UNIVERSIDADE FEDERAL DE SÃO CARLOS
CENTRO DE CIÊNCIAS EXATAS E TECNOLOGIA
PROGRAMA INTERINSTITUCIONAL DE PÓS-GRADUAÇÃO EM ESTATÍSTICA UFSCar-USP

Giovanni Pastori Piccirilli

Modelos de regressão misto para resposta limitada usando distribuições do tipo Johson-SB

Dissertação apresentada ao Departamento de Estatística – Des/UFSCar e ao Instituto de Ciências Matemáticas e de Computação – ICMC-USP, como parte dos requisitos para obtenção do título de Mestre ou Doutor em Estatística - Programa Interinstitucional de Pós-Graduação em Estatística UFSCar-USP.

Orientador: Prof. Dr. Jorge Luis Bazán Guzmán

São Carlos
Março de 2021

UNIVERSIDADE FEDERAL DE SÃO CARLOS
CENTRO DE CIÊNCIAS EXATAS E TECNOLOGIA
PROGRAMA INTERINSTITUCIONAL DE PÓS-GRADUAÇÃO EM ESTATÍSTICA UFSCar-USP

Giovanni Pastori Piccirilli

Bounded mixed regression models using Johnson-SB type distributions

Dissertation submitted to the Department of Statistics – DEs-UFSCar and to the Instituto de Ciências Matemáticas e de Computação – ICMC-USP, in partial fulfillment of the requirements for the degree of the Master Interagency Program Graduate in Statistics UFSCar-USP.

Advisor: Prof. Dr. Jorge Luis Bazán Guzmán

São Carlos
March 2021

UNIVERSIDADE FEDERAL DE SÃO CARLOS

Centro de Ciências Exatas e de Tecnologia
Programa Interinstitucional de Pós-Graduação em Estatística

Folha de Aprovação

Defesa de Dissertação de Mestrado do candidato Giovanni Pastori Piccirilli, realizada em 10/03/2021.

Comissão Julgadora:

Prof. Dr. Jorge Luis Bazán Guzmán (USP)

Prof. Dr. Caio Lucidius Naberezny Azevedo (UNICAMP)

Prof. Dr. Fernanda de Bastiani (UFPE)

O presente trabalho foi realizado com apoio da Coordenação de Aperfeiçoamento de Pessoal de Nível Superior - Brasil (CAPES) - Código de Financiamento 001.
O Relatório de Defesa assinado pelos membros da Comissão Julgadora encontra-se arquivado junto ao Programa Interinstitucional de Pós-Graduação em Estatística.

ACKNOWLEDGEMENTS

First of all, I thank my parents Paulo and Milene that made all this possible with their examples and teachings. For always believing and supporting me in all stages of my life, especially in this journey that made me feel closer to them than ever.

I would to expresses my gratitude to my advisor Jorge Luis Bazán for the time dedicated, support, and confidence in my research. His guidance has always guided me in the right way to the knowledge. Thank you for being a reference as a professional and making me acquire knowledge from this journey.

I would like to thank my committee members: Professor Fernanda De Bastiani and Professor Caio Lucidius Naberezny Azevedo, for their suggestions and time dedicated to the evaluation. Especially for professor Fernanda, who help in my research. My thanks also go to professors from PIPGEs, for the teachings that helped me along this journey. I will carry them with pride. To PIPGEs employees for administrative assistance and clarifications provided.

To my friends Cleidison, Matheus, college friends for life, Gilberto, the friend who was closest on this journey, and Erika, for the company of years. To friends from the PIPGEs that I had the pleasure to meet on this journey. Thank you for the hours of study, classes, and conversations in difficult times.

This work was carried out with financial support from the Coordenação de Aperfeiçoamento de Pessoal de Nível Superior - Brazil (CAPES).

“A ciência é um processo contínuo. Nunca termina. Não existe uma única e definitiva verdade a ser alcançada, após a qual todos os cientistas poderão se aposentar.”

(Carl Sagan)

RESUMO

PICCIRILLI, G. P. **Modelos de regressão misto para resposta limitada usando distribuições do tipo Johnson-SB**. 2021. 123 p. Dissertação (Mestrado em Estatística – Programa Interinstitucional de Pós-Graduação em Estatística) – Instituto de Ciências Matemáticas e de Computação, Universidade de São Paulo, São Carlos – SP, 2021.

Neste trabalho novas propriedades, métodos de estimação, análise de resíduos e extensões são desenvolvidas para modelos de regressão no intervalo (0,1) considerando as distribuições do tipo Johson SB (JSB). As extensões consideradas são o modelos inflacionados de zeros e uns e os modelos de regressão mista. Novos modelos mistos para dados longitudinais limitados no intervalo $(0,1)$ com base nas distribuições JSB são apresentados. Os estimadores de verossimilhança penalizada são obtidos maximizando a verossimilhança penalizada e calculados pelo algoritmo de Rigby e Stasinopoulos (RS). Na abordagem bayesiana, o algoritmo No-U-Turn-Sampler (NUTS) é usado para simular valores da distribuição a posterior. A análise de resíduos é realizada considerando os resíduos quantílicos. São apresentados estudos de simulação considerando a robustez a outliers das distribuições e extensões dos modelos para suportar observações 0 e 1. Três conjuntos de dados reais motivam o uso dos novos modelos. O primeiro conjunto de dados contém a proporção de indivíduos vulneráveis à pobreza dos 645 municípios do estado de São Paulo no Brasil e não contém nenhuma covariável. O segundo conjunto de dados contém a proporção de votos obtidos por um partido político em cinco eleições presidenciais brasileiras, a cada quatro anos, de 1994 a 2010, nos 75 municípios do estado de Sergipe no Brasil. O terceiro conjunto de dados é proveniente da área de saúde pública dos estados brasileiros. Ele contém as taxas de mortalidade por câncer brônquico e de pulmão nos 27 estados brasileiros nos últimos 30 anos. O objetivo é identificar se fatores como sexo, idade e Índice de Desenvolvimento Humano Municipal do estado podem influenciar na taxa de mortalidade. Os modelos de regressão misto JSB e o modelo misto Beta foram aplicados. Os modelos mistos JSB exibem valores mais baixos do que o modelo misto Beta para os critérios de comparação de modelos. Os resultados e a análise residual revelam que os modelos JSB podem ser uma alternativa ao modelo Beta.

Palavras-chave: Distribuição Johson-Sb, Modelos inflacionados, Modelos mistos, Resíduos quantilicos normalizados, Resposta limitada.

ABSTRACT

PICCIRILLI, G. P. **Bounded mixed regression models using Johnson-SB type distributions**. 2021. 123 p. Dissertação (Mestrado em Estatística – Programa Interinstitucional de Pós-Graduação em Estatística) – Instituto de Ciências Matemáticas e de Computação, Universidade de São Paulo, São Carlos – SP, 2021.

In this work, new properties, estimation methods, residual analysis and extensions are developed for regression models in the (0,1) interval considering the Johnson-SB type distributions (JSB). The extensions are zero-and-one inflated models and mixed regression models. New mixed-effects models for bounded longitudinal data in the interval $(0,1)$ based on the JSB distributions are presented. The penalized likelihood estimators are obtained by maximizing the penalized likelihood and are computed by the Rigby and Stasinopoulos (RS) algorithm. From the Bayesian perspective, the No-U-Turn-Sampler (NUTS) is used to sample from the posterior distribution. Residual analysis is performed considering randomized quantile residuals. Simulation studies considering robustness to outliers from the distributions and extensions of the models to support 0 and 1 observations are presented. Three real data sets motivate the use of the new models. The first dataset contains the proportion of individuals vulnerable to poverty of the 645 municipalities from São Paulo state in Brazil and with no covariate. The second dataset incorporates the proportion of votes obtained by a political party in five Brazilian presidential elections, every four years, from 1994 to 2010, from the 75 municipalities from Sergipe state in Brazil. The third dataset comes from the public health area in Brazilian states. It contains the mortality rates from bronchial and lung cancer from the 27 Brazilian states over the last 30 years. The aim is to identify if factors like sex, age, and the Municipal Human Development Index of the state can influence the mortality rate. The JSB mixed regression models and the Beta mixed model were applied. The JSB mixed models display better values than the Beta mixed model for the model comparison criteria. The results and the residual analysis reveal that the JSB models are an alternative to the Beta model.

Keywords: Bounded response, Johnson-Sb distribution, Inflated models, Mixed models, Normalized randomized quantile residuals.

LIST OF FIGURES

LIST OF TABLES

CONTENTS

CHAPTER 1

INTRODUCTION

Linear regression models have been the central tool to analyze data that contains a response variable related to covariates. However, with the limitation of linear regression models, new alternatives became more appropriate for some cases. For example, the generalized linear model (GLM) extends the regression models for any distribution from the exponential family, and it has brought improvement in the regression models research.

The first topic of this work is response variables in the $(0,1)$ interval. The linear regression model and the GLM are inappropriate to make inferences around these bounded variables. The response variable in the unit interval is frequent in many real problems. Examples of these variables are the country's unemployment proportion, poverty rate, illiteracy rate, and the percentage of the population using at least basic sanitation service.

The second topic of this work is longitudinal data. Longitudinal data are measures taken from the same subject over time, and they can be a particular case of multilevel or hierarchical data with two levels, level one consisting of repeated measures and the level two by individuals (OLIVEIRA, 2015). Longitudinal data may come from a medical study, that the purpose is to analyze the blood pressure of different individuals over a certain period. The data will contain the individuals' measurements over a given period.

Several regression models for bounded response variables have been proposed in recent years. Ferrari and Cribari-Neto (2004) suggested a regression model based on the Beta distribution, Lemonte and Bazán (2016), followed Johnson (1949) and presented a broad class of distribution with bounded support based on the symmetric family of distributions called generalized Johnson-S_b distribution. Migliorati, Brisco and Ongaro (2018) also presented regression models for bounded response variables based on flexible Beta distribution. All these references are improvements to bounded regression models. For the longitudinal data, a mixed model is a popular approach because of the flexibility that it offers to analyze clustered and longitudinal data. Bayes, Bazán and Castro (2017) proposed a quantile parametric mixed model based on the Kumarasawamy distribution, Figueroa-Zúñiga, Arellano-Valle and Ferrari (2013) showed a mixed beta regression under a Bayesian perspective, Verkuilen and Smithson (2012) also describe a mixed beta regression, but under a classical approach.

The primary condition of mixed models is that some subset of regression parameters vary randomly from one individual to another, then the individuals in a population have their trajectory. The most popular mixed model is the linear mixed model (LMM), an extension of the linear model that assumes normality for the response variable and accommodates random effects. The response variable from LMM is a linear combination of characteristics from the population shared with all individuals (fixed effects) and specific influences that are specific to each individual (random effects). However, sometimes we need to build a model that accommodates correlated data, and our response is not normally distributed. The use of random effects goes beyond normality, like the Generalized Linear Mixed Models (GLMM), a class of model that allows the use of random effects in distribution that belong to the exponential family (SEARLE; MCCULLOCH, 2001). The LMM and GLMM covariances of random effects have an impact on the structure of individuals' variances. It implies that assuming random effects, we capture the correlation that comes from repeated and clustered data.

This work considers a flexible mechanism for constructing probability distributions in a unit interval. This new flexible class of distribution in the $(0,1)$ interval is called the GF-quantile distributions. In order to introduce this new class of distributions, let W be a random variable with cumulative density function $G(w;\theta)$ and probability density function (pdf) $g(w;\theta)$ on the support R_W. Let X be other variable continues with cumulative density function (cdf) $H(x;\varphi)$ and quantile function $Q(y;\varphi) = H^{-1}(y;\varphi)$, $0 < y < 1$, where $R_X = R_W$ and φ is a know parameter from $H(x;\varphi)$ cdf Rodrigues, Bazán and Suzuki (2019).

To obtain the distribution of the GF-quantile class we define the composite probability distribution function given by:

$$G_h(y;\theta,\varphi) = G(Q(y;\varphi);\theta) = \int_{-\infty}^{Q(y;\varphi)} g(w;\theta)dw, y \in (0,1), \tag{1.1}$$

where $G_h(y;\theta,\varphi)$ is the cdf of the continuous random variable Y on the $(0,1)$ interval, with parameters θ and φ. The pdf of the GF-quantile family of distributions in (1.1) can be obtained by:

$$g_h(y;\theta,\varphi) = g(Q(y;\varphi);\theta)q(y;\varphi), y \in (0,1), \tag{1.2}$$

where $q(y;\varphi)$ is defined by $q(y;\varphi) = \frac{dQ(y;\varphi)}{dy}$.

From that method, we obtain many distributions in the interval $(0,1)$. We need a continuous baseline distribution and a quantile function of a transformed distribution, where the support of the two distributions are the same. Rodrigues, Bazán and Suzuki (2019) suggested other alternatives for H. However, a different choice of H brings particular support for the new distribution, and our purpose here is to extend the class of Johnson (JOHNSON, 1949) and generalized Johnson (LEMONTE; BAZÁN, 2016) distributions to a bounded response.

This work focus on distributions from the GF-quantile class obtained using transformed distributions with support on the real line. More specifically, a new derivation of GF-quantile distributions is introduced, which are an extension of Johnson-S_b and generalized JohnsonS_b distributions called the Johnson-SB type (JSB) distributions. Rodrigues, Bazán and Suzuki (2019) have proposed the regression models for the JSB distributions. However, the authors omitted many details about the classical estimation, and they have not presented the Bayesian estimation of the model. This work gives more details about the classical estimation, and the Bayesian estimation is presented for the model. A new model called the JSB mixed regression model is developed in this work, and it is not previously presented in the literature.

The work is organized as follows. Chapter 2 presents the JSB class of distributions, including definition, examples, and some properties. Chapter 3 shows the JSB mixed regression model for bounded response variables; Chapter 4 presents the estimation method for Bayesian and classical approaches, including model comparison criteria, and propose residual analysis considering randomized quantile residuals. Chapter 5 presents a simulation study to evaluate the robustness of the proposed distributions to outliers and a parameter recovery study to illustrate the performance of the classical estimates of an extension of the JSB distributions called zero-and-one inflated distribution. Chapter 6 presents the results of three applications in real datasets. And Chapter 7 shows final remarks, including some proposals for future extension of the model presented here.

CHAPTER 2

JOHNSON-SB TYPE DISTRIBUTIONS

This Chapter presents the GF-quantile distributions, which can be obtained using a baseline distribution with support in $\mathbb{R}$ and cumulative density function (cdf) given by $G(\boldsymbol{\theta})$, with $\boldsymbol{\theta} = (\mu, \sigma, \nu)'$, where μ is a location parameter, σ is a scale parameter and ν is a shape parameter. It is not restricted that the parameters μ, σ and ν are location, scale and shape parameters respectively. In this work, it is assumed the Logistic distribution with $\varphi = (0,1)$ as transformed distribution $H(\varphi)$, so the quantile function $H^{-1}(\varphi)$ is given by:

$$Q(y) = \log\left(\frac{y}{1-y}\right). \tag{2.1}$$

Because Q transforms the response, it is known as a transformation function, and models in the form of 1.2 are called transformation models. The following notation $f(.) := g_h(.)$ and $F(.) := G_h(.)$ will be used throughout this work to denote the pdf and cdf of the JSB distribution class. If the random variable Y belongs to the JSB class of distributions, we write $Y \sim JSB(\boldsymbol{\theta})$ with $\boldsymbol{\theta} = (\mu, \sigma, \nu)'$. From Equation (1.2), the $JSB(\boldsymbol{\theta})$ pdf is given by:

$$f(y;\boldsymbol{\theta}) = g\left(\log\left(\frac{y}{1-y}\right);\boldsymbol{\theta}\right)\frac{1}{y(1-y)}. \tag{2.2}$$

This new derivation of the GF-quantile distribution is an extension of the Johnson-S_b and the generalized Johnson-S_b. The following distributions were used as baseline distributions $G(\boldsymbol{\theta})$:

a) Gumbel distribution: $\boldsymbol{\theta} = (\mu, \sigma)$;

b) Reverse Gumbel distribution: $\boldsymbol{\theta} = (\mu, \sigma)$;

c) Exponential Gaussian distribution: $\boldsymbol{\theta} = (\mu, \sigma, \nu)$;

d) Skew normal type I distribution: $\boldsymbol{\theta} = (\mu, \sigma, \nu)$.

e) Power exponential distribution: $\boldsymbol{\theta} = (\mu, \sigma, \nu)$;

The motivation for these distributions is that they are distributions with support on $(-\infty, \infty)$ and have already been presented in Rodrigues, Bazán and Suzuki (2019). Also, they are available at the `gamlss.dist` package. These distributions are not new and some references about them are given in this Chapter. Here many unprecedented details and properties about these distributions will be presented. A comprehensive review of all baseline distributions, which include the pdf and cdf, can be found in the book Rigby *et al.* (2019).

