbb{P}(T_2 \in B)}.
$$

Since $f_{T_1}(t_1) = f_{Z_1}(\widetilde{t}_1)/(\sigma_1 t_1)$ and $\mathbb{P}(T_2 \in B) = \mathbb{P}\Big(\rho Z_1 + \sqrt{1-\rho^2} Z_2 \in B_0\Big)$, where B_0 is given in (1.7) with $\rho = 0$, the term on the right-hand side of the above identity is

$$
= \frac{1}{\sigma_1 t_1}\, f_{Z_1}(\widetilde{t}_1)\, \frac{\int_B f_{T_2}(t_2 \,|\, T_1 = t_1)\, \mathrm{d}t_2}{\mathbb{P}\Big(\rho Z_1 + \sqrt{1-\rho^2} Z_2 \in B_0\Big)}.
$$

By using the expression for $f_{T_2}(t_2 \,|\, T_1 = t_1)$ provided by Lemma 1.1.4, the previous expression is

$$= \frac{1}{t_1\sigma_1\sigma_2\sqrt{1-\rho^2}}\, f_{Z_1}(\widetilde{t}_1)\, \frac{\int_B \frac{1}{t_2}\, f_{Z_2}\Big(\frac{1}{\sqrt{1-\rho^2}}\,\widetilde{t}_2 - \frac{\rho}{\sqrt{1-\rho^2}}\,\widetilde{t}_1 \,\Big|\, Z_1 = \widetilde{t}_1\Big)\,\mathrm{d}t_2}{\mathbb{P}\Big(\rho Z_1 + \sqrt{1-\rho^2} Z_2 \in B_0\Big)},$$

where $\widetilde{t}_i$, $i = 1, 2$, are as in (1.1). Finally, by applying the change of variable $w = (\widetilde{t}_2 - \rho\,\widetilde{t}_1)/\sqrt{1-\rho^2}$, the above expression is

$$= \frac{1}{t_1\sigma_1}\, f_{Z_1}(\widetilde{t}_1)\, \frac{\int_{B_\rho} f_{Z_2}(w \,|\, Z_1 = \widetilde{t}_1)\,\mathrm{d}w}{\mathbb{P}\Big(\rho Z_1 + \sqrt{1-\rho^2} Z_2 \in B_0\Big)},$$

in which B_ρ is as in (1.7). Thus, we have completed the proof. □

Corollary 1.1.6 (Gaussian generator)**.** *Let $\boldsymbol{T} = (T_1, T_2) \sim \mathrm{BLS}(\boldsymbol{\theta}, g_c)$ and $g_c(x) = \exp(-x/2)$ be the generator of the bivariate log-normal distribution. Then, for each Borelian subset B of $(0,\infty)$, the PDF of $T_1 \,|\, T_2 \in B$ is given by (for $t_1 > 0$)*

$$f_{T_1}(t_1|T_2 \in B) = \tfrac{1}{t_1\sigma_1}\, \phi\Big(\log\Big[\Big(\tfrac{t_1}{\eta_1}\Big)^{1/\sigma_1}\Big]\Big)\, \frac{\Phi\Big(\frac{1}{\sqrt{1-\rho^2}}\,\log\Big[\Big(\frac{B}{\eta_2}\Big)^{1/\sigma_2}\Big] - \frac{\rho}{\sqrt{1-\rho^2}}\,\log\Big[\Big(\frac{t_1}{\eta_1}\Big)^{1/\sigma_1}\Big]\Big)}{\Phi\Big(\log\Big[\Big(\frac{B}{\eta_2}\Big)^{1/\sigma_2}\Big]\Big)},$$

where we are adopting the notation $\Phi(C) = \int_C \phi(x)\mathrm{d}x$, for $\phi(x) = g_c(x^2)/\sqrt{2\pi}$.

Proof. It is well-known that the bivariate normal distribution admits a stochastic representation like as in (1.4), where $Z_1 \sim N(0,1)$ and $Z_2 \sim N(0,1)$ are independent. Consequently, $Z_2|Z_1 = z \sim N(0,1)$ and $\rho Z_1 + \sqrt{1-\rho^2} Z_2 \sim N(0,1)$. Further, a simple algebraic manipulation shows that

$$\int_{B_\rho} f_{Z_2}(w \,|\, Z_1 = \widetilde{t}_1)\,\mathrm{d}w = \Phi(B_\rho),$$

where B_ρ is the Borelian set defined in (1.7). Then, by applying Theorem 1.1.5, the proof follows. □

Corollary 1.1.7 (Student-t generator)**.** *Let* $\boldsymbol{T} = (T_1, T_2) \sim \mathrm{BLS}(\boldsymbol{\theta}, g_c)$ *and* $g_c(x) = (1+(x/\nu))^{-(\nu+2)/2}$, $\nu > 0$, *be the generator of the bivariate log-Student-t distribution with* ν *degrees of freedom. Then, for each Borelian subset* B *of* $(0,\infty)$, *the PDF of* $T_1|T_2 \in B$ *is given by (for* $t_1 > 0$*)*

$$f_{T_1}(t_1|T_2 \in B) = \tfrac{1}{t_1\sigma_1}\, f_\nu\Big(\log\Big[\Big(\tfrac{t_1}{\eta_1}\Big)^{1/\sigma_1}\Big]\Big)\, \frac{F_{\nu+1}\Big(\sqrt{\frac{\nu+1}{\nu+\widetilde{t}_1^{\,2}}}\,\Big\{\frac{1}{\sqrt{1-\rho^2}}\,\log\Big[\Big(\frac{B}{\eta_2}\Big)^{1/\sigma_2}\Big] - \frac{\rho}{\sqrt{1-\rho^2}}\,\log\Big[\Big(\frac{t_1}{\eta_1}\Big)^{1/\sigma_1}\Big]\Big\}\Big)}{F_\nu\Big(\log\Big[\Big(\frac{B}{\eta_2}\Big)^{1/\sigma_2}\Big]\Big)},$$

where we are adopting the notation $F_\nu(C) = \int_C f_\nu(x)\mathrm{d}x$, *for*

$$f_\nu(x) = \frac{\Gamma(\frac{\nu+1}{2})}{\sqrt{\nu\pi}\,\Gamma(\frac{\nu}{2})}\, g_c(x^2).$$

Proof. It is well-known that the bivariate Student-t distribution has a stochastic representation like as in (1.4) (see Balakrishnan and Lai 2009), where $Z_1 = Z_1^*\sqrt{\nu/Q} \sim t_\nu$ and $Z_2 = Z_2^*\sqrt{\nu/Q} \sim t_\nu$, $Q \sim \chi^2_\nu$ (chi-square with ν degrees of freedom) is independent of Z_1^* and $\rho Z_1^* + \sqrt{1-\rho^2}Z_2^*$; whereas Z_1^* and Z_2^* are independent and identically distributed standard normal random variables.