2.1 Definition of the Johnson-SB type distributions

Definition 1. Let W be a random variable with support on the $\mathbb{R}_W$ and cdf defined by $G(w;\boldsymbol{\theta})$. Let X be another random variable following the standard logistic distribution with quantile function defined by $Q(y;\varphi)$ for $0 < y < 1$ and $\varphi = (0,1)$. To obtain the new flexible class of distribution on the (0,1) interval, called *JSB* distribution, we define the composite cumulative density function:

$$F(y;\boldsymbol{\theta},\varphi) = G(Q(y;\varphi);\boldsymbol{\theta}), \tag{2.3}$$

for $y \in (0,1)$. The $F(y;\boldsymbol{\theta},\varphi)$ function is the cdf of continuous random variable Y on the $(0,1)$ interval with parameters $\boldsymbol{\theta}$ and $\varphi = (0,1)$. The following notation $Y \sim JSB(\boldsymbol{\theta})$ will be used throughout this work to denote that the random variable Y following a probability distribution from the *JSB* class.

2.1.1 Logistic Gumbel

The Gumbel distribution belongs to a group of distribution called Extreme value distributions. The Gumbel, also called the generalized extreme value distribution Type-I, is used to model the maximum (or the minimum) distribution of some samples from various distributions. The Gumbel distribution is appropriate for moderately negative skew data.

Let the Gumbel distribution be the baseline distribution with cdf $G(y;\boldsymbol{\theta})$, $\boldsymbol{\theta} = (\mu,\sigma)$ where μ is the mode and the location parameter, σ is the scaling parameter and support in $(-\infty,\infty)$. Assume that the transformed distribution $H(\varphi)$ is the Logistic distribution with quantile function defined in (2.1), then the logistic Gumbel (LGU) cdf is given by:

$$F(y;\mu,\sigma) = 1 - \exp\left[-\exp\left(\frac{\log\left(\frac{y}{1-y}\right) - \mu}{\sigma}\right)\right], \tag{2.4}$$

for $0 < y < 1, -\infty < \mu < \infty$ and $\sigma > 0$, which can be readily inverted to give the quantile function:

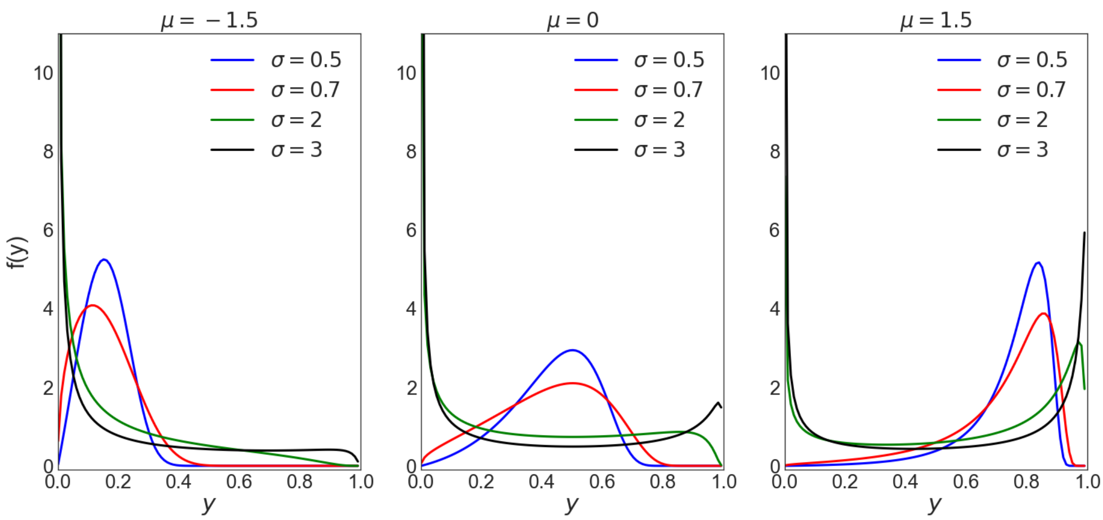
$$Q_Y(p) = F_Y^{-1}(p) = \frac{\exp\{\mu + \sigma \log(-\log(1-p))\}}{1+\exp\{\mu + \sigma \log(-\log(1-p))\}}. \tag{2.5}$$

The pdf of the logistic Gumbel distribution is given by:

$$f(y;\mu,\sigma) = \frac{1}{\sigma}\exp\left\{\left(\frac{\log\left(\frac{y}{1-y}\right)-\mu}{\sigma}\right) - \exp\left(\frac{\log\left(\frac{y}{1-y}\right)-\mu}{\sigma}\right)\right\}\frac{1}{y(1-y)}. \tag{2.6}$$

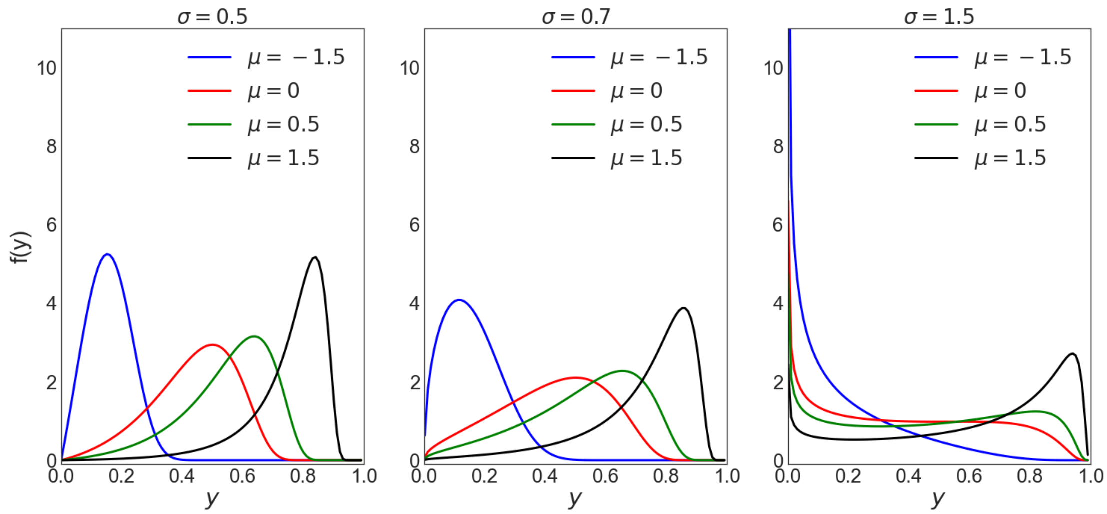
Figures 1 and 2 show the shapes for some values of μ and σ. It is possible to see that μ is the location parameter. The location of the density curve occurs as the μ values change. The density curve flattens as the σ values change, and this shows the dispersion effect of the σ parameter. For the high values of σ (2 and 3) the density is allocated at the extreme points of the axis y. For low and moderate values of σ (0.5 and 0.7) the density curve is more behaved.

Figure 1 – Logistic Gumbel probability density function for parameter μ = -1.5, 0 and 1.5 and some choices of parameter σ.



Source: Elaborated by the author.

Figure 2 – Logistic Gumbel probability density function for parameter σ = 0.5, 0.7 and 1.5 and some choices of parameter μ.



Source: Elaborated by the author.

2.1.2 Logistic reverse Gumbel

The reverse Gumbel is a special case of the generalized extreme value distribution. If a random variable $T \sim RG(\mu,\sigma)$ and another random variable $U = -T$, then $U \sim GU(-\mu,\sigma)$. The reverse Gumbel distribution is appropriate for moderately positive skew data.

Assuming the reverse Gumbel distribution as the baseline distribution with cdf $G(y;\boldsymbol{\theta})$, $\boldsymbol{\theta} = (\mu,\sigma)$ where μ is the mode and the location parameter, σ is the scaling parameter and support on $(-\infty,\infty)$ the logistic reverse Gumbel (LRG) cdf is given by:

$$F(y;\mu,\sigma) = \exp\left\{-\exp\left[-\left(\frac{\log\left(\frac{y}{1-y}\right)-\mu}{\sigma}\right)\right]\right\}, \tag{2.7}$$

for $0 < y < 1, -\infty < \mu < \infty$, and $\sigma > 0$. The quantile function is given by:

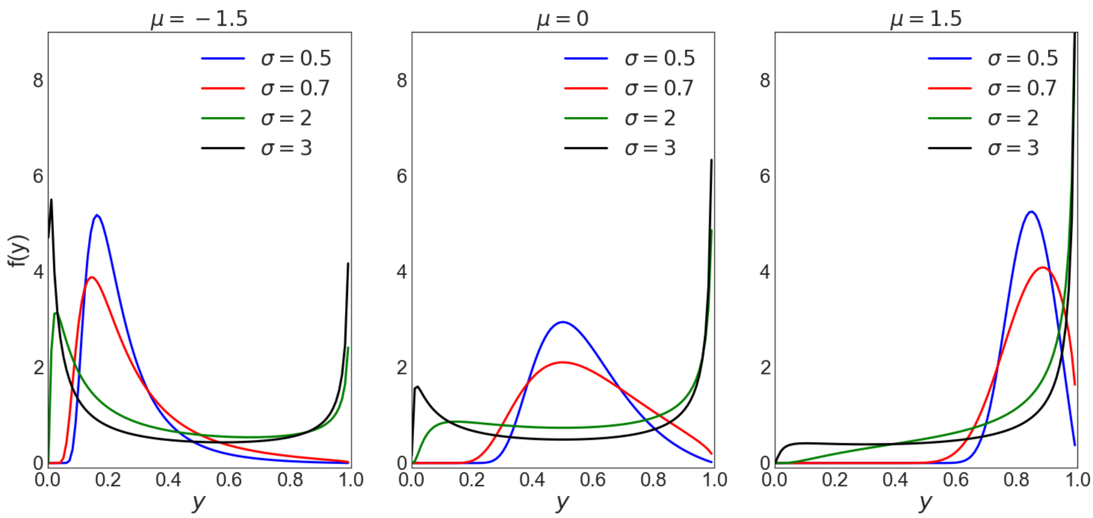
$$Q_Y(p) = F_Y^{-1}(p) = \frac{\exp\{\mu - \sigma\log(-\log(p))\}}{1+\exp\{\mu-\sigma\log(-\log(p))\}}. \tag{2.8}$$

The pdf of the logistic reverse Gumbel distribution is given by

$$f(y;\mu,\sigma) = \frac{1}{\sigma}\exp\left\{-\left(\frac{\log\left(\frac{y}{1-y}\right)-\mu}{\sigma}\right)-\exp\left[-\left(\frac{\log\left(\frac{y}{1-y}\right)-\mu}{\sigma}\right)\right]\right\}\frac{1}{y(1-y)}, \tag{2.9}$$

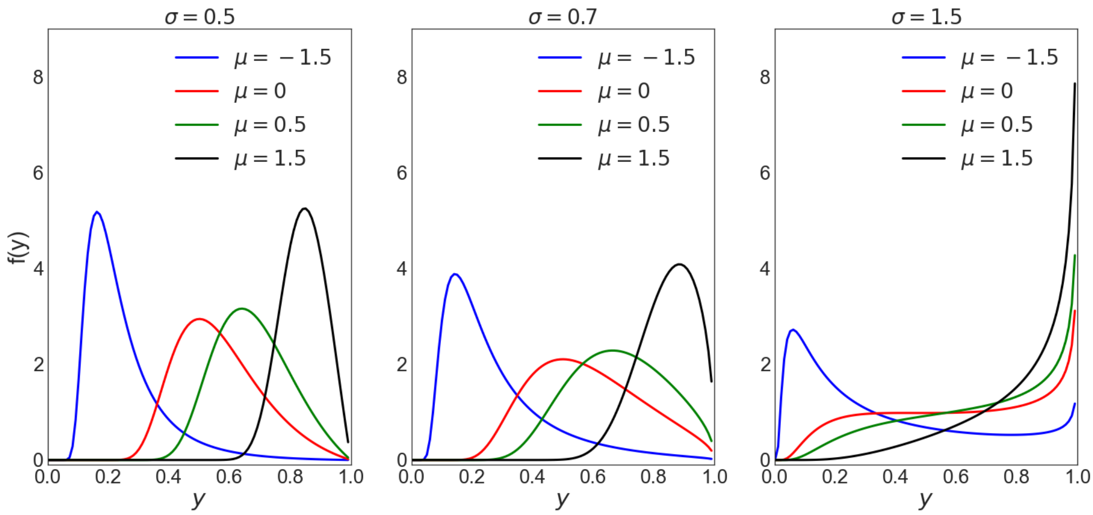
Figures 3 and 4 show the shapes for some values of μ and σ. The density of the logistic Gumbel and the logistic Reverse Gumbel are very similar, then it is possible to see the same effects of the parameters. The parameter μ controls the distribution curve location, and σ refers to the dispersion.

Figure 3 – Logistic reverse Gumbel probability density function for parameter μ = -1.5, 0 and 1.5 and some choices of parameter σ.



Source: Elaborated by the author.

Figure 4 – Logistic reverse Gumbel probability density function for parameter σ = 0.5, 0.7 and 1.5 and some choices of parameter μ.



Source: Elaborated by the author.

2.1.3 Logistic exponential Gaussian

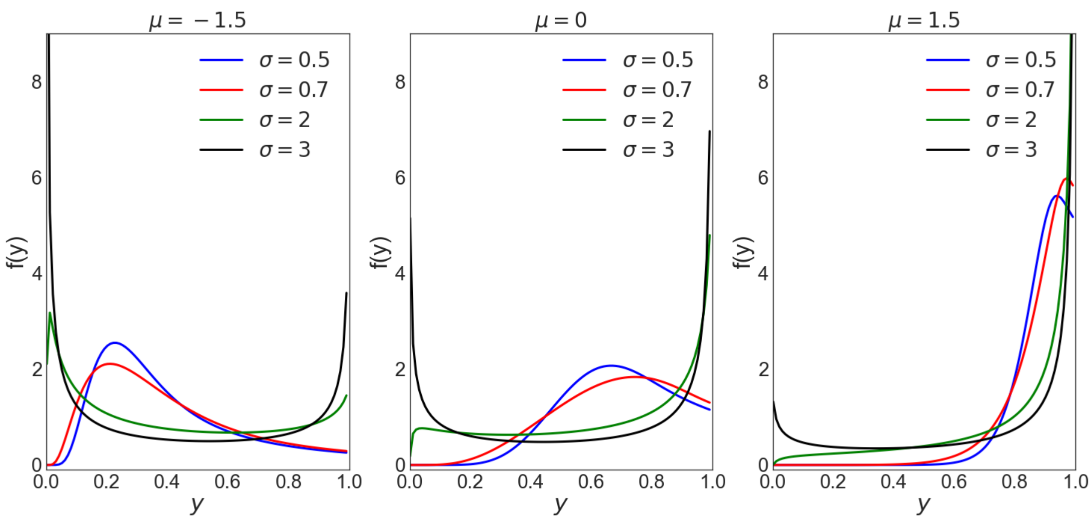
If two random variables U_1 and U_2 defined by $U_1 \sim N(\mu, \sigma^2)$ and $U_2 \sim Exp(\nu)$ with U_1 independent of U_2, then their sum $U = U_1 + U_2$ follows an exponential Gaussian distribution. The exponential Gaussian does not have a closed form for the cdf, and it is appropriate for positive skew data (LOVISON; SCHINDLER, 2014).