Since, $\rho Z_1^* + \sqrt{1-\rho^2}Z_2^* \sim N(0,1)$, we have

$$\rho Z_1 + \sqrt{1-\rho^2}Z_2 = (\rho Z_1^* + \sqrt{1-\rho^2}Z_2^*)\sqrt{\frac{\nu}{Q}} \sim t_\nu.$$

Then

$$\mathbb{P}\Big(\rho Z_1 + \sqrt{1-\rho^2}Z_2 \in B_0\Big) = F_\nu(B_0).$$

On the other hand, if $\boldsymbol{X} = \begin{pmatrix} X_1 \\ X_2 \end{pmatrix} \sim \mathrm{BES}(\boldsymbol{\theta}_*, g_c)$, by Remark 3.7 of Saulo et al. (2022), we have

$$\sqrt{\frac{\nu+1}{(\nu+r^2)(1-\rho^2)}}\left(\frac{X_2-\mu_2}{\sigma_2}-\rho r\right)\,\Bigg|\,\frac{X_1-\mu_1}{\sigma_1}=r\sim t_{\nu+1}.$$

Equivalently,

$$\begin{aligned}\mathbb{P}(T_{\nu+1}\leqslant x)&=\mathbb{P}\left(\sqrt{\frac{\nu+1}{(\nu+r^2)(1-\rho^2)}}\left(\frac{X_2-\mu_2}{\sigma_2}-\rho r\right)\leqslant x\,\Bigg|\,\frac{X_1-\mu_1}{\sigma_1}=r\right)\\&=\mathbb{P}\left(Z_2\leqslant\sqrt{\frac{\nu+r^2}{\nu+1}}\,x\,\Bigg|\,Z_1=r\right),\quad T_{\nu+1}\sim t_{\nu+1}.\end{aligned}$$

By taking $x=\sqrt{(\nu+1)/(\nu+r^2)}\,w$ with $r=\widetilde{t}_1$, we reach at

$$\mathbb{P}\left(T_{\nu+1}\leqslant\sqrt{\frac{\nu+1}{\nu+\widetilde{t}_1^{\,2}}}\,w\right)=\mathbb{P}(Z_2\leqslant w\,|\,Z_1=\widetilde{t}_1).$$

So, differentiating the above identity with respect to w we have

$$\sqrt{\frac{\nu+1}{\nu+\widetilde{t}_1^{\,2}}}\,f_{\nu+1}\left(\sqrt{\frac{\nu+1}{\nu+\widetilde{t}_1^{\,2}}}\,w\right)=f_{Z_2}(w\,|\,Z_1=\widetilde{t}_1).$$

Hence,

$$\begin{aligned}\int_{B_\rho}f_{Z_2}(w\,|\,Z_1=\widetilde{t}_1)\,\mathrm{d}w&=\sqrt{\frac{\nu+1}{\nu+\widetilde{t}_1^{\,2}}}\int_{B_\rho}f_{\nu+1}\left(\sqrt{\frac{\nu+1}{\nu+\widetilde{t}_1^{\,2}}}\,w\right)\mathrm{d}w\\&=F_{\nu+1}\left(\sqrt{\frac{\nu+1}{\nu+\widetilde{t}_1^{\,2}}}\,B_\rho\right).\end{aligned}$$

Finally, by applying Theorem 1.1.5, the proof follows. □

1.1.5 The squared Mahalanobis Distance

The squared Mahalanobis distance of a random vector $\boldsymbol{T} = (T_1, T_2)$ and the vector $\log(\boldsymbol{\eta}) = (\log(\eta_1), \log(\eta_2))$ of a bivariate log-symmetric distribution is defined as

$$d^2(\boldsymbol{T}, \log(\boldsymbol{\eta})) = \frac{\widetilde{T_1}^2 - 2\rho\widetilde{T_1}\widetilde{T_2} + \widetilde{T_2}^2}{1-\rho^2}, \quad \widetilde{T_i} = \log\left[\left(\frac{T_i}{\eta_i}\right)^{1/\sigma_i}\right], \ \eta_i = \exp(\mu_i), \ i = 1, 2.$$

In what follows we derive formulas for the CDF and PDF of the random variable $d^2(\boldsymbol{T}, \log(\boldsymbol{\eta}))$.

Proposition 1.1.8. *If* $\boldsymbol{T} \sim \mathrm{BLS}(\boldsymbol{\theta}, g_c)$ *then the CDF of* $d^2(\boldsymbol{T}, \log(\boldsymbol{\eta}))$*, denoted by* $F_{d^2(\boldsymbol{T}, \log(\boldsymbol{\eta}))}$*, is expressed as*

$$F_{d^2(\boldsymbol{T}, \log(\boldsymbol{\eta}))}(x) = 4\int_0^{\sqrt{x}} \left[F_{Z_2}\left(\sqrt{x - z_1^2} \,\middle|\, Z_1 = z_1\right) - \frac{1}{2}\right] f_{Z_1}(z_1)\,\mathrm{d}z_1 \cdot \mathbb{1}_{(0,\infty)}(x) \tag{1.8}$$

$$= \frac{4}{Z_{g_c}} \int_0^{\sqrt{x}} \left[\int_0^{\sqrt{x - z_1^2}} g_c(z_1^2 + z_2^2)\,\mathrm{d}z_2\right] \mathrm{d}z_1 \cdot \mathbb{1}_{(0,\infty)}(x), \tag{1.9}$$

where Z_{g_c} *is as in Proposition* 1.1.1.

Proof. Since $\boldsymbol{T} = (T_1, T_2)$ admits the stochastic representation given in Proposition 1.1.2, there are Z_1 and Z_2 so that $\widetilde{T_1} = Z_1$ and $\widetilde{T_2} = \rho Z_1 + \sqrt{1-\rho^2} Z_2$. Then, a simple algebraic manipulation shows that

$$d^2(\boldsymbol{T}, \log(\boldsymbol{\eta})) = Z_1^2 + Z_2^2. \tag{1.10}$$

Hence, by using the law of total expectation we have (for $x > 0$)

$$
\begin{aligned}
F_{d^2(\boldsymbol{T},\log(\boldsymbol{\eta}))}(x) &= \mathbb{E}[\mathbb{E}(\mathbb{1}_{\{Z_1^2+Z_2^2\leqslant x\}} \,|\, Z_1)] \\
&= \mathbb{E}\Big[\mathbb{E}\Big(\mathbb{1}_{\{|Z_2|\leqslant\sqrt{x-Z_1^2}\}} \,\Big|\, Z_1 \mathbb{1}_{\{|Z_1|\leqslant\sqrt{x}\}}\Big)\Big] \\
&= \int_{-\sqrt{x}}^{\sqrt{x}} \left[\int_{-\sqrt{x-z_1^2}}^{\sqrt{x-z_1^2}} f_{Z_2}(z_2 \,|\, Z_1 = z_1)\,\mathrm{d}z_2\right] f_{Z_1}(z_1)\,\mathrm{d}z_1.
\end{aligned} \tag{1.11}
$$

Since $f_{Z_2}(z_2 \,|\, Z_1 = z_1)$ and $f_{Z_1}(z_1)$ are even functions (see Proposition 1.1.3), from (1.11) the proof of the first equality (1.8) follows. The second equality (1.9) follows by using in (1.11) the joint PDF f_{Z_1,Z_2} given in Proposition 1.1.3.

As $f_{Z_2}(z_2 \,|\, Z_1 = z_1)$ and $f_{Z_1}(z_1)$ are both even functions (see Proposition 1.1.3), the proof of the first equality in (1.8) follows from (1.11). The second equality in (1.9) follows by using the joint PDF f_{Z_1,Z_2} given in Proposition 1.1.3 in (1.11). □

Proposition 1.1.9. *If* $\boldsymbol{T} \sim \mathrm{BLS}(\boldsymbol{\theta}, g_c)$ *then the PDF of* $d^2(\boldsymbol{T}, \log(\boldsymbol{\eta}))$*, denoted by* $f_{d^2(\boldsymbol{T},\log(\boldsymbol{\eta}))}$*, is written as*

$$
f_{d^2(\boldsymbol{T},\log(\boldsymbol{\eta}))}(x) = \frac{\pi}{Z_{g_c}}\, g_c(x), \quad x > 0,
$$

where Z_{g_c} *is as in Proposition* 1.1.1.

Proof. The proof is immediate, it follows by differentiating (1.9) with respect to x and then by using the following known formula (Leibniz integral rule):

$$
\frac{\mathrm{d}}{\mathrm{d}x}\int_{a(x)}^{b(x)} h(x,y)\,\mathrm{d}y = h(x,b(x))b'(x) - h(x,a(x))a'(x) + \int_{a(x)}^{b(x)} \frac{\partial h(x,y)}{\partial x}\,\mathrm{d}y.
$$

□

Remark 1.1.10. • *Gaussian generator. By taking* $g_c(x) = \exp(-x/2)$ *and* Z_{g_c} *as in*

Table 1.1, *and by applying Proposition* 1.1.9, *we get*

$$f_{d^2(\boldsymbol{T},\boldsymbol{\eta})}(x) = \frac{1}{2}\exp\left(-\frac{x}{2}\right) = \frac{1}{2^{k/2}\Gamma(k/2)}\, x^{(k/2)-1}\exp\left(-\frac{x}{2}\right), \quad \text{with } k = 2.$$

But the formula on the right is the PDF of a random variable following the chi-squared distribution with k degrees of freedom (χ^2_k). Hence, $d^2(\boldsymbol{T}, \log(\boldsymbol{\eta})) \sim \chi^2_2$.

- *Student-t generator. By taking $g_c(x) = (1 + (x/\nu))^{-(\nu+2)/2}$ and Z_{g_c} as in Table* 1.1, *and by applying Proposition* 1.1.9, *we have*

$$\begin{aligned} f_{d^2(\boldsymbol{T},\boldsymbol{\eta})}(x) &= \frac{\Gamma((\nu+2)/2)}{\Gamma(\nu/2)\nu}\left(1 + \frac{x}{\nu}\right)^{-(\nu+2)/2} \\ &= \frac{1}{2}\,\frac{\sqrt{\frac{[d_1(x/2)]^{d_1} d_2^{d_2}}{[d_1(x/2)+d_2]^{d_1+d_2}}}}{(x/2)\mathrm{B}(d_1/2, d_2/2)}, \quad \text{with } d_1 = 2 \text{ and } d_2 = \nu. \end{aligned}$$

Here, $\mathrm{B}(x, y) = \Gamma(x)\Gamma(y)/\Gamma(x + y)$, $x > 0, y > 0$, is the beta function. Notice that the formula on the second identity above is the PDF of a random variable $2X$, where X follows the F-distribution with d_1 and d_2 degrees of freedom (F_{d_1,d_2}). Hence, for abuse of language, we write $d^2(\boldsymbol{T}, \log(\boldsymbol{\eta})) \sim 2F_{2,\nu}$.

1.1.6 Independence

Proposition 1.1.11. *Let $\boldsymbol{T} \sim \mathrm{BLS}(\boldsymbol{\theta}, g_c)$. If $\rho = 0$ and the density generator g_c in* (1.1) *satisfies*

$$g_c(x^2 + y^2) = g_{c_1}(x^2)g_{c_2}(y^2), \quad \forall (x, y) \in \mathbb{R}^2, \tag{1.12}$$

for some density generators g_{c_1} and g_{c_2}, then T_1 and T_2 are statistically independent.

Proof. When $\rho = 0$, by (1.12) the joint density of (T_1, T_2) satisfies

$$f_{T_1,T_2}(t_1, t_2; \boldsymbol{\theta}) = f_1(t_1; \mu_1, \sigma_1) f_2(t_2; \mu_2, \sigma_2), \quad \forall (t_1, t_2) \in (0, \infty) \times (0, \infty),$$

and consequently $Z_{g_c} = Z_{g_{c_1}} Z_{g_{c_2}}$, where

$$f_i(t_i; \mu_i, \sigma_i) = \frac{1}{t_i \sigma_i Z_{g_{c_i}}} g_{c_i}(\widetilde{t_i}^2), \ t_i > 0, \quad \text{and} \quad Z_{g_{c_i}} = \int_{-\infty}^{\infty} g_{c_i}({z_i}^2) \, \mathrm{d}z_i, \quad i = 1, 2,$$

and $\widetilde{t_i}$ as in (1.1). A simple calculation shows that f_1 and f_2 are densities functions (in fact, f_1 and f_2 are densities associated to two univariate continuous and symmetric random variables; see Vanegas and Paula (2016). Then, from Proposition 2.5 of James (2004) it follows that T_1 and T_2 are independent, and even more, that $f_i = f_{T_i}$, for $i = 1, 2$. □

Remark 1.1.12. *Notice that, in Table* 1.1*, the density generator of the bivariate log-normal is the unique one that satisfies the condition* (1.12)*.*

1.1.7 Real moments

Proposition 1.1.13. *Let* $\boldsymbol{X} = \begin{pmatrix} X_1 \\ X_2 \end{pmatrix} \sim \mathrm{BES}(\boldsymbol{\theta}_*, g_c)$ *and* $\boldsymbol{T} = \begin{pmatrix} T_1 \\ T_2 \end{pmatrix} \sim \mathrm{BLS}(\boldsymbol{\theta}, g_c)$. *If the moment-generating function (MGF) of X_i, denoted by $M_{X_i}(s_i)$, $i = 1, 2$, exists, then the real moments of T_i are*

$$\mathbb{E}(T_i^r) = \eta_i^r \vartheta(\sigma_i^2 r^2), \quad \text{with } \eta_i = \exp(\mu_i), \ i = 1, 2, \ r \in \mathbb{R},$$

for some scalar function ϑ, which is called the characteristic generator (see p. 32 in Fang et al. 1990).

For example, when $g_c(x) = \exp(-x/2)$ (Gaussian generator), $\vartheta(x) = \exp(x/2)$, and when $g_c(x) = (1 + (x/\nu))^{-(\nu+2)/2}$, $\nu > 0$ (Student-t generator), ϑ don't exists.

Proof. We only show the case $i = 1$, because the other one follows an analogous reasoning.

Indeed, since the random variable X_1 has a MGF $M_{X_1}(s_1)$, the domain of the characteristic function $\varphi_{X_1}(t)$ can be extended to the complex plane, and

$$M_{X_1}(s_1) = \varphi_{X_1}(-is_1).$$

Since $X_1 = \log(T_1)$, by using the above identity we get

$$\begin{aligned}\mathbb{E}(T_1^r) = \mathbb{E}[\exp(rX_1)] = M_{X_1}(r) &= \exp(r\mu_1)M_{S_1}(\sigma_1 r)\\ &= \exp(r\mu_1)\varphi_{S_1}(-i\sigma_1 r)\\ &= \exp(r\mu_1)\varphi_{S_1,S_2}(-i\sigma_1 r, 0), \qquad (1.13)\end{aligned}$$

with $\begin{pmatrix} S_1 \\ S_2 \end{pmatrix} \sim \text{BES}(\boldsymbol{\theta}_{*0}, g_c)$, $\boldsymbol{\theta}_{*0} = (0, 0, 1, 1, \rho)$; and $\varphi_{S_1,S_2}(s_1, 0)$ is the marginal characteristic function. On the other hand, the characteristic function of the BES distribution is given by (see Item 13.10, p. 595 in Balakrishnan and Lai 2009)

$$\varphi_{S_1,S_2}(s_1, s_2) = \vartheta(s_1^2 + 2\rho s_1 s_2 + s_2^2), \qquad (1.14)$$

where ϑ is the characteristic generator specified in the statement of the proposition. Finally, by using (1.14) in the right-hand side of (1.13), the proof follows. □