Let the exponential Gaussian distribution be the baseline distribution with cdf $G(y;\boldsymbol{\theta})$, $\boldsymbol{\theta} = (\mu, \sigma, \nu)$ where μ is the mean of the normal component, σ is the standard deviation of the normal component, ν is the mean of the exponential component and support in $(-\infty, \infty)$. The density of the logistic exponential Gaussian (LEG) distribution is given by:

$$f(y;\mu,\sigma,\nu) = \frac{1}{\nu}\exp\left\{\frac{\mu - \log\left(\frac{y}{1-y}\right)}{\nu} + \frac{\sigma^2}{2\nu^2}\right\}\Phi\left(\frac{\log\left(\frac{y}{1-y}\right) - \mu}{\sigma} - \frac{\sigma}{\nu}\right), \tag{2.10}$$

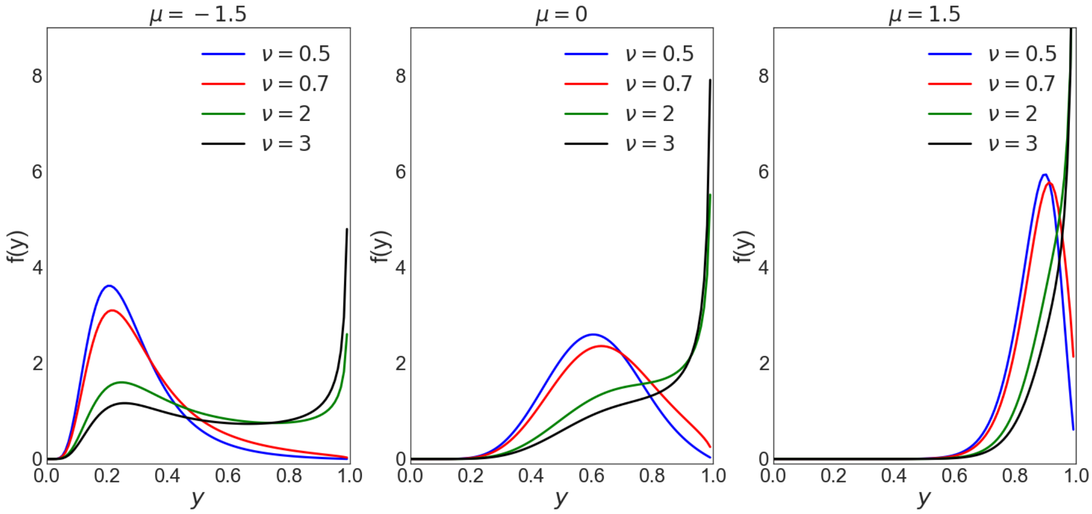
for $0 < y < 1$, $-\infty < \mu < \infty$, $\sigma > 0$ and $\nu > 0$. The logistic exponential Gaussian does not have a closed form for cdf. The Figures 5, 6, 7, 8, 9 and 10 show the density curve shapes for some values of μ, σ and ν. The parameter μ controls the location of the density curve as can see in 7 and 9 since the location of the density change as the μ values change. The σ parameter controls the dispersion as we can see in 5 and 10 because the density curve flattens as the σ values change. The last parameter ν seems to control the Kurtosis and some degree of Skewness of distribution. The Kurtosis of the curve density seems low for lower values of ν.

Figure 5 – Logistic exponential Gaussian probability density function for parameter μ = -1.5, 0 and 1.5,ν = 1 and some choices of parameter σ.



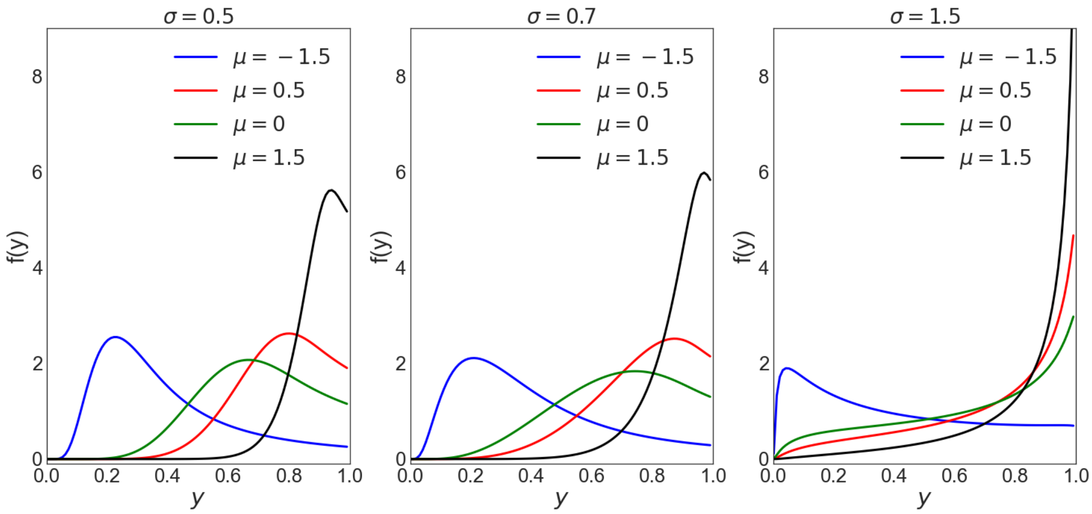
Source: Elaborated by the author.

Figure 6 – Logistic exponential Gaussian probability density function for parameter μ = -1.5, 0 and 1.5, σ = 0.5 and some choices of parameter ν.



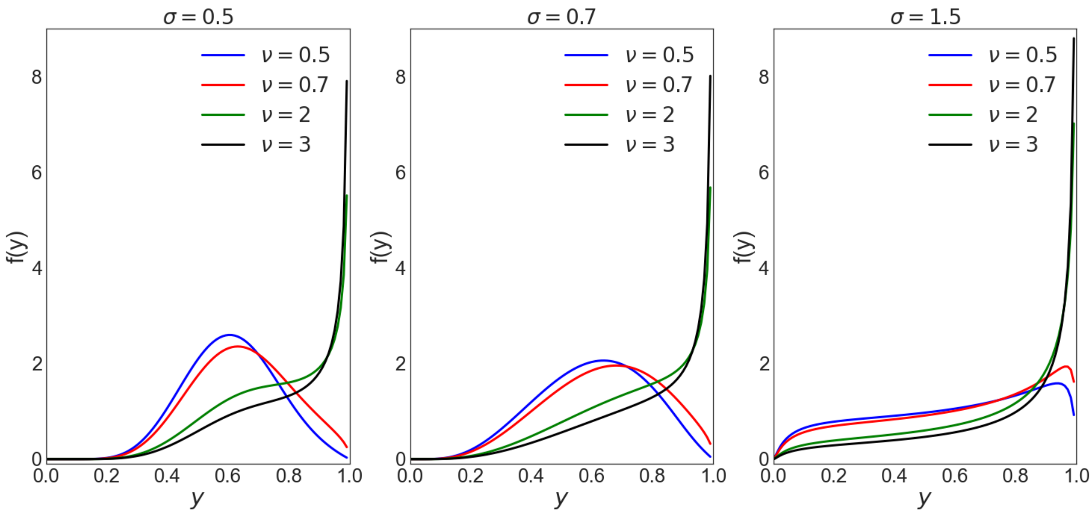
Source: Elaborated by the author.

Figure 7 – Logistic exponential Gaussian probability density function for parameter σ = 0.5, 0.7 and 1.5, ν = 1 and some choices of parameter μ.



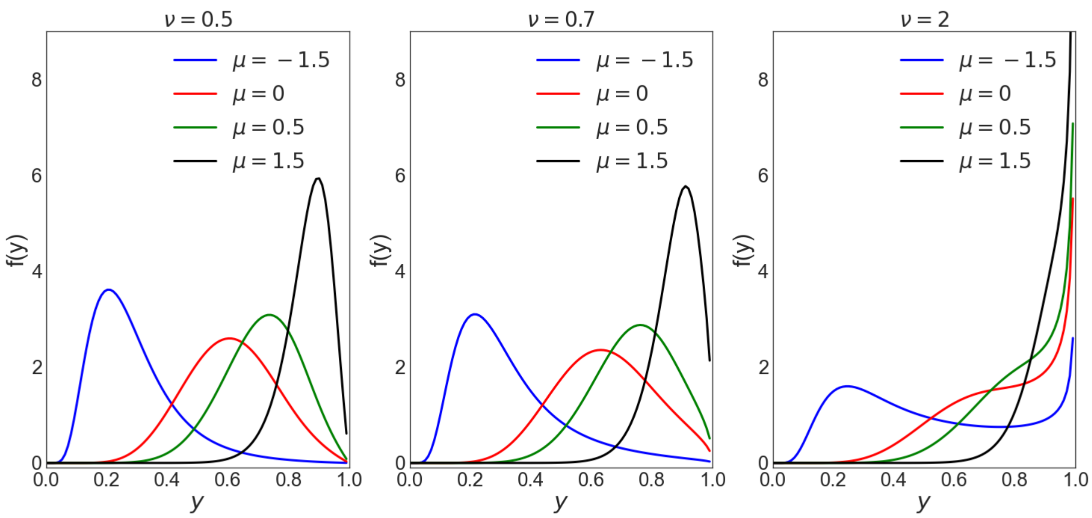
Source: Elaborated by the author.

Figure 8 – Logistic exponential Gaussian probability density function for parameter σ = 0.5, 0.7 and 1.5, μ = 0 and some choices of parameter ν.



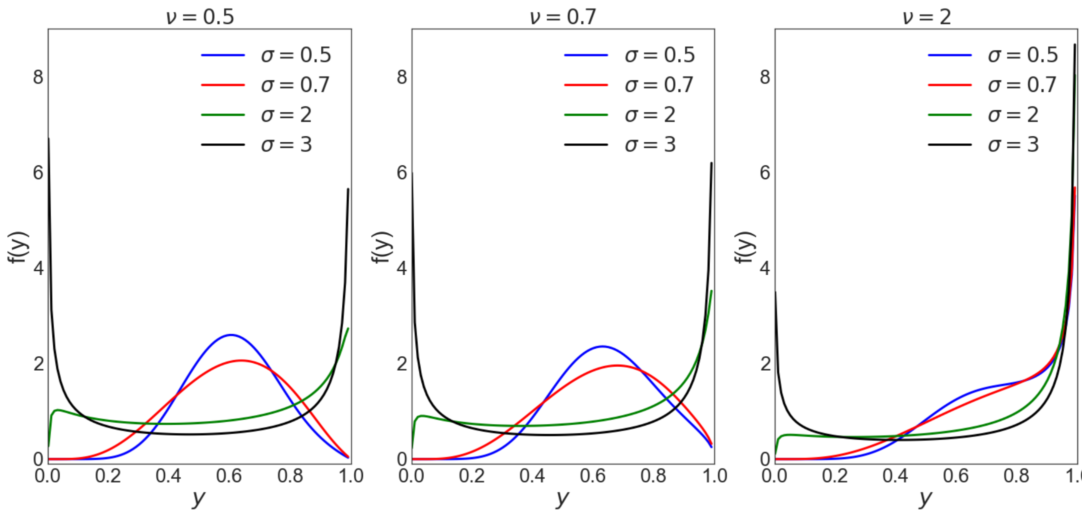
Source: Elaborated by the author.

Figure 9 – Logistic exponential Gaussian probability density function for parameter ν = 0.5, 0.7 and 2, σ = 0.5 and some choices of parameter μ.



Source: Elaborated by the author.

Figure 10 – Logistic exponential Gaussian probability density function for parameter ν = 0.5, 0.7 and 2, μ = 0 and some choices of parameter σ.



Source: Elaborated by the author.

2.1.4 Logistic skew normal

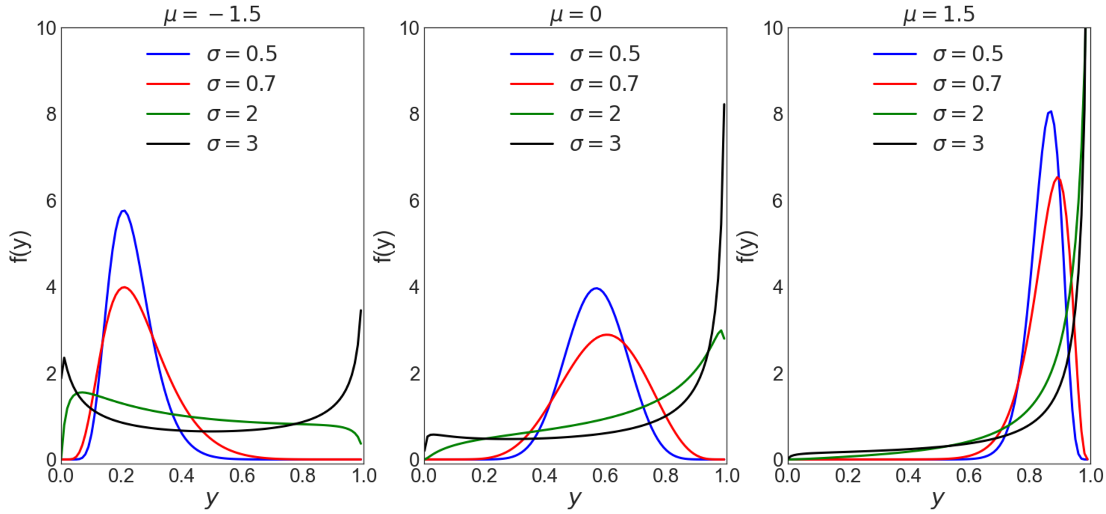
The skew normal type I distribution is presented by Azzalini (1985). The skew normal type I distribution includes the normal and the half-normal distributions as particular cases, and it does not have a closed-form for the cdf.

Assuming the skew normal type I distribution as the baseline distribution with cdf $G(y;\boldsymbol{\theta})$, $\boldsymbol{\theta} = (\mu, \sigma, \nu)$ where μ is the location parameter, σ is the scaling parameter, ν is the skewness parameter and support $(-\infty, \infty)$ then the density of the logistic skew normal (LSN) distribution is given by:

$$f(y;\mu,\sigma,\nu) = \frac{2}{\sigma}\phi(z)\Phi(\nu z)\frac{1}{y(1-y)}, \tag{2.11}$$

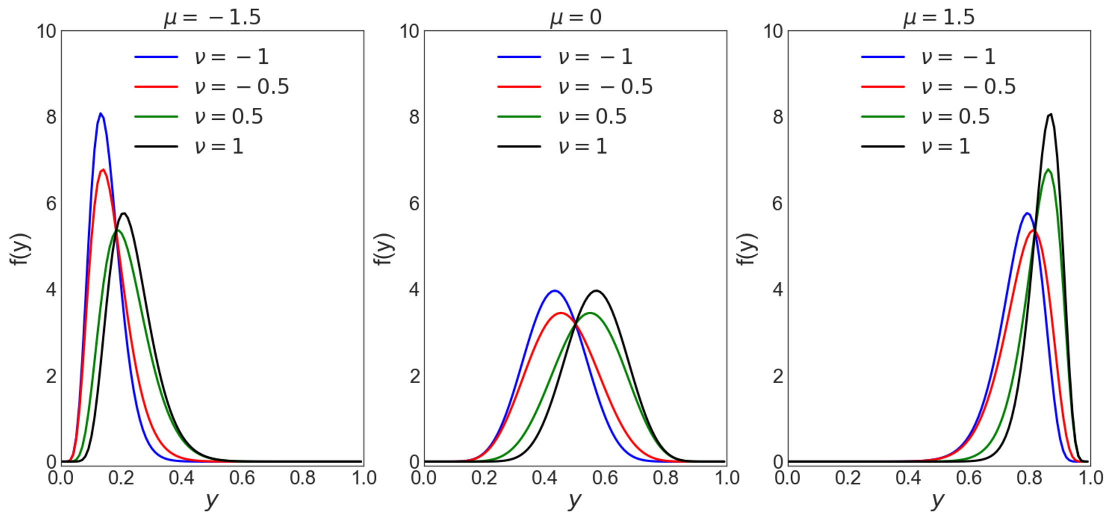
for $0 < y < 1$ with $z = \left(\log\left(\frac{y}{1-y}\right) - \mu\right)/\sigma$. The logistic skew normal does not have a closed form for the cdf. The Figures 11, 12, 13, 14, 15 and 16 show the shapes for some values μ, σ and ν. The parameter μ controls the location of the density curve as can see in 13 and 15. Clearly as the μ parameters is changed the the location of the curve on the graph is changed. The σ controls the dispersion as can see in 11 and 16, since he density curve flattens as the σ values change. The ν parameter affect the Kurtosis of the curve, however this parameters seems to has a little impact on the Skewnees as can see in 12 and 14.

Figure 11 – Logistic skew normal probability density function for parameter μ = -1.5, 0 and 1.5,ν = 1 and some choices of parameter σ.



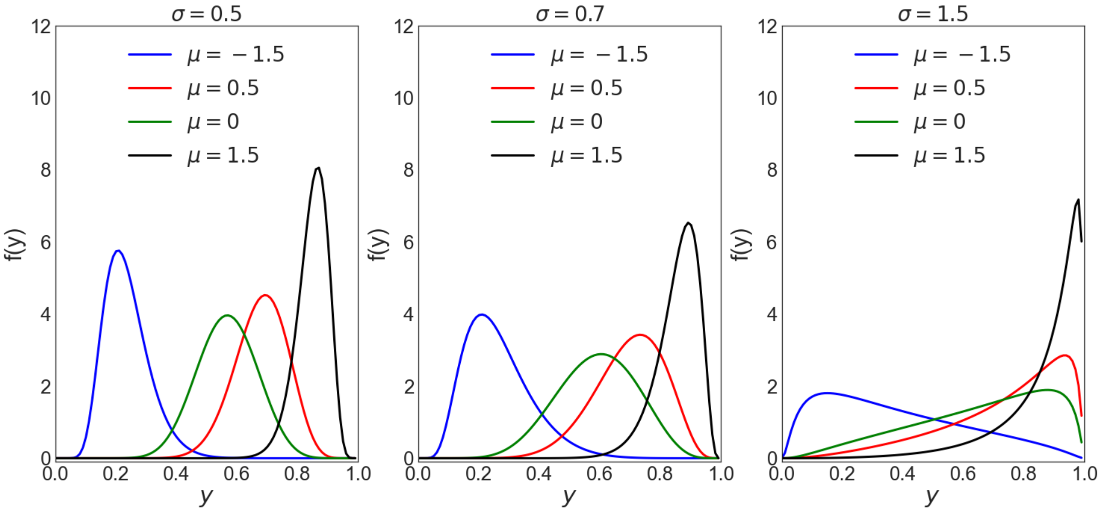
Source: Elaborated by the author.