1.1.8 Correlation function

By using the stochastic representation (Proposition 1.1.2) of $\boldsymbol{T} = \begin{pmatrix} T_1 \\ T_2 \end{pmatrix}$ and the law of total expectation, we have

$$\begin{aligned}\mathbb{E}(T_1T_2) &= \mathbb{E}(\mathbb{E}(T_1T_2 \,|\, T_1)) = \mathbb{E}(T_1\mathbb{E}(T_2 \,|\, T_1)) \\ &= \eta_1\eta_2\mathbb{E}\Big[\exp\Big((\sigma_1+\sigma_2\rho)Z_1\Big)\,\mathbb{E}\Big(\exp\Big(\sigma_2\sqrt{1-\rho^2}Z_2\Big)\,\Big|\,Z_1\Big)\Big],\end{aligned}$$

where Z_1 and Z_2 are defined in Proposition 1.1.2.

Hence, from formula of moments (Proposition 1.1.13) we get the next formula for the correlation function of T_1 and T_2:

$$\rho(T_1,T_2) = \frac{\mathbb{E}\Big[\exp\Big((\sigma_1+\sigma_2\rho)Z_1\Big)\,\mathbb{E}\Big(\exp\Big(\sigma_2\sqrt{1-\rho^2}Z_2\Big)\,\Big|\,Z_1\Big)\Big] - \vartheta(\sigma_1^2)\vartheta(\sigma_2^2)}{\sqrt{\vartheta(4\sigma_1^2)-\vartheta^2(\sigma_1^2)}\,\sqrt{\vartheta(4\sigma_2^2)-\vartheta^2(\sigma_2^2)}},$$

where ϑ is a scalar function stated in Proposition 1.1.13.

It is a simple task to verify that, when $g_c(x) = \exp(-x/2)$ (Gaussian generator), $\rho(T_1,T_2) = [\exp(\sigma_1\sigma_2\rho) - 1]/\Big[\sqrt{\exp(\sigma_1^2)-1}\,\sqrt{\exp(\sigma_2^2)-1}\Big]$, and when $g_c(x) = (1+(x/\nu))^{-(\nu+2)/2}$, $\nu>0$ (Student-t generator), $\rho(T_1,T_2)$ does not exist.

1.1.9 Other properties

If $\boldsymbol{T} = \begin{pmatrix} \mathrm{T}_1 \\ \mathrm{T}_2 \end{pmatrix} \sim \mathrm{BLS}(\boldsymbol{\theta}, g_c)$ then, in analogy to stated by Vanegas and Paula (2016), the following properties follow immediately as a consequence of the definition of the BLS distribution:

(P1) The CDF of $\boldsymbol{T}$ is written as $F_{T_1,T_2}(t_1,t_2;\boldsymbol{\theta}) = F_{S_1,S_2}\big(\widetilde{t}_1,\widetilde{t}_2;\boldsymbol{\theta}_{*0}\big)$, with $\begin{pmatrix} \mathrm{S}_1 \\ \mathrm{S}_2 \end{pmatrix} \sim \mathrm{BES}(\boldsymbol{\theta}_{*0}, g_c)$ and $\boldsymbol{\theta}_{*0} = (0,0,1,1,\rho)$.

(P2) The random vector $\begin{pmatrix} \mathrm{T}_1^* \\ \mathrm{T}_2^* \end{pmatrix} = ([T_1/\eta_1]^{1/\sigma_1}, [T_2/\eta_2]^{1/\sigma_2})$ follows standard BLS distribution. In other words, $\begin{pmatrix} \mathrm{T}_1^* \\ \mathrm{T}_2^* \end{pmatrix} \sim \mathrm{BLS}(\boldsymbol{\theta}_0, g_c)$ with $\boldsymbol{\theta}_0 = (1, 1, 1, 1, \rho)$.

(P3) $\begin{pmatrix} \mathrm{c}_1 T_1 \\ \mathrm{c}_2 T_2 \end{pmatrix} \sim \mathrm{BLS}(c_1\eta_1, c_2\eta_2, \sigma_1, \sigma_2, g_c)$ for all constants $c_1, c_2 > 0$.

(P4) $\begin{pmatrix} \mathrm{T}_1^{c_1} \\ \mathrm{T}_2^{c_2} \end{pmatrix} \sim \mathrm{BLS}(\eta_1^{c_1}, \eta_2^{c_2}, c_1^2\sigma_1, c_2^2\sigma_2, g_c)$ for all constants $c_1 \neq 0$ and $c_2 \neq 0$.

Proposition 1.1.14. *If* $\boldsymbol{T} = \begin{pmatrix} T_1 \\ T_2 \end{pmatrix} \sim \mathrm{BLS}(\boldsymbol{\theta}, g_c)$ *then the random vectors* $(\eta_1/T_1, \eta_2/T_2)$ *and* $(T_1/\eta_1, T_2/\eta_2)$ *are identically distributed.*

Furthemore, $(\eta_1/T_1, \eta_2/T_2) \sim \mathrm{BLS}(\boldsymbol{\theta}_\bullet, g_c)$ *and* $(T_1/\eta_1, T_2/\eta_2) \sim \mathrm{BLS}(\boldsymbol{\theta}_\bullet, g_c)$*, with* $\boldsymbol{\theta}_\bullet = (1, 1, \sigma_1, \sigma_2, \rho)$.

Proof. By using the well-known identity for two random variables X and Y (see e.g. p. 59 of James 2004): for all $a_1 < b_1$ and $a_2 < b_2$,

$$\mathbb{P}(a_1 < X \leqslant b_1, a_2 < Y \leqslant b_2) = F_{X,Y}(b_1, b_2) - F_{X,Y}(b_1, a_2) - F_{X,Y}(a_1, b_2) + F_{X,Y}(a_1, a_2);$$

with $a_1 = \eta_1/w_1$, $b_1 = \infty$, $a_2 = \eta_2/w_2$ and $b_2 = \infty$, for all $(w_1, w_2) \in (0, \infty)^2$, we get

$$\mathbb{P}\Big(\frac{\eta_1}{T_1} \leqslant w_1, \frac{\eta_2}{T_2} \leqslant w_2\Big) = 1 - F_{T_2}\Big(\frac{\eta_2}{w_2}\Big) - F_{T_1}\Big(\frac{\eta_1}{w_1}\Big) + F_{T_1,T_2}\Big(\frac{\eta_1}{w_1}, \frac{\eta_2}{w_2}\Big). \tag{1.15}$$

Since

$$\mathbb{P}\Big(\frac{T_1}{\eta_1} \leqslant w_1, \frac{T_2}{\eta_2} \leqslant w_2\Big) = F_{T_1,T_2}(\eta_1 w_1, \eta_2 w_2), \tag{1.16}$$

by taking partial derivatives with respect to w_1 and w_2 in (1.15) and (1.16), we have that the joint PDF of η_1/T_1 and η_2/T_2, and the joint PDF of T_1/η_1 and T_2/η_2, are related as follows

$$f_{\frac{\eta_1}{T_1},\frac{\eta_2}{T_2}}(w_1,w_2) = f_{\frac{T_1}{\eta_1},\frac{T_2}{\eta_2}}(w_1,w_2) = \frac{1}{w_1 w_2 \sigma_1 \sigma_2 \sqrt{1-\rho^2} Z_{g_c}}\, g_c\left(\frac{\widetilde{w}_1^2 - 2\rho\widetilde{w}_1\widetilde{w}_2 + \widetilde{w}_2^2}{1-\rho^2}\right),$$

with $w_i > 0$ and $\widetilde{w_i} = \log({w_i}^{1/\sigma_i})$, $i = 1, 2$. This completes the proof of proposition. $\square$

Chapter 2

Maximum likelihood estimation

Let $\{(T_{1i}, T_{2i}) : i = 1, \dots, n\}$ be a bivariate random sample of size n from the BLS$(\boldsymbol{\theta}, g_c)$ distribution with PDF as given in (1.1), and let (t_{1i}, t_{2i}) be the correspondent observations of (T_{1i}, T_{2i}). Then, the log-likelihood function for $\boldsymbol{\theta} = (\eta_1, \eta_2, \sigma_1, \sigma_2, \rho)$, without the additive constant, is expressed as

$$\ell(\boldsymbol{\theta}) = -n\log(\sigma_1\sigma_2) - \frac{n}{2}\log\left(1-\rho^2\right) + \sum_{i=1}^{n}\log g_c\left(\frac{\widetilde{t_{1i}}^2 - 2\rho\widetilde{t}_{1i}\widetilde{t}_{2i} + \widetilde{t_{2i}}^2}{1-\rho^2}\right), \quad t_{1i}, t_{2i} > 0,$$

$$\widetilde{t}_{ki} = \log\left[\left(\frac{t_{ki}}{\eta_k}\right)^{1/\sigma_k}\right],\ \eta_k = \exp(\mu_k),\ k = 1, 2;\ i = 1, \dots, n.$$

In the case that a supremum $\widehat{\boldsymbol{\theta}} = (\widehat{\eta_1}, \widehat{\eta_2}, \widehat{\sigma_1}, \widehat{\sigma_2}, \widehat{\rho})$ exists, it must satisfy the following likelihood equations:

$$\left.\frac{\partial\ell(\boldsymbol{\theta})}{\partial\eta_1}\right|_{\boldsymbol{\theta}=\widehat{\boldsymbol{\theta}}} = 0, \quad \frac{\partial\ell(\boldsymbol{\theta})}{\partial\eta_2} = 0, \quad \left.\frac{\partial\ell(\boldsymbol{\theta})}{\partial\sigma_1}\right|_{\boldsymbol{\theta}=\widehat{\boldsymbol{\theta}}} = 0, \quad \left.\frac{\partial\ell(\boldsymbol{\theta})}{\partial\sigma_2}\right|_{\boldsymbol{\theta}=\widehat{\boldsymbol{\theta}}} = 0, \quad \left.\frac{\partial\ell(\boldsymbol{\theta})}{\partial\rho}\right|_{\boldsymbol{\theta}=\widehat{\boldsymbol{\theta}}} = 0, \quad (2.1)$$

with

$$\frac{\partial \ell(\boldsymbol{\theta})}{\partial \eta_1} = \frac{2}{\sigma_1 \eta_1 (1-\rho^2)} \sum_{i=1}^{n} \Big(\rho \widetilde{t}_{2i} - \widetilde{t}_{1i}\Big) G(\widetilde{t}_{1i}, \widetilde{t}_{2i}),$$

$$\frac{\partial \ell(\boldsymbol{\theta})}{\partial \eta_2} = \frac{2}{\sigma_2 \eta_2 (1-\rho^2)} \sum_{i=1}^{n} \Big(\rho \widetilde{t}_{1i} - \widetilde{t}_{2i}\Big) G(\widetilde{t}_{1i}, \widetilde{t}_{2i}),$$

$$\frac{\partial \ell(\boldsymbol{\theta})}{\partial \sigma_1} = -\frac{n}{\sigma_1} + \frac{2}{\sigma_1 (1-\rho^2)} \sum_{i=1}^{n} \widetilde{t}_{1i} \Big(\rho \widetilde{t}_{2i} - \widetilde{t}_{1i}\Big) G(\widetilde{t}_{1i}, \widetilde{t}_{2i}),$$

$$\frac{\partial \ell(\boldsymbol{\theta})}{\partial \sigma_2} = -\frac{n}{\sigma_2} + \frac{2}{\sigma_2 (1-\rho^2)} \sum_{i=1}^{n} \widetilde{t}_{2i} \Big(\rho \widetilde{t}_{1i} - \widetilde{t}_{2i}\Big) G(\widetilde{t}_{1i}, \widetilde{t}_{2i}),$$

$$\frac{\partial \ell(\boldsymbol{\theta})}{\partial \rho} = \frac{n\rho}{1-\rho^2} - \frac{2}{(1-\rho^2)^2} \sum_{i=1}^{n} \Big(\rho \widetilde{t}_{1i} - \widetilde{t}_{2i}\Big)\Big(\rho \widetilde{t}_{2i} - \widetilde{t}_{1i}\Big) G(\widetilde{t}_{1i}, \widetilde{t}_{2i}), \tag{2.2}$$

where we are denoting

$$G(\widetilde{t}_{1i}, \widetilde{t}_{2i}) = g_c' \left(\frac{\widetilde{t_{1i}}^{2} - 2\rho \widetilde{t}_{1i} \widetilde{t}_{2i} + \widetilde{t_{2i}}^{2}}{1-\rho^2} \right) \Big/ g_c \left(\frac{\widetilde{t_{1i}}^{2} - 2\rho \widetilde{t}_{1i} \widetilde{t}_{2i} + \widetilde{t_{2i}}^{2}}{1-\rho^2} \right), \quad i = 1, \ldots, n.$$

A simple observation shows that the likelihood equations (2.1) can be written as follows

$$\sum_{i=1}^{n} \widetilde{t}_{1i}\, G(\widetilde{t}_{1i}, \widetilde{t}_{2i}) \Big|_{\boldsymbol{\theta} = \widehat{\boldsymbol{\theta}}} = 0,$$

$$\sum_{i=1}^{n} \Big(\widetilde{t}^{2}_{1i} - \widetilde{t}^{2}_{2i}\Big)\, G(\widetilde{t}_{1i}, \widetilde{t}_{2i}) \Big|_{\boldsymbol{\theta} = \widehat{\boldsymbol{\theta}}} = 0,$$

$$\sum_{i=1}^{n} \widetilde{t_{2i}} \Big[2\rho \widetilde{t_{2i}} - (1+\rho^2) \widetilde{t}_{1i}\Big]\, G(\widetilde{t}_{1i}, \widetilde{t}_{2i}) \Big|_{\boldsymbol{\theta} = \widehat{\boldsymbol{\theta}}} = -\frac{n\widehat{\rho}(1-\widehat{\rho}^2)}{2}\,.$$

Any nontrivial root $\widehat{\boldsymbol{\theta}}$ of the above likelihood equations is known as an ML estimate in the loose sense. When the parameter value provides the absolute maximum of the log-likelihood function, it is called an ML estimate in the strict sense.

In the following proposition we study the existence of the ML estimate $\widehat{\rho}$ when the other parameters are known.