Figure 12 – Logistic skew normal probability density function for parameter μ = -1.5, 0 and 1.5, σ = 0.5 and some choices of parameter ν.



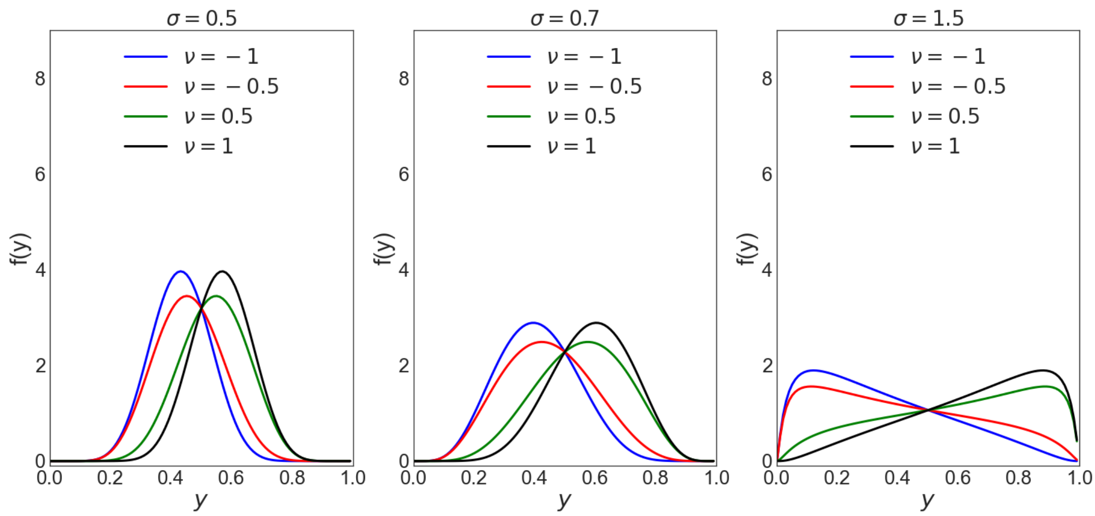
Source: Elaborated by the author.

Figure 13 – Logistic skew normal probability density function for parameter $\sigma = 0.5$, 0.7 and 1.5, $\nu = 1$ and some choices of parameter μ.



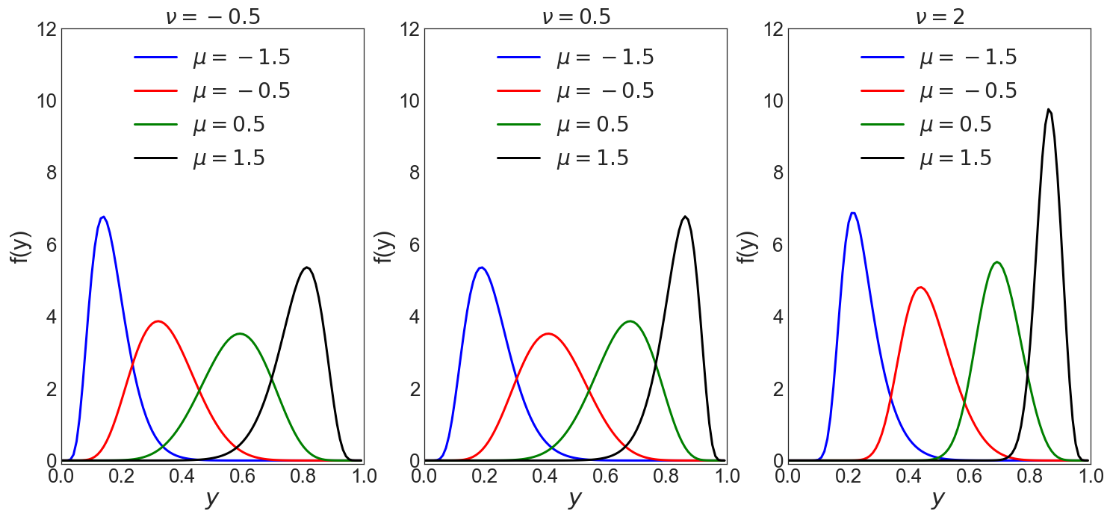
Source: Elaborated by the author.

Figure 14 – Logistic skew normal probability density function for parameter $\sigma = 0.5$, 0.7 and 1.5, $\mu = 0$ and some choices of parameter ν.



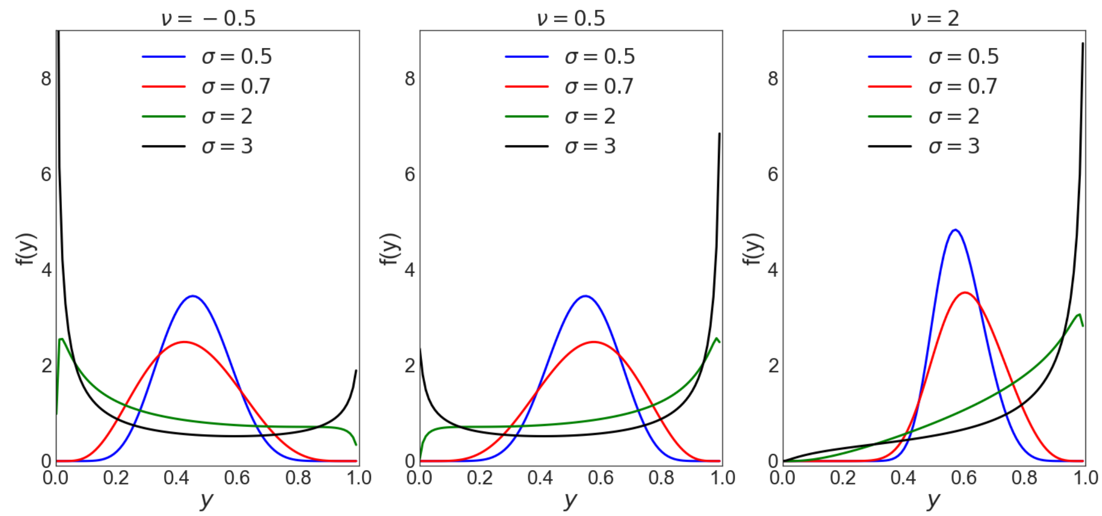
Source: Elaborated by the author.

Figure 15 – Logistic skew normal probability density function for parameter ν = -0.5, 0.5 and 2, σ = 0.5 and some choices of parameter μ.



Source: Elaborated by the author.

Figure 16 – Logistic skew normal probability density function for parameter ν = 0.5, -0.5 and 2, μ = 0 and some choices of parameter σ.



Source: Elaborated by the author.

2.1.5 Logistic power exponential

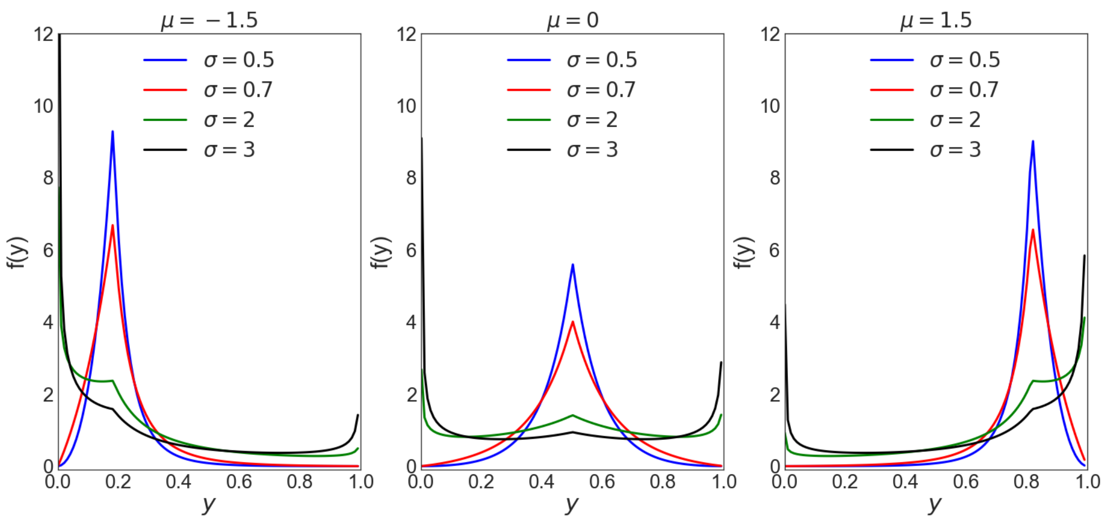
The parametrization of the power exponential distribution used here is presented by Nelson (1991). The power exponential distribution includes the Laplace and normal distribution as particular cases. The power exponential distribution is suitable for leptokurtic as well as platykurtic data.

Assuming the power exponential distribution as the baseline distribution with cdf $G(y;\boldsymbol{\theta})$, $\boldsymbol{\theta} = (\mu, \sigma, \nu)$ where μ is the mean, median, mode, σ is the standard deviation, ν the Kurtosis and support $(-\infty, \infty)$ then the logistic power exponential (LPE) density is given by:

$$f(y;\mu,\sigma,\nu) = \frac{\nu \exp(-|z|^{\nu})}{2c\sigma\Gamma(1/\nu)} \frac{1}{(1-y)y}, \tag{2.12}$$

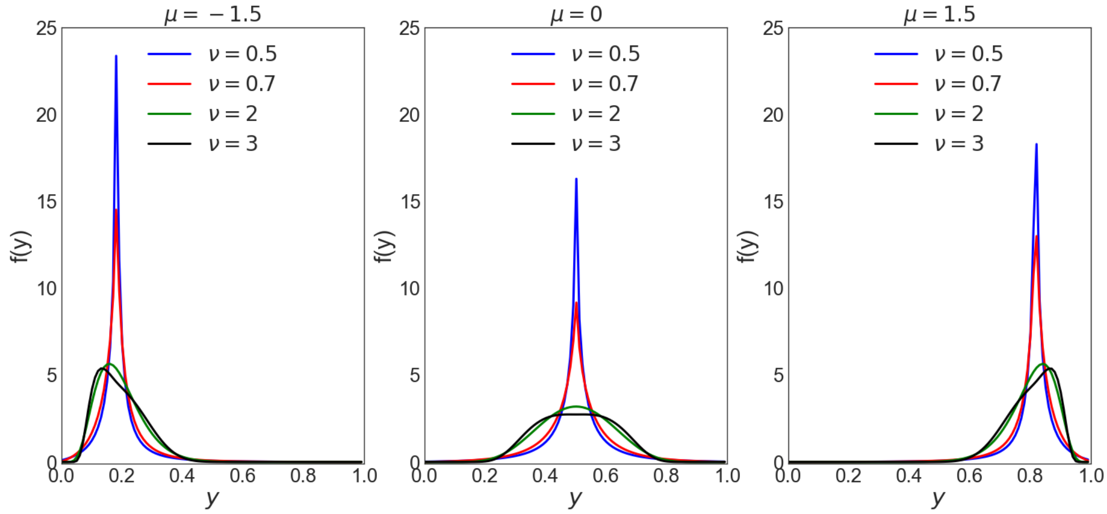
for $0 < y < 1, -\infty < \mu < \infty, \sigma > 0, \nu > 0, z = (\log\left(\frac{y}{1-y}\right) - \mu)/(c\sigma)$ and $c^2 = \Gamma(1/\nu)[\Gamma(3/\nu)^{-1}]$. The logistic power exponential does not have a closed form for cdf. The Figures 17, 18, 19, 20, 21 and 22 show the shape of logistic power exponential for some values of μ, σ and ν. Like the previously distributions, the μ parameter affect the location of the density curve. An important point showed by the Figures is the symmetry of the density for $\mu = 0$. The parameter ν, like in the Logistic Skew Normal, impacts the Kurtosis more than the Skewness.

Figure 17 – Logistic power exponential probability density function for parameter μ = -1.5, 0 and 1.5, ν = 1 and some choices of parameter σ.



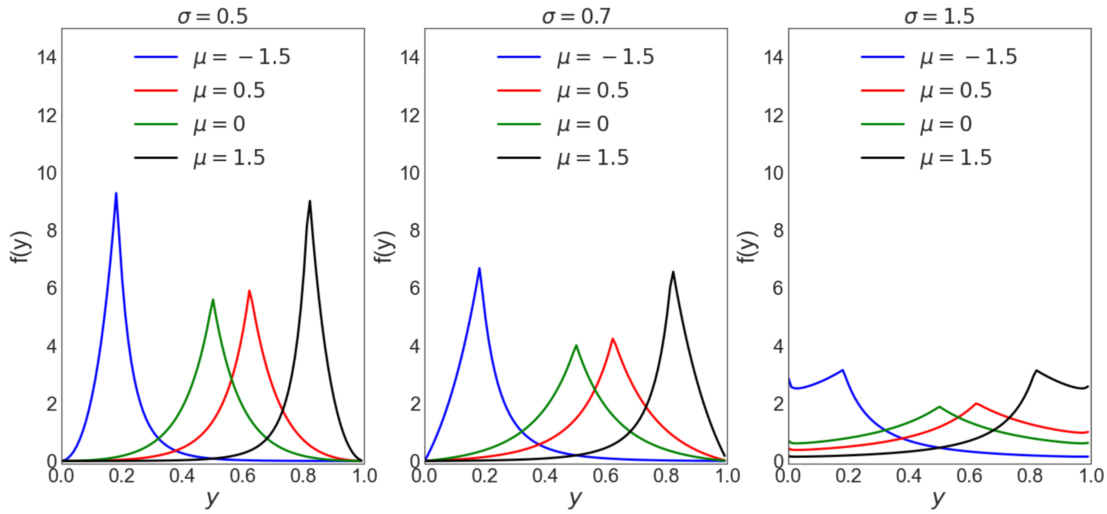
Source: Elaborated by the author.

Figure 18 – Logistic power exponential probability density function for parameter μ = -1.5, 0 and 1.5, σ = 0.5 and some choices of parameter ν.



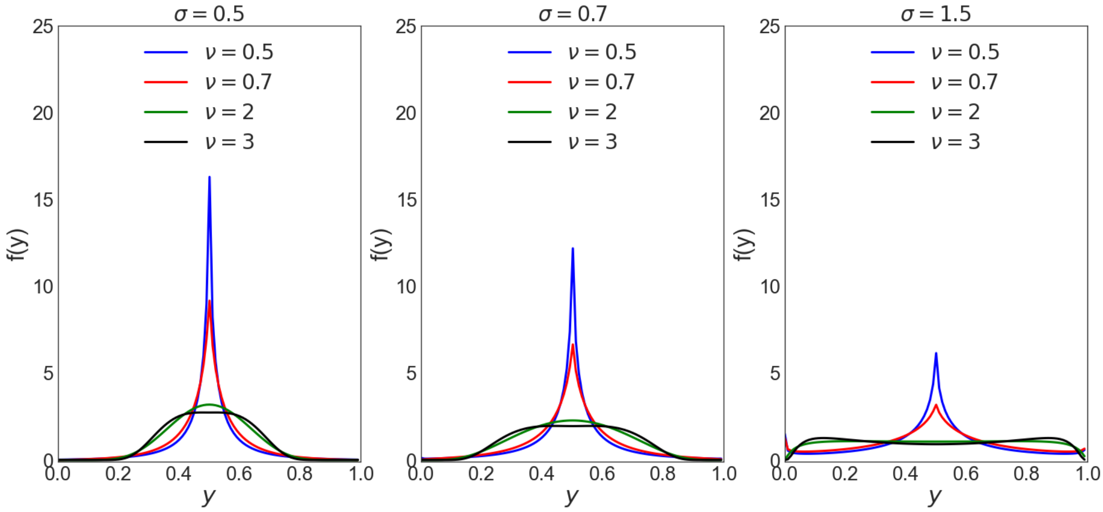
Source: Elaborated by the author.