Proposition 2.0.1. *Let g_c be a density generator satisfying the following condition:*

$$g_c'(x) = r(x)g_c(x), \quad -\infty < x < \infty, \tag{2.3}$$

for some real-valued function $r(x)$ so that $\lim_{\rho\to\pm 1} r(x_{\rho,i}) = c \in (-\infty, 0)$, where $x_{\rho,i} = (\widetilde{t}_{1i}^2 - 2\rho\widetilde{t}_{1i}\widetilde{t}_{2i} + \widetilde{t}_{2i}^2)/(1-\rho^2)$, $i = 1, \ldots, n$. If the parameters η_1, η_2, σ_1 and σ_2 are known, then the equation (2.2) has at least one root on the interval $(-1, 1)$.

Proof. Since $g_c'(x_{\rho,i}) = r(x_{\rho,i})g_c(x_{\rho,i})$, we have $G(\widetilde{t}_{1i}, \widetilde{t}_{2i}) = r(x_{\rho,i})$. Then, by using the condition $\lim_{\rho\to\pm 1} r(x_{\rho,i}) = c < 0$, from (2.2) we can easily see that

$$\lim_{\rho\to 1^-} \frac{\partial \ell(\boldsymbol{\theta})}{\partial \rho} = -\infty \quad \text{and} \quad \lim_{\rho\to -1^+} \frac{\partial \ell(\boldsymbol{\theta})}{\partial \rho} = +\infty.$$

So, by intermediate value theorem, there exists at least one solution on the interval $(-1, 1)$. □

Remark 2.0.2. *Notice that, in Table 1.1, the density generators of the Bivariate Log-normal (or Bivariate Log-power-exponential with $\xi = 0$) and the Bivariate Log-Kotz type (with $\delta = 1$) satisfy the hypotheses of Proposition 2.0.1 with $r(x_{\rho,i}) = -1/2$ and $r(x_{\rho,i}) = (-\lambda x_{\rho,i} + \xi - 1)/x_{\rho,i} \longrightarrow -\lambda$ as $\rho \to \pm 1$, $\forall i = 1, \ldots, n$, respectively. Then, Proposition 2.0.1 can be applied to guarantee the existence of an ML estimator $\widehat{\rho}$ of ρ in the loose sense.*

On the other hand, the density generators of the Bivariate Log-Kotz type (with $\delta < 1$), the Bivariate Log-Student-t, the Bivariate Log-Pearson Type VII and the Bivariate Log-power-exponential (with $\xi \neq 0$) satisfy the condition (2.3) with $r(x) = (-\lambda\delta x^{\delta} + \xi - 1)/x$, $r(x) = -(\nu+2)/2(1+\frac{x}{\nu})$, $r(x) = -\xi/(\theta + x)$ and $r(x) = -x^{-\xi/(\xi+1)}/2(\xi+1)$, respectively, but in all these cases $r(x_{\rho,i}) \longrightarrow 0$ as $\rho \to \pm 1$, $\forall i = 1, \ldots, n$.

For the BLS model no closed-form solution to the maximization problem is known or available, and an MLE can only be found via numerical optimization. Under mild regularity conditions (Cox and Hinkley 1974 or Davison 2008), the asymptotic distribution of ML estimator $\widehat{\boldsymbol{\theta}}$ of $\boldsymbol{\theta}$ is easily determined by the convergence in law: $(\widehat{\boldsymbol{\theta}} - \boldsymbol{\theta}) \xrightarrow{\mathscr{D}} N(\mathbf{0}, I^{-1}(\boldsymbol{\theta}))$, where $\mathbf{0}$ is the zero mean vector and $I^{-1}(\boldsymbol{\theta})$ is the inverse expected Fisher information matrix. The main use of the last convergence is to construct confidence regions and to perform hypothesis testing for $\boldsymbol{\theta}$ (Davison 2008).

Chapter 3

Monte Carlo simulation

In this chapter, we carry out a Monte Carlo (MC) simulation study to evaluate the performance of the previously proposed maximum likelihood estimators for the BLS models. We use different sample sizes and parameter settings, using the following distributions: log-slash, log-power-exponential and log-normal.

The simulation scenario considers the following setting: 1,000 MC replications, sample size $n \in (25, 50, 100, 150)$, vector of true parameters $(\eta_1, \eta_2, \sigma_1, \sigma_2) = (1, 1, 0.5, 0.5)$, and $\rho \in \{0, 0.25, 0.5, 0.75, 0.95\}$. The extra parameters of the chosen distributions are assumed to be fixed.

The performance and recovery of the ML estimators were evaluated through the empirical bias and the mean square error (MSE), which are calculated from the MC replicates, as shown below,

$$\text{Bias}(\widehat{\theta}) = \frac{1}{N}\sum_{i=1}^{N} \widehat{\theta}^{(i)} - \theta, \qquad \text{MSE}(\widehat{\theta}) = \frac{1}{N}\sum_{i=1}^{N} (\widehat{\theta}^{(i)} - \theta)^2, \tag{3.1}$$

where θ and $\widehat{\theta}^{(i)}$ are the true value of the parameter and its respective ith estimate, and N is the number of MC replications. The steps for the MC simulation study are described in Algorithm 1.

Algorithm 1. Simulation
1. Choose the BLS distribution based on Table 1.1 and define the value of the parameters of the chosen distribution.
2. Generate 1,000 samples of size n based on the chosen model.
3. Estimate the model parameters using the ML method for each sample.
4. Compute the empirical bias and MSE.

The simulation results are shown in Tables 3.1, 3.2 and 3.3. It is possible to observe in the simulations that the results produced for the chosen distributions were as expected. As the sample size increases, the bias and MSE tend to decrease. In general, the results do not seem to depend on the parameter ρ.

Table 3.1: Monte Carlo simulation results for the bivariate log-slash distribuition with $\nu = 4$.

n	ρ	MLE									
		$\widehat{\eta}_1$		$\widehat{\eta}_2$		$\widehat{\sigma}_1$		$\widehat{\sigma}_2$		$\widehat{\rho}$	
		Bias	MSE	Bias	MSE	Bias	MSE	Bias	MSE	Bias	MSE
25	0.0	0.0093	0.0138	0.0085	0.0145	-0.1040	0.0154	-0.1004	0.0143	-0.0053	0.0426
	0.25	0.0078	0.0137	0.0096	0.0150	-0.1053	0.0148	-0.0988	0.0140	-0.0079	0.0412
	0.50	0.0050	0.0126	0.0060	0.0133	-0.1008	0.0142	-0.1046	0.0149	-0.0145	0.0276
	0.95	0.0072	0.0150	0.0064	0.0148	-0.1055	0.0151	-0.1048	0.0150	-0.0046	0.0007
50	0.0	0.0056	0.0067	0.0030	0.0071	-0.0967	0.0113	-0.0947	0,0110	-0.0065	0.0230
	0.25	0.0053	0.0069	0.0017	0.0072	-0.0963	0.0114	-0.0968	0.0114	-0.0071	0.0230
	0.50	0.0033	0.0065	0.0041	0.0073	-0.0959	0.0112	-0.0951	0.0111	-0.0013	0.0130
	0.95	0.0060	0.0071	0.0066	0.0070	-0.0944	0.0109	-0.0938	0.0108	-0.0003	0.0002
100	0.0	0.0006	0.0037	0.0023	0.0034	-0.0955	0.0101	-0.0944	0.0098	0.0015	0.0122
	0.25	-0.0016	0.0033	-0.0007	0.0035	-0.0929	0.0096	-0.0943	0.0098	-0.0056	0.0095
	0.50	0.0003	0.0035	0.0023	0.0034	-0.0949	0.0100	-0.0962	0.0102	-0.0047	0.0066
	0.95	0.0009	0.0036	0.0015	0.0035	-0.0955	0.0101	-0.0960	0.0102	-0.0012	0.0001
150	0.0	0.0034	0.0022	0.0033	0.0024	-0.0953	0.0098	-0.0941	0.0095	0.0050	0.0075
	0.25	0.0000	0.0023	-0.0004	0.0022	-0.0933	0.0093	-0.0934	0.0094	-0.0045	0.0066
	0.50	0.0027	0.0022	0.0038	0.0023	-0.0942	0.0095	-0.0932	0.0094	-0.0015	0.0039
	0.95	0.0007	0.0023	0.0011	0.0023	-0.0936	0,0094	-0.0937	0.0094	-0.0006	0.0001

Table 3.2: Monte Carlo simulation results for the bivariate log-power-exponential distribuition with $\xi = 0.3$.

n	ρ	MLE									
		$\widehat{\eta}_1$		$\widehat{\eta}_2$		$\widehat{\sigma}_1$		$\widehat{\sigma}_2$		$\widehat{\rho}$	
		Bias	MSE	Bias	MSE	Bias	MSE	Bias	MSE	Bias	MSE
25	0.00	0.0076	0.0174	0.0133	0.0173	-0.0466	0.0138	-0.0488	0.0139	-0.0038	0.0411
	0.25	0.0048	0.0155	0.0076	0.0169	-0.0434	0.0079	-0.0432	0.0082	-0.0098	0.0399
	0.50	0.0066	0.0170	0.0099	0.0192	-0.0436	0.0097	-0.0413	0.0101	-0.0047	0.0268
	0.95	0.0108	0.0170	0.0113	0.0171	-0.0423	0.0095	-0.0431	0.0097	-0.0025	0.0006
50	0.00	0.0026	0.0083	0.0060	0.0084	-0.0735	0.0415	-0.0712	0.0418	0.0016	0.0192
	0.25	0.0062	0.0090	0.0017	0.0081	-0.0572	0.0243	-0.0597	0.0258	0.0003	0.0181
	0.50	0.0055	0.0085	0.0027	0.0076	-0.0452	0.0128	-0.0460	0.0125	-0.0026	0.0134
	0.95	0.0062	0.0086	0.0068	0.0088	-0.0364	0.0047	-0.0358	0.0047	-0.0010	0.0003
100	0.00	0.0034	0.0044	0.0022	0.0043	-0.0475	0.0173	-0.0453	0.0175	0.0011	0.0107
	0.25	0.0023	0.0041	0.0025	0.0042	-0.0331	0.0044	-0.0319	0.0044	0.0011	0.0088
	0.50	0.0011	0.0045	0.0000	0.0041	-0.0313	0.0035	-0.0345	0.0035	-0.0058	0.0063
	0.95	0.0034	0.0043	0.0027	0.0043	-0.0344	0.0041	-0.0344	0.0040	-0.0009	0.0001
150	0.00	-0.0002	0.0029	0.0038	0.0029	-0.0319	0.0020	-0.0318	0.0020	-0.0008	0.0066
	0.25	0.0001	0.0028	0.0036	0.0026	-0.0300	0.0019	-0.0303	0.0019	-0.0018	0.0067
	0.50	0.0050	0.0026	0.0027	0.0027	-0.0331	0.0036	-0.0347	0.0037	-0.0035	0.0041
	0.95	0.0014	0.0029	0.0014	0.0029	-0.0306	0.0019	-0.0299	0.0018	-0.0002	0.0001

Table 3.3: Monte Carlo simulation results for the bivariate log-normal distribuition.

n	ρ	MLE									
		$\widehat{\eta}_1$		$\widehat{\eta}_2$		$\widehat{\sigma}_1$		$\widehat{\sigma}_2$		$\widehat{\rho}$	
		Bias	MSE	Bias	MSE	Bias	MSE	Bias	MSE	Bias	MSE
25	0.00	-0.0007	0.0100	0.0034	0.0100	-0.0312	0.0182	-0.0265	0.0178	-0.0044	0.0411
	0.25	0.0095	0.0107	0.0085	0.0101	-0.0336	0.0180	-0.0320	0.0220	-0.0151	0.0367
	0.50	0.0120	0.0109	0.0159	0.0112	-0.0313	0.0205	-0.0282	0.0198	-0.0106	0.0252
	0.95	-0.0013	0.0101	-0.0011	0.0100	-0.0289	0.0197	-0.0296	0.0199	-0.0026	0.0005
50	0.00	-0.0012	0.0051	0.0055	0.0050	-0.0394	0.0373	-0.0433	0.0387	0.0006	0.0187
	0.25	0.0014	0.0053	0.0023	0.0050	-0.0445	0.0391	-0.0446	0.0387	-0.0014	0.0174
	0.50	0.0023	0.0051	-0.0027	0.0052	-0.0370	0.0329	-0.0363	0.0338	0.0044	0.0114
	0.95	0.0020	0.0051	0.0026	0.0051	-0.0151	0.0127	-0.0148	0.0125	0.0000	0.0002
100	0.00	0.0021	0.0025	0.0020	0.0024	-0.0043	0.0022	-0.0067	0.0021	0.0079	0.0096
	0.25	-0.0025	0.0025	-0.0031	0.0025	-0.0063	0.0042	-0.0073	0.0039	0.0008	0.0091
	0.50	-0.0019	0.0025	-0.0007	0.0023	-0.0035	0.0013	-0.0052	0.0013	0.0001	0.0057
	0.95	0.0025	0.0024	0.0029	0.0024	-0.0061	0.0038	-0.0063	0.0039	-0.0001	0.0001
150	0.00	0.0000	0.0018	-0.0018	0.0016	-0.0101	0.0080	-0.0096	0.0081	-0.0024	0.0064
	0.25	0.0003	0.0017	0.0001	0.0017	-0.0041	0.0027	-0.0035	0.0027	0.0016	0.0058
	0.50	0.0000	0.0016	-0.0001	0.0017	-0.0088	0.0066	-0.0082	0.0064	-0.0021	0.0040
	0.95	0.0013	0.0018	0.0016	0.0018	-0.0115	0.0100	-0.0118	0.0100	-0.0006	0.0001

Chapter 4

Application to real data

In this chapter we will illustrate the proposed methodology and use a real data set to apply the bivariate log-symmetric models. The data is based on the article by Marchant et al. (2015), in which the authors proposed a multivariate Birnbaum-Saunders regression model to describe fatigue data. The authors describes fatigue as the process of material failure, which is caused by cyclic stress. Thus, fatigue is composed of crack initiation and propagation, until the material fractures. The calculation of fatigue life is important for determining the reliability of components or structures. Here, we consider the variables *Von Mises stress* (T_1, in N/mm^2) and *die lifetime* (T_2, in number of cycles). According to Marchant et al. (2015), die fracture is the fatigue of metal caused by cyclic stress in the course of the service life cycle of dies (die lifetime).

Table 4.1 provides descriptive statistics for the variables *Von mises stress* (T_1) and *die lifetime* (T_2), including the minimum, median, mean, maximum, standard deviation (SD), coefficient of variation (CV), coefficient of skewness (CS), and coefficient of kurtosis (CK). We can observe in the variable *Von mises stress*, that the mean and median are respectively 1.247 and 1.130, i.e. the mean is larger than the median, which indicates a possitively skewed feature in the data distribution. The CV is 56.172%, which means a moderate level of dispersion around the mean. Furthermore, the CS value confirms the skewed nature. The variable *die lifetime* has mean equal to 23.761 and median equal to 19.000. These values also indicate the positively skewed feature in the distribution of the data. Moreover, the CV value is 71.967%, which shows us the moderate level of dispersion around the mean. The CS confirms the skewed nature and the CK value indicates the high kurtosis feature in the data distribution.