Figure 19 – Logistic power exponential probability density function for parameter σ = 0.5, 0.7 and 1.5, ν = 1 and some choices of parameter μ.



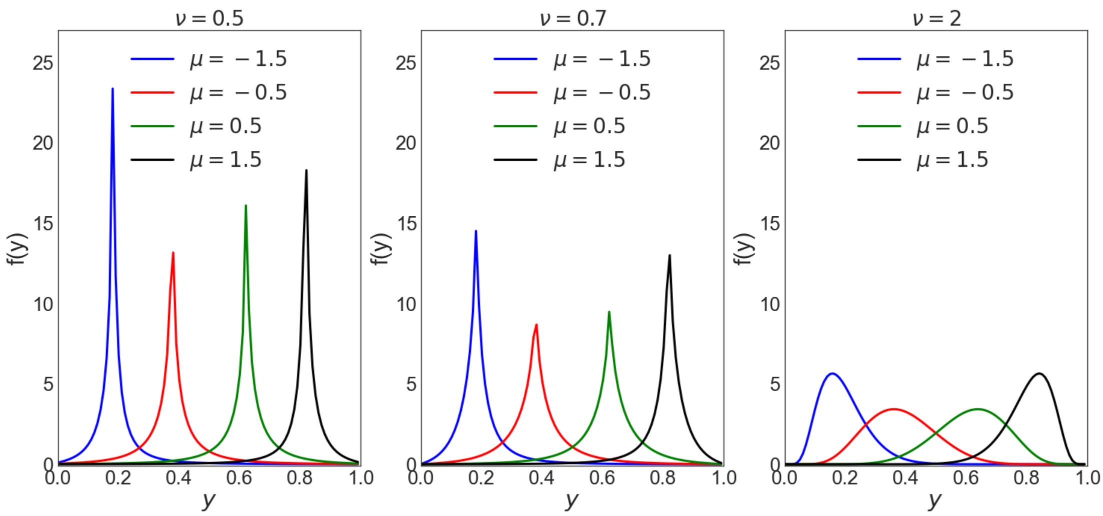
Source: Elaborated by the author.

Figure 20 – Logistic power exponential probability density function for parameter σ = 0.5, 0.7 and 1.5, μ = 0 and some choices of parameter ν.



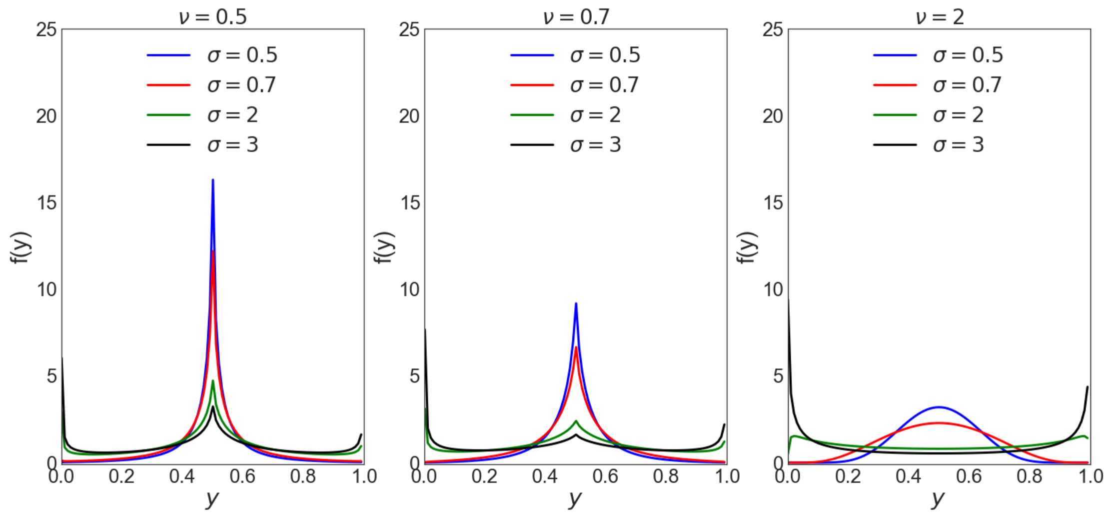
Source: Elaborated by the author.

Figure 21 – Logistic power exponential probability density function for parameter ν = 0.5, 0.7 and 2, σ = 0.5 and some choices of parameter μ.



Source: Elaborated by the author.

Figure 22 – Logistic power exponential probability density function for parameter ν = 0.5, 0.7 and 2, μ = 0 and some choices of parameter σ.



Source: Elaborated by the author.

2.2 Properties

This session give some properties from logistc Gumbel, logistic reverse Gumbel, logistic exponential gaussian, logistic skew normal and logistic power exponential. These properties are contributions from this work.

Proposition 1. Let Y be a random variable such that $Y \sim LGU(\mu,\sigma)$ with support on the (0,1) interval. Define $Z = \frac{\log\left(\frac{y}{1-y}\right)-\mu}{\sigma}$. Then, the random variable $Z \sim \text{Gumbel}(\mu,\sigma)$ with $\mu = 0$ and $\sigma = 1$.

Proof. Consider the transformation $Z = \frac{\log\left(\frac{y}{1-y}\right)-\mu}{\sigma}$. Then, the pdf of Z is given by:

$$
\begin{aligned}
f_Z(z) &= f_Y\left(\frac{\exp\{z\sigma+\mu\}}{1+\exp\{z\sigma+\mu\}}\right)\left|\frac{\partial\left(\frac{\exp\{z\sigma+\mu\}}{1+\exp\{z\sigma+\mu\}}\right)}{\partial z}\right| \\
&= \frac{1}{\sigma}\exp\left\{\left(\frac{\log\left(\frac{(1+\exp(z\sigma+\mu))\exp(z\sigma+\mu)}{(1+\exp(z\sigma+\mu))}\right)-\mu}{\sigma}\right)-\exp\left[\left(\frac{\log\left(\frac{(1+\exp(z\sigma+\mu))\exp(z\sigma+\mu)}{(1+\exp(z\sigma+\mu))}\right)-\mu}{\sigma}\right)\right]\right\} \\
&\times\frac{1}{\left(\frac{\exp\{z\sigma+\mu\}}{1+\exp\{z\sigma+\mu\}}\right)}\frac{1}{\left[1-\left(\frac{\exp\{z\sigma+\mu\}}{1+\exp\{z\sigma+\mu\}}\right)\right]}\frac{\exp\{z\sigma+\mu\}\sigma}{(1+\exp\{z\sigma+\mu\})^2} \\
&= \exp\left\{\frac{z\sigma+\mu-\mu}{\sigma}-\exp\left[\frac{z\sigma+\mu-\mu}{\sigma}\right]\right\} = \exp\{z-\exp[z]\},
\end{aligned}
\tag{2.13}
$$

which is the corresponding pdf of the Gumbel distribution with $\mu = 0$ and $\sigma = 1$. □

Proposition 2. Let Y be a random variable such that $Y \sim LRG(\mu,\sigma)$ with support on (0,1) interval. Define $Z = \frac{\log\left(\frac{y}{1-y}\right)-\mu}{\sigma}$. Then, the random variable $Z \sim \text{ReverseGumbel}(\mu,\sigma)$ with $\mu = 0$ and $\sigma = 1$.

Proof. Consider the transformation $Z = \frac{\log\left(\frac{y}{1-y}\right)-\mu}{\sigma}$. Then the pdf of Z is given by:

$$
\begin{aligned}
f_Z(z) &= f_Y\left(\frac{\exp\{z\sigma+\mu\}}{1+\exp\{z\sigma+\mu\}}\right)\left|\frac{\partial\left(\frac{\exp\{z\sigma+\mu\}}{1+\exp\{z\sigma+\mu\}}\right)}{\partial z}\right| \\
&= \frac{1}{\sigma}\exp\left\{-\left(\frac{\log\left(\exp(z\sigma+\mu)-\mu\right.}{\sigma}\right)-\exp\left[-\left(\frac{\log\left(\exp(z\sigma+\mu)\right)-\mu}{\sigma}\right)\right]\right\} \\
&\times\frac{1}{\left(\frac{\exp\{z\sigma+\mu\}}{1+\exp\{z\sigma+\mu\}}\right)}\frac{1}{\left[1-\left(\frac{\exp\{z\sigma+\mu\}}{1+\exp\{z\sigma+\mu\}}\right)\right]}\frac{\exp\{z\sigma+\mu\}\sigma}{(1+\exp\{z\sigma+\mu\})^2} \\
&= \exp\left\{\frac{-z\sigma}{\sigma}-\exp\left[\frac{-z\sigma}{\sigma}\right]\right\} = \exp\{-z-\exp[-z]\},
\end{aligned}
\tag{2.14}
$$

which is the correspondent pdf of the reverse Gumbel distribution with $\mu = 0$ and $\sigma = 1$. □

Proposition 3. Let Y be a random variable such that $Y \sim LSN(\mu,\sigma,\nu)$ with support on (0,1) interval. Define $Z = \frac{\log\left(\frac{y}{1-y}\right)-\mu}{\sigma}$. Then, the random variable $Z \sim \mathrm{N}(0,1)$ for $\nu = 0$.

Proof. Consider the transformation $Z = \frac{\log\left(\frac{y}{1-y}\right)-\mu}{\sigma}$. Then the pdf of Z is given by:

$$
\begin{aligned}
f_Z(z) &= f_Y\left(\frac{\exp\{z\sigma+\mu\}}{1+\exp\{z\sigma+\mu\}}\right)\left|\frac{\partial\left(\frac{\exp\{z\sigma+\mu\}}{1+\exp\{z\sigma+\mu\}}\right)}{\partial z}\right| \\
&= \frac{2}{\sigma}\phi\left(\frac{u\sigma+\mu-\mu}{\sigma}\right)\Phi\left(\nu\left(\frac{u\sigma+\mu-\mu}{\sigma}\right)\right) \\
&\times\frac{1}{\left(\frac{\exp\{z\sigma+\mu\}}{1+\exp\{z\sigma+\mu\}}\right)}\frac{1}{\left[1-\left(\frac{\exp\{z\sigma+\mu\}}{1+\exp\{z\sigma+\mu\}}\right)\right]}\frac{\exp\{z\sigma+\mu\}\sigma}{(1+\exp\{z\sigma+\mu\})^2} \\
&= 2\phi(z)\Phi(\nu z) = \phi(z),
\end{aligned}
\tag{2.15}
$$

with $\nu = 0$, which is the correspondent pdf of the standard normal distribution. □

Proposition 4. Let Y be a random variable such that $Y \sim LPE(\mu,\sigma,\nu)$ with support on (0,1) interval. Define $Z = \frac{\log\left(\frac{y}{1-y}\right)-\mu}{\sigma}$. Then, the random variable $Z \sim \mathrm{N}(0,1)$ for $\nu = 2$.

Proof. Consider the transformation $Z = \frac{\log\left(\frac{y}{1-y}\right)-\mu}{\sigma}$. Then the pdf of Z is given by:

$$\begin{aligned} f_Z(z) &= f_Y\left(\frac{\exp\{z\sigma+\mu\}}{1+\exp\{z\sigma+\mu\}}\right)\left|\frac{\partial\left(\frac{\exp\{z\sigma+\mu\}}{1+\exp\{z\sigma+\mu\}}\right)}{\partial z}\right| \\ &= \frac{\nu\exp\left[-\left|\frac{z\sigma+\mu-\mu}{c\sigma}\right|^{\nu}\right]}{2c\sigma\Gamma\left(\frac{1}{\nu}\right)}\frac{1}{\left(\frac{\exp\{z\sigma+\mu\}}{1+\exp\{z\sigma+\mu\}}\right)}\frac{1}{\left[1-\left(\frac{\exp\{z\sigma+\mu\}}{1+\exp\{u\sigma+\mu\}}\right)\right]}\frac{\exp\{z\sigma+\mu\}\sigma}{(1+\exp\{z\sigma+\mu\})^2} \\ &= \frac{\nu\exp\left[-\left|\frac{z}{c}\right|^2\right]}{2c\Gamma\left(\frac{1}{\nu}\right)} = \frac{2\exp\left[-\left(\frac{z}{c}\right)^2\right]}{2c\sqrt{\pi}} = \frac{\exp\left[-\left(\frac{z}{2}\right)^2\right]}{\sqrt{2\pi}}, \end{aligned} \tag{2.16}$$

with $\nu = 2$, which is the correspondent pdf of the Standard Normal. □

Proposition 5. Let Y be a random variable such that $Y \sim LGU(\mu,\sigma)$ with support on (0,1) interval and median $m = Q_Y(p)$, where $p = 0.5$. If $Y \sim LGU(\mu,\sigma)$ with location parameter $\mu = \log\left(\frac{m}{1-m}\right) - \sigma\log(-\log(0.5))$, then the pdf of the alternative parameterization of the logistic Gumbel distribution, denoted by $Y \sim LGU(m,\sigma)$, is given by:

$$\begin{aligned} &f(y;m,\sigma) = \\ &\frac{1}{\sigma}\exp\left\{\left(\frac{\log\left(\frac{y}{1-y}\right)-\log\left(\frac{m}{1-m}\right)+\sigma\log(-\log(0.5))}{\sigma}\right)-\exp\left(\frac{\log\left(\frac{y}{1-y}\right)-\log\left(\frac{m}{1-m}\right)+\sigma\log(-\log(0.5))}{\sigma}\right)\right\} \\ &\times\frac{1}{y(1-y)}. \end{aligned} \tag{2.17}$$

Proof. The median of LGU distribution is given by:

$$m = \frac{\exp\{\mu+\sigma\log(-\log(0.5))\}}{1+\exp\{\mu+\sigma\log(-\log(0.5))\}}. \tag{2.18}$$

Then

$$\mu = \log\left(\frac{m}{1-m}\right) - \sigma\log(-\log(0.5)). \tag{2.19}$$

□

This property is especially relevant since a quantile model can be considered assuming that $Y \sim LGU(m,\sigma)$. So, the contribution from this property is the formulation of a new quantile model.

Proposition 6. Let Y be a random variable such that $Y \sim LRG(\mu,\sigma)$ with support on (0,1) interval and median $m = Q_Y(p)$, where $p = 0.5$. If $Y \sim LRG(\mu,\sigma)$ with location parameter $\mu = \log\left(\frac{m}{1-m}\right) + \sigma\log(-\log(0.5))$, then the pdf of the alternative parameterization of logistic reverse Gumbel distribution, denoted by $Y \sim LRG(m,\sigma)$, is given by:

$$f(y;m,\sigma) = \frac{1}{\sigma}\exp\left\{-\left(\frac{\log\left(\frac{y}{1-y}\right)-\log\left(\frac{m}{1-m}\right)-\sigma\log(-\log(0.5))}{\sigma}\right)-\exp\left[-\left(\frac{\log\left(\frac{y}{1-y}\right)-\log\left(\frac{m}{1-m}\right)-\sigma\log(-\log(0.5))}{\sigma}\right)\right]\right\} \times \frac{1}{y(1-y)}. \tag{2.20}$$

Proof. The median of *LRG* is given by:

$$m = \frac{\exp\{\mu-\sigma\log(-\log(0.5))\}}{1+\exp\{\mu-\sigma\log(-\log(0.5))\}}. \tag{2.21}$$

Then

$$\mu = \log\left(\frac{m}{1-m}\right)+\sigma\log(-\log(0.5)). \tag{2.22}$$

□

Proposition 7. The extension of the JSB distributions to the random variable Z with support on bounded interval (c,d) have the pdf $f_Z(z;\boldsymbol{\theta})$ defined by:

$$f_Z(z;\boldsymbol{\theta}) = f_Y\left(\frac{z-c}{d-c};\boldsymbol{\theta}\right) = g\left(\log\left(\frac{z-c}{d-z}\right);\boldsymbol{\theta}\right)\frac{1}{(z-c)(d-z)} \tag{2.23}$$

Proof. Consider the transformation $Z = Y(d-c)+c, c,d \in \mathbb{R}$ such that $Y \sim JSB(\boldsymbol{\theta})$ and

$$f(y) = g\left(\log\left(\frac{y}{1-y}\right);\theta\right)\frac{1}{y(1-y)},$$

we can obtain the correspondent pdf $f_Z(z;\boldsymbol{\theta})$ of the random variable Z. □

Proposition 8. Let Y be a random variable such that $Y \sim LEG(\mu,\sigma,\nu)$ with support on (0,1). An alternative parameterization of LEG with location parameter $0 < m < 1$, denoted by $Y \sim LEG(m,\sigma,\nu)$ is given by :

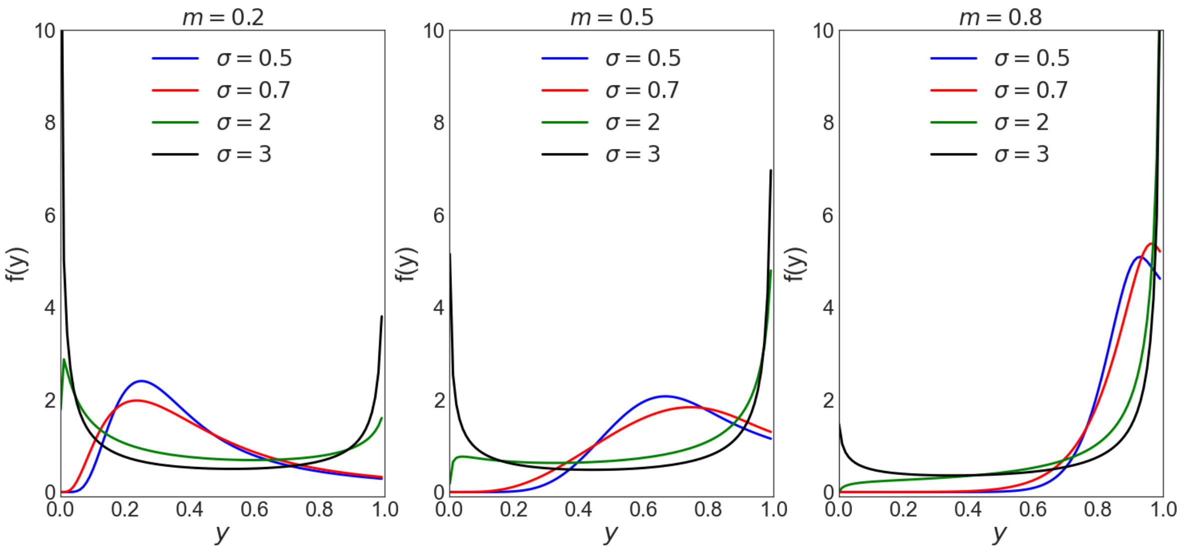
$$f(y;m,\sigma,\nu) = \frac{1}{\nu}\exp\left\{\frac{\log\left(\frac{m}{1-m}\right)-\log\left(\frac{y}{1-y}\right)}{\nu}+\frac{\sigma^2}{2\nu^2}\right\}\Phi\left(\frac{\log\left(\frac{y}{1-y}\right)-\log\left(\frac{m}{1-m}\right)}{\sigma}-\frac{\sigma}{\nu}\right) \times \frac{1}{y(1-y)}, \tag{2.24}$$

for $0 < y < 1$, $0 < m < 1$, $\sigma > 0$ and $\nu > 0$.

Proof. Considering the reparameterization $m = \frac{\exp(\mu)}{1+\exp(\mu)}$, $m \in (0,1)$, we obtain the pdf of *LEG* with parameter vector (m, σ, ν). □

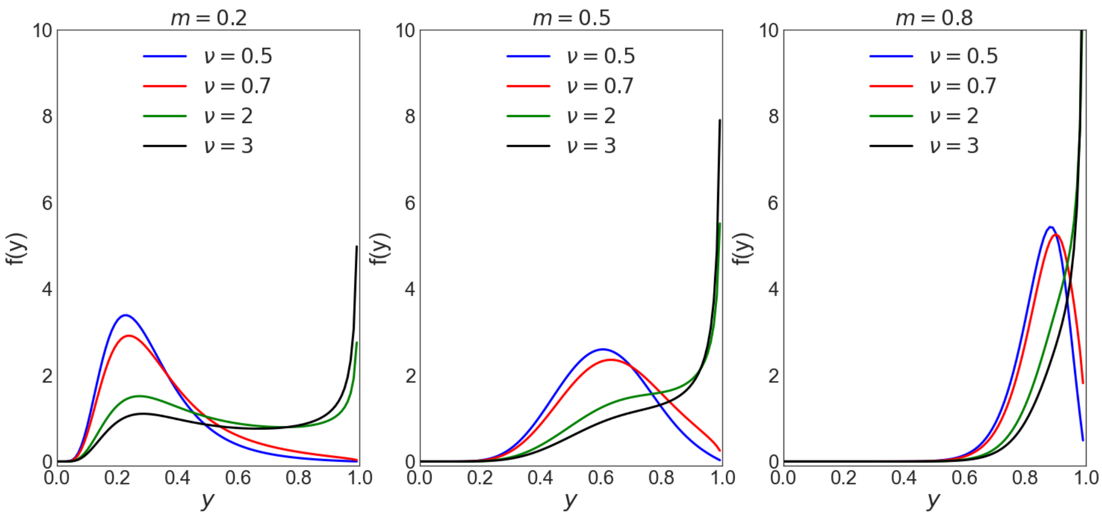
The Figure 23 and 24 show the density curve for different values of the location parameter m. Like the μ parameter, the location of the density change as the m values change.

Figure 23 – Logistic exponential Gaussian probability density function for parameter m = 0.2, 0.5 and 0.8, σ = 0.5 and some choices of parameter ν.



Source: Elaborated by the author.

Figure 24 – Logistic exponential Gaussian probability density function for parameter ν = 0.5, 0.7 and 2, μ = 0 and some choices of parameter σ.



Source: Elaborated by the author.

Proposition 9. Let Y be a random variable such that $Y \sim LSN(\mu, \sigma, \nu)$ with support on (0,1). An alternative parameterization of LSN with location parameter $0 < m < 1$, denoted by $Y \sim LSN(m, \sigma, \nu)$ is given by:

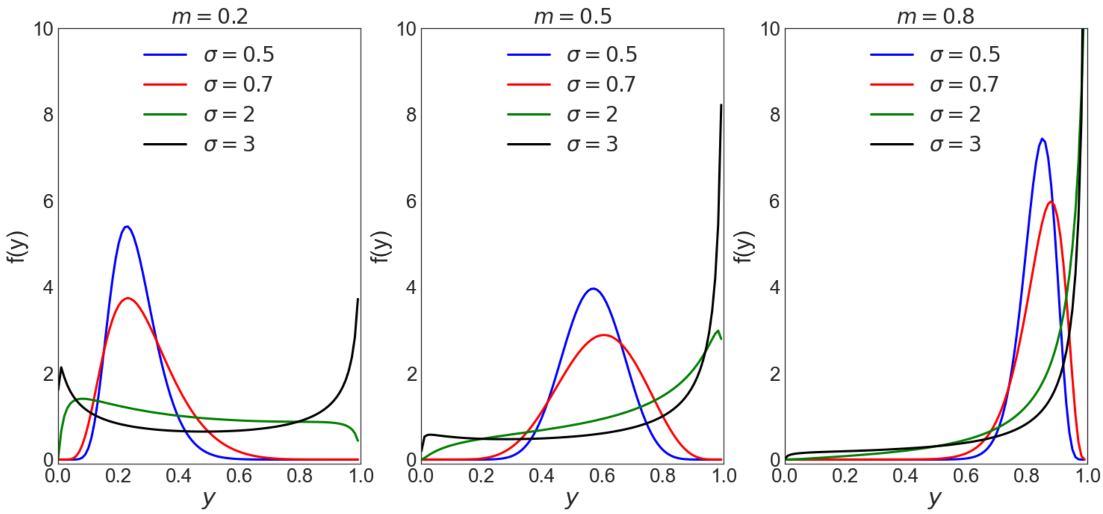
$$f(y; m, \sigma, \nu) = \frac{2}{\sigma}\phi(z)\Phi(\nu z)\frac{1}{y(1-y)}, \tag{2.25}$$

for $0 < y < 1$ with $z = \left(\log\left(\frac{y}{1-y}\right) - \log\left(\frac{\mu}{1-\mu}\right)\right)/\sigma$, $0 < m < 1$, $\sigma > 0$ and $-\infty < \nu < \infty$.

Proof. Considering the reparameterization $m = \frac{\exp(\mu)}{1+\exp(\mu)}$, $m \in (0,1)$ we obtain the pdf of *LSN* with parameter vector (m, σ, ν). □

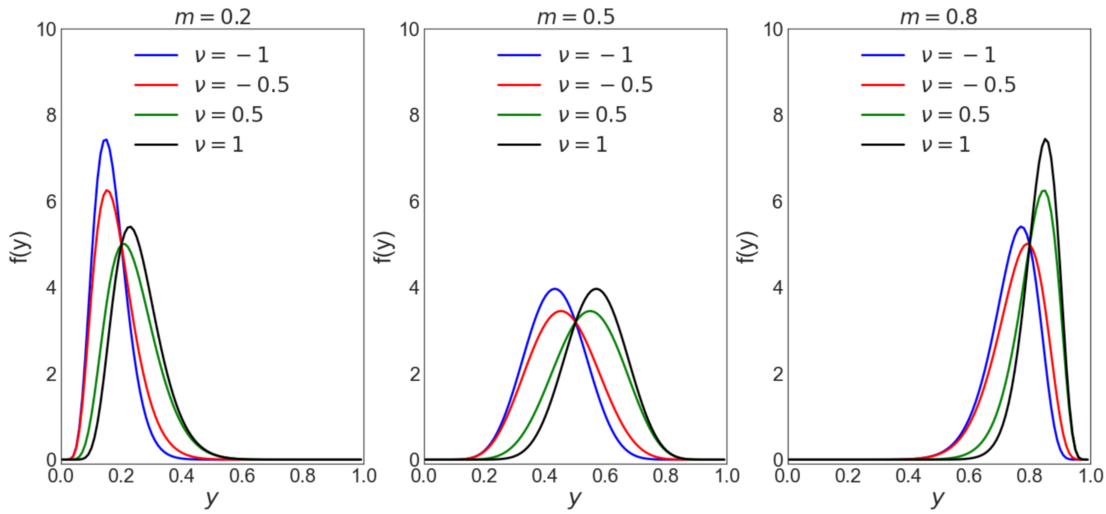
The Figure 25 and 24 show the density curve for different values of the location parameter m. The visualization of location density is better for the location parameter m in the $(0,1)$ interval.

Figure 25 – Logistic skew normal probability density function for m = 0.2, 0, 0.8, ν = 1 and some choices of parameter σ.



Source: Elaborated by the author.

Figure 26 – Logistic skew normal probability density function for m = 0.2, 0.5, 0.8 σ = 0.5 and some choices of parameter ν.



Source: Elaborated by the author.

Proposition 10. Let Y be a random variable such that $Y \sim LPE(\mu,\sigma,\nu)$ with support on (0,1). An alternative parameterization of LPE with location parameter $0 < m < 1$, denoted by $Y \sim LPE(m,\sigma,\nu)$ is given by:

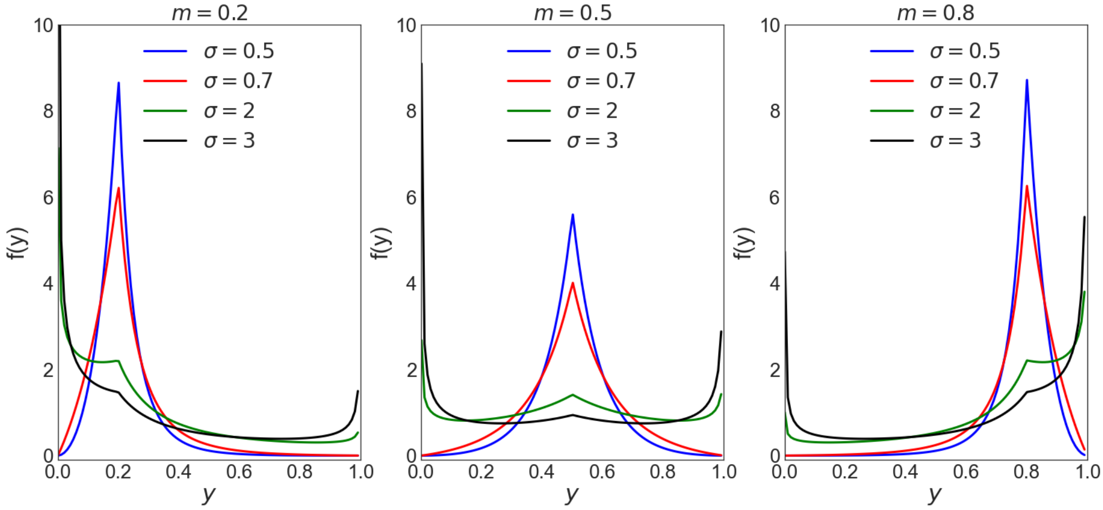
$$f(y;m,\sigma,\nu) = \frac{\nu \exp(-|z|^{\nu})}{2c\sigma\Gamma(1/\nu)} \frac{1}{y(1-y)}, \tag{2.26}$$

for $0 < y < 1, 0 < m < 1, \sigma > 0,\ \nu > 0,\ z = \frac{\log\left(\frac{y}{1-y}\right) - \log\left(\frac{m}{1-m}\right)}{(c\sigma)}$ and $c^2 = \Gamma(1/\nu)[\Gamma(3/\nu)^{-1}]$.

Proof. Considering the reparameterization $m = \frac{\exp(\mu)}{1+\exp(\mu)}$, $m \in (0,1)$ we obtain the pdf of LPE with parameter vector (m,σ,ν). □

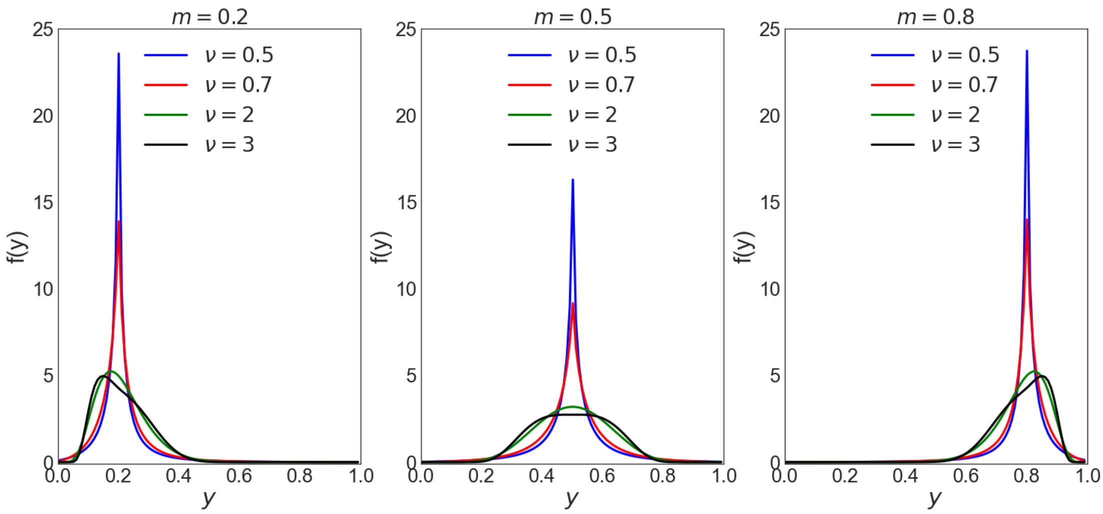
The Figure 27 and 28 show the density curve for different values of the location parameter m.

Figure 27 – Logistic power exponential probability density function for $m = 0.2$, 0.5 and 0.8, $\nu = 1$ and some choices of parameter σ.



Source: Elaborated by the author.

Figure 28 – Logistic power exponential probability density function for $m = 0.2$, 0.5 and 0.8, $\sigma = 0.5$ and some choices of parameter ν.



Source: Elaborated by the author.

Proposition 11. If $Y \sim LSN(\mu,\sigma,\nu)$, then for $\nu = 0$ the distribution Y has logistic normal distribution (JOHNSON, 1949) as a special case.

Proof. The logistic skew normal distribution includes the logistic normal distribution as a special case when $\nu = 0$.

$$f(y;\mu,\sigma) = \frac{1}{\sqrt{2\pi}\sigma}\frac{1}{y(1-y)}\exp\left\{-\frac{1}{2}\left(\frac{\log\left(\frac{y}{1-y}\right)-\mu}{\sigma}\right)^2\right\}. \tag{2.27}$$

□

Proposition 12. If $Y \sim LPE(\mu, \sigma, \nu)$, then for $\nu = 2$ the distribution Y has the logistic normal distribution as a special case.

Proof. The logistic power exponential includes the logistic normal distribution as a special case when $\nu = 2$.

$$f(y;\mu,\sigma) = \frac{1}{\sqrt{2\pi}\sigma}\frac{1}{y(1-y)}\exp\left\{-\frac{1}{2}\left(\frac{\log\left(\frac{y}{1-y}\right)-\mu}{\sigma}\right)^2\right\}. \tag{2.28}$$

□

Proposition 13. The *LSN* density is symmetric when $\mu = 0$ and $\nu = 0$ for all values of σ.