Table 4.1: Summary statistics for the indicated data set.

Variables	n	Minimum	Median	Mean	Maximum	SD	CV	CS	CK
T_1	15	0.243	1.130	1.247	2.430	0.700	56.172	0.209	-1.466
T_2	15	6.420	19.900	23.761	74.800	17.100	71.967	1.631	2.495

Table 4.2 presents the ML estimates and the standard errors (in parentheses) for the bivariate log-symmetric model parameters. This table also reports the log-likelihood value, and the values of the Akaike (AIC) and Bayesian (BIC) information criteria. The extra parameters were estimated using the profile log-likelihood. From Table 4.2, we observe that the log-Laplace model provides better adjustment than other models based on the values of log-likelihood, AIC and BIC.

Table 4.2: ML estimates (with standard errors in parentheses), log-likehood, AIC and BIC values for the indicated bivariate log-symmetric models.

Distribuiton	$\widehat{\eta}_1$	$\widehat{\eta}_2$	$\widehat{\sigma}_1$	$\widehat{\sigma}_2$	$\widehat{\rho}$	$\widehat{\nu}$	Log-likehood	AIC	BIC
Log-normal	1.0362*	19.4824*	0.6536*	0.6210*	-0.9390*	-	-58.117	126.23	129.78
	(0.0175)	(3.1239)	(0.01192)	(0.1133)	(0.0305)				
Log-Student-t	1.0188*	20.1932*	0.6111*	0.5508*	-0.9514*	7	-57.915	125.83	129.37
	(0.1339)	(2.0685)	(0.2220)	(0.1881)	(0.0207)				
Log-Pearson Type VII	1.0211*	20.1218*	0.3712*	0.3362*	-0.9502*	$\xi = 5$, $\theta = 22$	-57.917	125.83	129.37
	(0.1806)	(3.2095)	(0.0761)	(0.0698)	(0.0280)				
Log-hyperbolic	1.0175*	20.1900*	0.6843*	0.6201*	-0.9504*	2	-57.922	125.84	129.38
	(0.1910)	(3.2818)	(0.0376)	(0.0333)	(0.0276)				
Log-Laplace	**1.0594***	**20.9110***	**0.7748***	**0.6809***	**-0.9471***	**-**	**-57.585**	**125.17**	**128.71**
	(0.0023)	(0.0105)	(0.2032)	(0.1745)	(0.0342)				
Log-slash	1.0207*	20.1854*	0.5158*	0.4648*	-0.9515*	5	-57.945	125.89	129.43
	(0.1783)	(3.1973)	(0.1030)	(0.0955)	(0.0277)				
Log-power-exponential	1.0298*	19.9461*	0.4516*	0.4154*	-0.9432*	0.37	-57.984	125.97	129.51
	(0.18445)	(3.2182)	(0.0935)	(0.0852)	(0.0294)				
Log-logistic	1.0498*	18.8904*	0.7651*	0.7488*	-0.9315*	-	-58.672	127.34	130.89
	(0.1650)	(3.0396)	(0.1212)	(0.1231)	(0.0316)				

* significant at 5% level.

Concluding Remarks

In this paper, we have introduced a class of bivariate log-symmetric models, which is the result of an exponential transformation on a variable that follows a bivariate symmetric distribution. We have studied the main statistical properties, proposed the maximum likelihood estimators for the model parameters. A Monte Carlo simulation study has been carried out to numerically evaluate the maximum likelihood estimators. The simulation results have showed the good performance for the estimators, obtaining empirical bias values close to zero, as shown in Tables 3.1-3.3. We have applied the proposed models to a real fatigue data set. The results are seen to be favorable to the log-Laplace model. As part of future research, it will be of interest to propose bivariate log-symmetric regression models. Furthermore, the study of some hypothesis and misspecification tests via Monte Carlo simulation can be investigated. Work on these problems is currently in progress and we hope to report these findings in future.

References

[1] Abdous, B., Fougères, A.-L., Ghoudi, K., "Extreme behaviour for bivariate elliptical distributions", *Canadian Journal of Statistics*, pp. 317–334, 2005.

[2] Balakrishnan, N., Lai, C-D., *Continuous Bivariate Distributions*, Springer-Verlag, New York, 2009.

[3] Birnbaum, Z. W., Saunders, S. C., "A new family of life distributions", *J. Appl. Probab.* V.6, pp. 319-327, 1969.

[4] Cox, D. R., Hinkley, D. V., *Theoretical statistics*, London, UK: Chapman and Hall, 1974.

[5] Crow, E. L., Shimizu, K., *Lognormal Distributions: Theory and Applications*, New York, US: Dekker, 2008.

[6] Cysneiros, F. J. A., Paula, G. A., Galea, M., *Modelos simétricos aplicados.*, Escola de Modelos de Regressao, V. 9, 2005.

[7] Davison, A. C., *Statistical Models* (First ed.). Cambridge Series in Statistical and Probabilistic 408 Mathematics, Cambridge University Press, 2008.

[8] Diaz-Garcia, J., Leiva, V., "A new family of life distributions based on elliptically contoured" J. Stat. Plan. Inference, V. 128, pp. 445-457, 2005.

[9] Fang, K. T., Kotz, S., Ng, K. W., *Symmetric multivariate and related distributions*, Chapman and Hall, London, 1990.

[10] James, B. R., *Probabilidade: um curso em nível intermediário*, Projeto Euclides, Brazil, 2004.

[11] Johnson, N., Kotz, S., Balakrishnan, N., *Continuous Univariate Distributions*, V. 1, New York, US: Wiley, 1994.

[12] Jones, M. C., “On reciprocal symmetry”, *J. Stat. Plan. Inference*, v. 138, pp. 3039-3043, 2008.

[13] Kotz, S., Kozubowski, T. J., Podgórski, K., *The Laplace Distribution and Generalizations*, 2001.

[14] Marshall, A., Olkin, I., *Life Distributions*. New York, US: Springer, 2007.

[15] Marchant, C., Cysneiros, F. J. A., *A multivariate log-linear model for Birnbaum-Saunders distributions*. IEEE Transactions on Reliability, 65(2), 816-827, 2015.

[16] Puig, P., *A note on the harmonic law: a two-parameter family of distributions for ratios.*, Statistics probability letters, v. 78, n. 3, pp. 320-326, 2008.

[17] Rieck, J., Nedelman, J., “A log-linear model for the Birnbaum-Saunders distribution”, *Technometrics*, V. 3, pp. 51-60, 1991.

[18] Rohatgi, V. K., Saleh, A. K. Md. E., *An introduction to probability theory and mathematical statistics*, John Wiley & Sons, 3rd edition, 2015.

[19] Saulo, H., Balakrishnan, N., Zhu, X., Gonzales, J. F. B., Leão, J., “Estimation in generalized bivariate Birnbaum-Saunders models", *Metrika*, pp. 427-453, 2017.

[20] Saulo, H., Vila, R., Cordeiro, S. S., Leiva, V., “Bivariate symmetric Heckman models and their characterization”, *Journal of Multivariate Analysis*, 105097, 2022.

[21] Vanegas, L. H., Paula, G. A., “Log-symmetric distributions: Statistical properties and parameter estimation", *Brazilian Journal of Probability and Statistics*, pp. 196-220, 2016.