Proof. To show that the pdf of Y is symmetric when $\mu = 0$ and for any value of σ it is sufficient show that $f(0.5 - y) = f(0.5 + y)$

$$\begin{aligned} f(y+0.5) &= \frac{1}{\sqrt{2\pi}\sigma}\exp\left\{-\frac{1}{2}\left(\frac{\log(y+0.5)-\log(0.5-y)}{\sigma}\right)^2\right\}\frac{1}{(y+0.5)}\frac{1}{(0.5-y)} \\ &= \frac{1}{\sqrt{2\pi}\sigma}\exp\left\{-\frac{1}{2}\left(\frac{-\log(y+0.5)+\log(0.5-y)}{\sigma}\right)^2\right\}\frac{1}{(y+0.5)}\frac{1}{(0.5-y)} \\ &= f(0.5-y) \end{aligned} \tag{2.29}$$

□

Proposition 14. The *LPE* density is symmetric when $\mu = 0$ for all values of σ.

Proof. To show that the pdf of Y is symmetric when $\mu = 0$ and for any value of σ it is sufficient show that $f(0.5 - y) = f(0.5 + y)$

$$\begin{aligned} f(y+0.5) &= \frac{\nu\exp\left[-\left|\frac{\log(y+0.5)-\log(0.5-y)}{c\sigma}\right|^{\nu}\right]}{2\sigma c\Gamma(1/\nu)}\frac{1}{(y+0.5)(0.5-y)} \\ &= \frac{\nu\exp\left[-\left|\frac{\log(0.5-y)-\log(y+0.5)}{c\sigma}\right|^{\nu}\right]}{2\sigma c\Gamma(1/\nu)}\frac{1}{(0.5-y)(y+0.5)} \\ &= f(0.5-y) \end{aligned} \tag{2.30}$$

□

The Table 1 shows a summary of the properties. All distributions can be extended to a bounded interval (c, d). The LRG and LGU supports a quantile parameterization. This parametrization enables a quantile regression model. Only the LSN and LPE are symmetrical for some parameters.

Table 1 – A summary of the properties.

	LGU	LRG	LEG	LSN	LPE
Reversibility	✓	✓		✓	✓
Quantile parameterization	✓	✓			
Transformation in location parameter			✓	✓	✓
Extension to a bounded interval (c,d)	✓	✓	✓	✓	✓
Particular cases				✓	✓
Symmetry				✓	✓

2.3 Mode and Moments

The mode and the moments of the distributions from this chapter do not have a closed form. The mode and moments of the distributions were computed numerically.

All baseline distributions are unimodal, see Johnson, Kotz and Balakrishnan (1995) for LGU and LRG, Jr and Kutner (1976) for LEG, Azzalini (1985) for LSN, and Nelson (1991) for LPE. Only one mode was found for each *JSB* distribution by the numerical method presented by Nelder and Mead (1965) available in the `stats` package R Core Team (2020) with the function *optim*.

Definition 2. The mode y_0 of the JBS distribution is the solution of the equation

$$\frac{\partial f(y;\mu,\sigma,\nu)}{\partial y} = 0. \tag{2.31}$$

Definition 3. If $Y \sim JSB(\boldsymbol{\theta})$, then the moments of Y about zero are given by:

$$E(Y^k) = \int_0^1 y^k f_Y(y)dy, \tag{2.32}$$

that can be written as:

$$E(Y^k) = \int_{-\infty}^{\infty} \left(\frac{\exp(x)}{1+\exp(x)}\right) f_X(x)dx. \tag{2.33}$$

The moments from the *JSB* distributions do not have a closed form. The expression 2.32 have been computed numerically.

2.3.1 Logistic Gumbel

Table 3 displays that E(Y) evaluated by the Monte Carlo integration. Unfortunately, the E(Y) does not have an analytical form. It is possible to observe a relation between the parameters and the E(Y) by numerical values. For example, the E(Y) increases as μ increases, showing that the location parameter impacts the E(Y) of the distribution. Despite not having a big impact on the E(Y), the dispersion parameter correlates with Var(Y). Table 3 shows the E(Y) and Var(Y) for some parameters, and we can take only simple conclusions from how μ and σ impact on E(Y) and Var(Y). Table 2 reveals that for small values of σ, like 0.5 and 0.7, the mode of the LGU is similar to the E(Y).

Table 2 – Mode of the logistic Gumbel calculated by the function *optim* from the R language for different values of μ and σ.

μ	σ	y_0	μ	σ	y_0	μ	σ	y_0
	0.5	0.15273		0.5	0.5		0.5	0.83831
-1.5	0.7	0.14801	0	0.7	0.50001	1.5	0.7	0.85609
	2	0.00006		2	0.00006		2	0.97408

Source: Elaborated by the author.

Table 3 – E(Y) and Var(Y) evaluated by the Monte Carlo integration of the logistic Gumbel distribution for some choices of μ and σ.

	$\mu = -1.5$		
	$\sigma = 0.5$	$\sigma = 0.7$	$\sigma = 2$
E(Y)	0,1583	0,1576	0,1894
Var(Y)	0,0047	0,0082	0,0470
	$\mu = 0$		
	$\sigma = 0.5$	$\sigma = 0.7$	$\sigma = 2$
E(Y)	0,4392	0,4215	0,3765
Var(Y)	0,0186	0,0304	0,0960
	$\mu = 1.5$		
	$\sigma = 0.5$	$\sigma = 0.7$	$\sigma = 2$
E(Y)	0,7402	0,7135	0,5864
Var(Y)	0,0253	0,0367	0,1140

Source: Elaborated by the author.

2.3.2 Logistic reverse Gumbel

The logistic reverse Gumbel is similar to the logistic Gumbel, and we can see from Table 5 a similar behavior of E(Y) and Var(Y) from the logistic Gumbel distribution concerto to the parameters. So, to look just at the numerical values from the 5 shows that the location parameter impacts the E(Y) of the distribution. The E(Y) increases as μ increases, and Var(Y) increases as σ increases. Table 4 displays that the LRG mode is not similar to the E(Y).

Table 4 – Mode of the logistic reverse Gumbel calculated by the function *optim* from the R language for different values of μ and σ.

μ	σ	y_0	μ	σ	y_0	μ	σ	y_0
	0.5	0.16168		0.5	0.49999		0.5	0.84726
-1.5	0.7	0.14391	0	0.7	0.49999	1.5	0.7	0.88519
	2	0.02591		2	0.99993		2	0.99993

Source: Elaborated by the author.

Table 5 – E(Y) and Var(Y) evaluated by the Monte Carlo integration of the logistic reverse Gumbel distribution for some choices of μ and σ.

	$\mu = -1.5$		
	$\sigma = 0.5$	$\sigma = 0.7$	$\sigma = 2$
E(Y)	0.2427	0.2722	0.4256
Var(Y)	0.0156	0.0299	0.1159
	$\mu = 0$		
	$\sigma = 0.5$	$\sigma = 0.7$	$\sigma = 2$
E(Y)	0,5604	0.5772	0.6434
Var(Y)	0.0203	0.0319	0.0911
	$\mu = 1.5$		
	$\sigma = 0.5$	$\sigma = 0.7$	$\sigma = 2$
E(Y)	0,8361	0,8420	0,8575
Var(Y)	0,0091	0,0087	0.0162

Source: Elaborated by the author.

2.3.3 Logistic exponential Gaussian

The logistic exponential Gaussian has one more parameter, and it does not have an analytical form for E(Y) and Var(Y). To make conclusions from the relation between the E(Y) and Var(Y) with the parameters is more laborious. Table 7 shows the $E(Y)$ for tree divergent values of ν. Like the previously distributions, we can take only simple conclusions. The numerical values show a impact of μ and ν at $E(Y)$, that is, the E(Y) increases as both parameters increase. But, we are not able to delve into this relation just by looking at the numerical values. Table 6 shows the numerical values for the mode. To look at them, we can see that the mode increases as ν increases. It is an interesting observation since the ν can be seen as a shape parameter.

Table 6 – Mode of the logistic exponential Gaussian calculated by the function *optim* from the R language for different values of μ and σ and $\nu = 1$.

$\nu = 0.5$								
μ	σ	y_0	μ	σ	y_0	μ	σ	y_0
	0.5	0.20543		0.5	0.60477		0.5	0.89626
-1.5	0.7	0.18385	0	0.7	0.63833	1.5	0.7	0.92181
	2	0.00598		2	0.99		2	0.99
$\nu = 1$								
μ	σ	y_0	μ	σ	y_0	μ	σ	y_0
	0.5	0.22596		0.5	0.66601		0.5	0.93837
-1.5	0.7	0.20973	0	0.7	0.74336	1.5	0.7	0.97020
	2	0.00690		2	0.99993		2	0.99993
$\nu = 2$								
μ	σ	y_0	μ	σ	y_0	μ	σ	y_0
	0.5	0.24602		0.5	0.99		0.5	0.99
-1.5	0.7	0.23798	0	0.7	0.99	1.5	0.7	0.99
	2	0.00806		2	0.99		2	0.99

Source: Elaborated by the author.

Table 7 – E(Y) and Var(Y) evaluated by the Monte Carlo integration of the logistic exponential Gaussian distribution for some choices of μ, σ and ν.

$\nu = 0.5$									
	$\mu = -1.5$			$\mu = 0$			$\mu = 1.5$		
	$\sigma = 0.5$	$\sigma = 0.7$	$\sigma = 2$	$\sigma = 0.5$	$\sigma = 0.7$	$\sigma = 2$	$\sigma = 0.5$	$\sigma = 0.7$	$\sigma = 2$
E(Y)	0.2839	0.2932	0.3547	0.6122	0.6073	0.5758	0.8674	0.8592	0.7762
Var(Y)	0.0202	0.0274	0,0910	0.0193	0.0291	0.0978	0.0015	0.0052	0.0613
$\nu = 1$									
	$\mu = -1.5$			$\mu = 0$			$\mu = 1.5$		
	$\sigma = 0.5$	$\sigma = 0.7$	$\sigma = 2$	$\sigma = 0.5$	$\sigma = 0.7$	$\sigma = 2$	$\sigma = 0.5$	$\sigma = 0.7$	$\sigma = 2$
E(Y)	0.3844	0.3895	0.4255	0.6908	0.6848	0.6386	0.9007	0.8937	0.8183
Var(Y)	0.0260	0.0341	0.0955	0.0462	0.0523	0.1055	0.0005	0.0035	0.0516
$\nu = 2$									
	$\mu = -1.5$			$\mu = 0$			$\mu = 1.5$		
	$\sigma = 0.5$	$\sigma = 0.7$	$\sigma = 2$	$\sigma = 0.5$	$\sigma = 0.7$	$\sigma = 2$	$\sigma = 0.5$	$\sigma = 0.7$	$\sigma = 2$
E(Y)	0.5344	0.5355	0.5388	0.7843	0.7785	0.7271	0.9266	0.9210	0.8579
Var(Y)	0.0781	0.0824	0.1211	0.0288	0.0349	0.0869	0.0068	0.0093	0.0479

Source: Elaborated by the author.

2.3.4 Logistic skew normal

In the logistic skew normal distribution the ν parameter takes values from $-\infty$ to ∞. So, it is harder then LEG to describe the numerical values of E(Y). Table 9 shows the E(Y) values for three different values of ν. The numerical values show a impact of μ in E(Y). However, the ν has a smaller impact on E(Y) compared to the conclusions from LEG. Table 8 displays that for low values o σ, like 0.5 and 0.7, the mode is similar to $E(Y)$.

Table 8 – Mode of the logistic skew normal calculated by the function *optim* from the R language for different values of μ and σ and $\nu = 1$.

$\nu = -1$								
μ	σ	y_0	μ	σ	y_0	μ	σ	y_0
-1.5	0.5	0.13269	0	0.5	0.43149	1.5	0.5	0.79286
	0.7	0.10735		0.7	0.39597		0.7	0.79085
	2	0.00379		2	0.00185		2	0.93041
$\nu = 0.5$								
μ	σ	y_0	μ	σ	y_0	μ	σ	y_0
-1.5	0.5	0.18809	0	0.5	0.54831	1.5	0.5	0.86186
	0.7	0.17777		0.7	0.57613		0.7	0.88847
	2	0.01607		2	0.98444		2	0.99
$\nu = 1$								
μ	σ	y_0	μ	σ	y_0	μ	σ	y_0
-1.5	0.5	0.207131	0	0.5	0.56850	1.5	0.5	0.867304
	0.7	0.209141		0.7	0.60402		0.7	0.89264
	2	0.06958		2	0.98141		2	0.99621

Source: Elaborated by the author.

Table 9 – E(Y) and Var(Y) evaluated by the Monte Carlo integration of the logistic skew normal distribution for some choices of μ, σ and ν.

ν = -1									
	μ = -1.5			μ = 0			μ = 1.5		
	σ = 0.5	σ = 0.7	σ = 2	σ = 0.5	σ = 0.7	σ = 2	σ = 0.5	σ = 0.7	σ = 2
E(Y)	0.1521	0.1448	0.1357	0.4298	0.4082	0.3188	0.7602	0.7343	0.5654
Var(Y)	0.0026	0.0047	0,0270	0.0104	0.0171	0.0656	0.0083	0.0138	0.0800
ν = 0.5									
	μ = -1.5			μ = 0			μ = 1.5		
	σ = 0.5	σ = 0.7	σ = 2	σ = 0.5	σ = 0.7	σ = 2	σ = 0.5	σ = 0.7	σ = 2
E(Y)	0.1952	0.2045	0.2849	0.4969	0.4978	0.5000	0.8043	0.7963	0.7164
Var(Y)	0.0056	0.0115	0,0756	0.0154	0.0258	0.0986	0.0076	0.0123	0.0753
ν = 1									
	μ = -1.5			μ = 0			μ = 1.5		
	σ = 0.5	σ = 0.7	σ = 2	σ = 0.5	σ = 0.7	σ = 2	σ = 0.5	σ = 0.7	σ = 2
E(Y)	0.2383	0.2642	0.4341	0.5641	0.5875	0.6813	0.8484	0.8583	0.8673
Var(Y)	0.049	0.0110	0,0798	0.0114	0.0185	0.0659	0.0031	0.0079	0.0250

Source: Elaborated by the author.

2.3.5 Logistic power exponential

Table 11 shows a less expressive influence from ν, compared to LEG and LSN, at $E(Y)$. However, like other cases, μ and σ are strictly related to $E(Y)$ and $Var(Y)$ respectively. The 10 reveals that ν does not have a huge impact on the mode of LPE.

Table 10 – Mode of the logistic power exponential calculated by the function *optim* from the R language for different values of μ and σ and $\nu = 0.5$.

$\nu = 0.5$								
μ	σ	y_0	μ	σ	y_0	μ	σ	y_0
	0.5	0.18243		0.5	0.5		0.5	0.81757
-1.5	0.7	0.18244	0	0.7	0.5	1.5	0.7	0.81755
	2	0.18243		2	0.5		2	0.81757
$\nu = 1$								
μ	σ	y_0	μ	σ	y_0	μ	σ	y_0
	0.5	0.18242		0.5	0.5		0.5	0.81758
-1.5	0.7	0.18240	0	0.7	0.5	1.5	0.7	0.81759
	2	0.00006		2	0.5		2	0.99994
$\nu = 2$								
μ	σ	y_0	μ	σ	y_0	μ	σ	y_0
	0.5	0.15831		0.5	0.5		0.5	0.84168
-1.5	0.7	0.13497	0	0.7	0.5	1.5	0.7	0.86503
	2	0.00429		2	0.5		2	0.99

Source: Elaborated by the author.

Table 11 – E(Y) and Var(Y) evaluated by the Monte Carlo integration of the logistic power exponential distribution for some choices of μ, σ and ν.

$\nu = 0.5$									
	μ = -1.5			μ = 0			μ = 1.5		
	σ = 0.5	σ = 0.7	σ = 2	σ = 0.5	σ = 0.7	σ = 2	σ = 0.5	σ = 0.7	σ = 2
E(Y)	0.1915	0.1978	0.2402	0.5009	0.5006	0.5004	0.8137	0.8058	0.7609
Var(Y)	0.0058	0.0105	0.0461	0.0099	0.0165	0.0553	0.0014	0.0074	0.0451
$\nu = 1$									
	μ = -1.5			μ = 0			μ = 1.5		
	σ = 0.5	σ = 0.7	σ = 2	σ = 0.5	σ = 0.7	σ = 2	σ = 0.5	σ = 0.7	σ = 2
E(Y)	0.1960	0.2038	0.2667	0.4980	0.4981	0.4984	0.8043	0.7957	0.7306
Var(Y)	0.005	0.0113	0,0640	0.0140	0.0232	0.0813	0.0078	0.0135	0.0659
$\nu = 2$									
	μ = -1.5			μ = 0			μ = 1.5		
	σ = 0.5	σ = 0.7	σ = 2	σ = 0.5	σ = 0.7	σ = 2	σ = 0.5	σ = 0.7	σ = 2
E(Y)	0.1955	0.2043	0.2847	0.4977	0.4981	0.4992	0.8028	0.7939	0.7132
Var(Y)	0.0054	0.0114	0,0756	0.0153	0.0260	0.0987	0.0084	0.0137	0.0768

Source: Elaborated by the author.

2.4 Kurtosis and Skewness

2.4.1 Logistic Gumbel

Figures 29 and 30 present the behavior of Skewness (third standardized moment) and Kurtosis (fourth standardized moment) for some values of μ and σ. The Skewness is a measure of asymmetry. A distribution is called left-skewed if the Skewness is negative. A distribution is called right-skewed if the Skewness is positive. The line on the value 0 is displayed in the Figures for better visualization. It is common to compare the kurtosis values with value 3 (the kurtosis of the univariate normal distribution). We can see that the logistic Gumbel distribution accommodates moderate negative and positive skew data. The Kurtosis is negative for many scenarios, like a platykurtic distribution with thinner tails. The logistic Gumbel distribution accommodates data with moderately negative Kurtosis data.

Figure 29 – Measure of Kurtosis of the logistic Gumbel distribution in function of dispersion and location parameter.

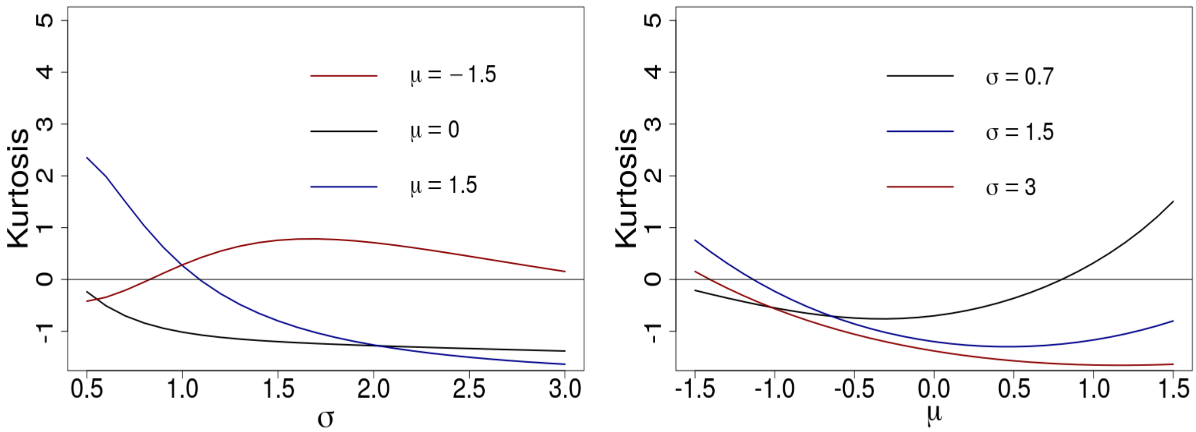
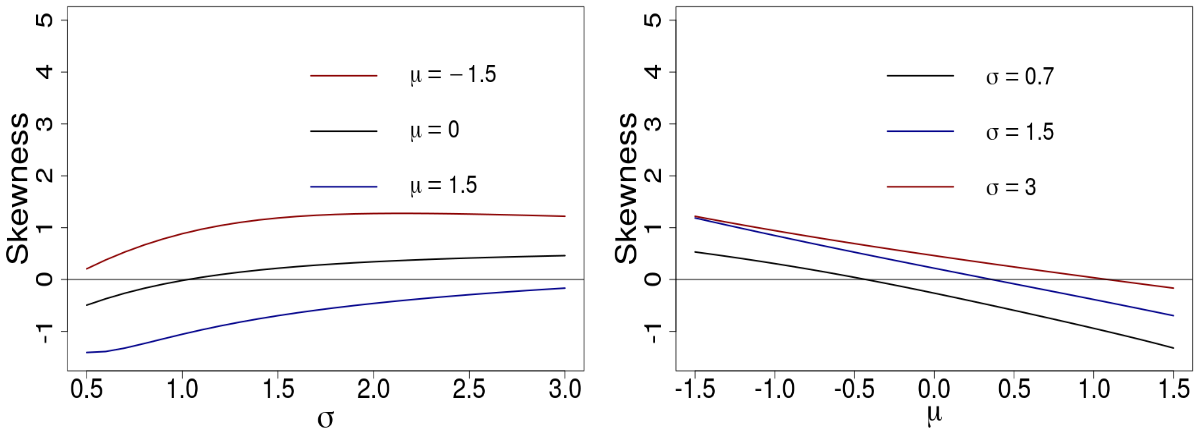


Figure 30 – Measure of Skewness of the logistic Gumbel distribution in function of dispersion and location parameter.



2.4.2 Logistic reverse Gumbel

Figures 31 and 32 present the behavior of Skewness and Kurtosis for some values of μ and σ. We can see a similar interpretation from the logistic Gumbel. The logistic reverse Gumbel accommodates moderate negative and positive skew data and with moderately negative Kurtosis data.

Figure 31 – Measure of Kurtosis of logistic reverse Gumbel distribution in function of dispersion and location parameter.

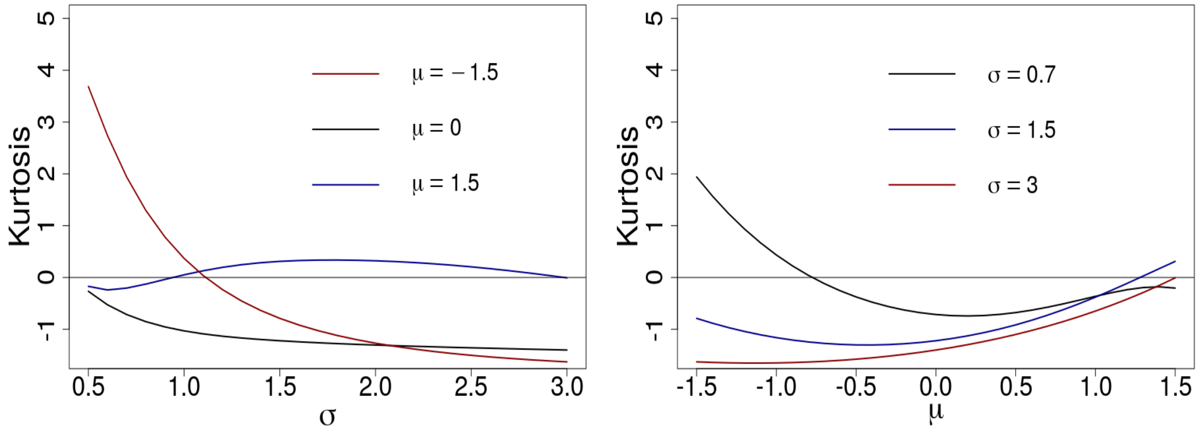
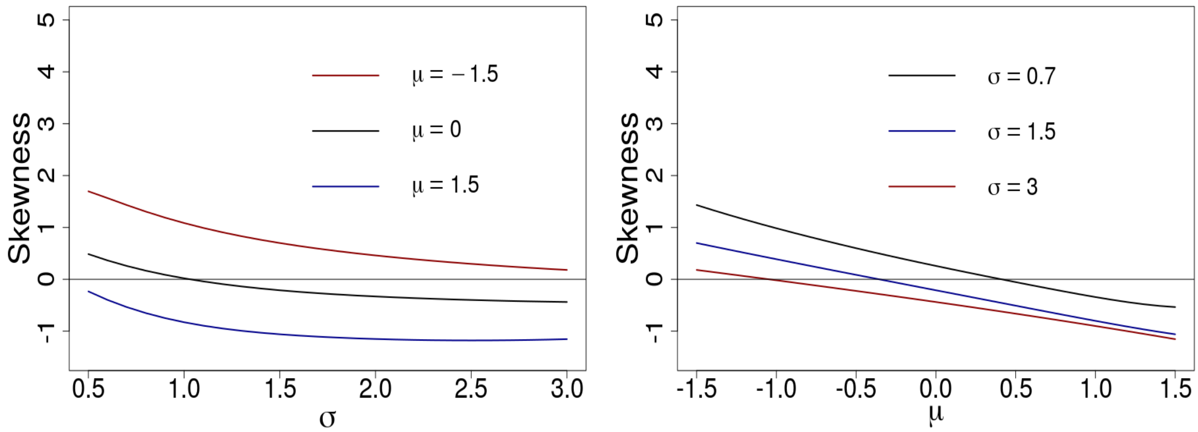


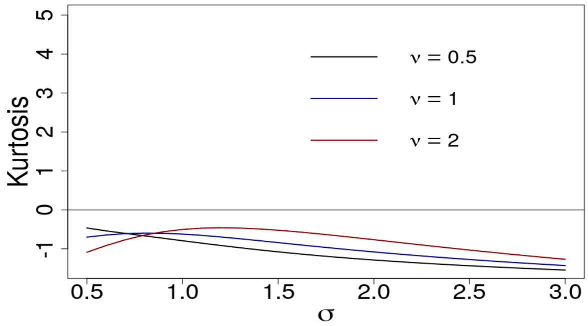
Figure 32 – Measure of Skewness of logistic reverse Gumbel distribution in function of dispersion and location parameter.



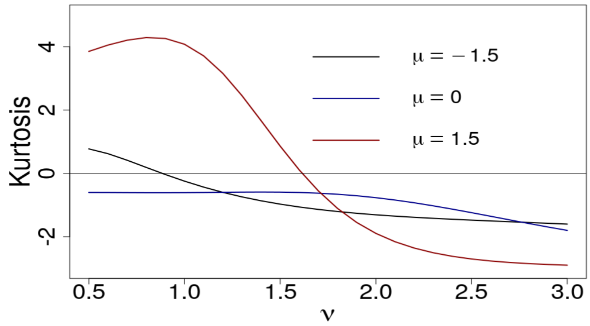
2.4.3 *Logistic exponential Gaussian*

Figure 33 presents the Kurtosis for the logistic exponential Gaussian. The Kurtosis is moderately negative for many scenarios, and it increases with a raise in μ. So the logistic exponential Gaussian has thinner tails but is more flexible than logistic Gumbel and logistic reverse Gumbel for data with positive Kurtosis, principally because of the third parameter ν. Figure 34 shows that the logistic exponential Gaussian accommodates positive and negative skew data.

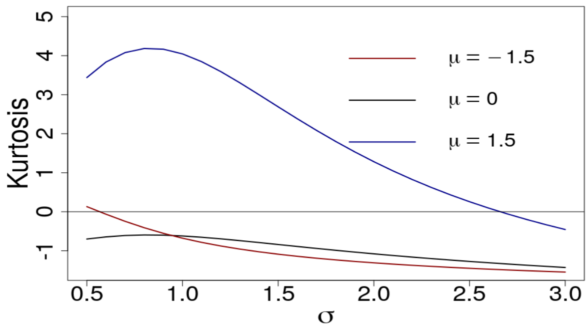
Figure 33 – Measure of Kurtosis of the logistic exponential Gaussian distribution in function of dispersion, location and shape parameter.



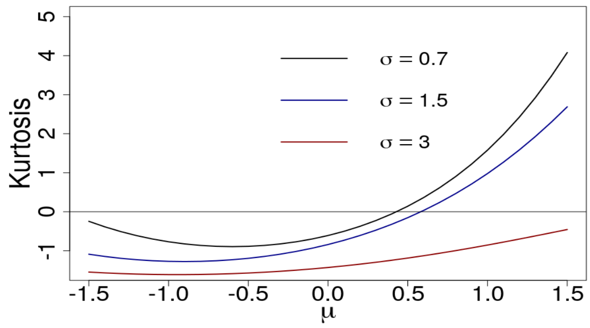
(a) Measure of Kurtosis of the logistic exponential Gaussian distribution for $\mu = 0$.



(b) Measure of Kurtosis of the logistic exponential Gaussian distribution for $\sigma = 0.7$.

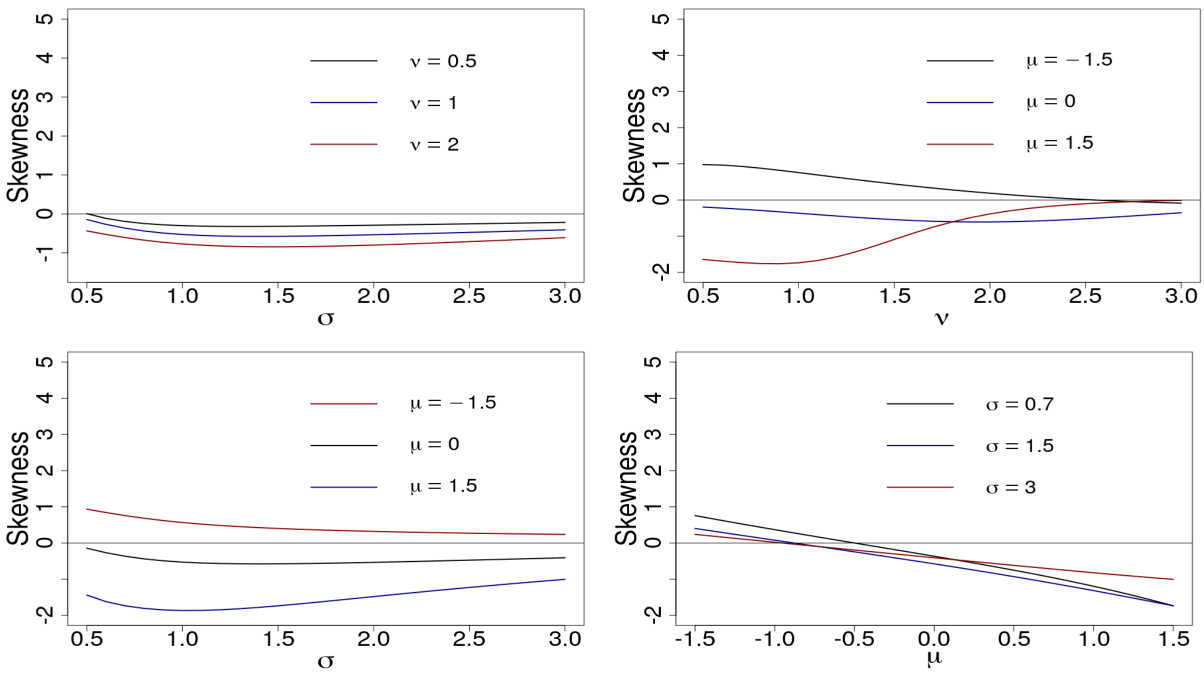


(c) Measure of Kurtosis of the logistic exponential Gaussian distribution for $\nu = 1$.



(d) Measure of Kurtosis of the logistic exponential Gaussian distribution for $\nu = 0$.

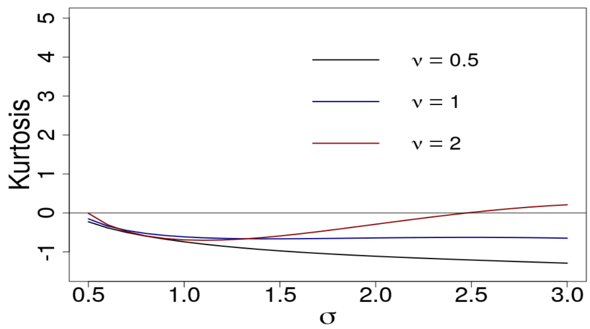
Figure 34 – Measure of Skewness of the logistic exponential Gaussian distribution in function of dispersion and location parameter.



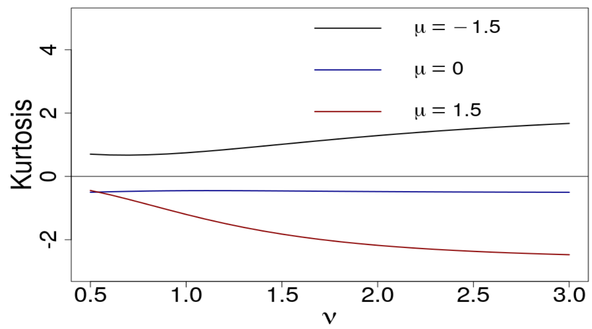
2.4.4 Logistic skew normal type I

Figure 35 present the Kurtosis for the logistic skew normal. The Kurtosis is moderately negative for many scenarios, and it seems that the logistic skew normal is not flexible for high Kurtosis data. However, we can see from Figure 36 that for $\mu = 0$ logistic skew normal is symmetric, so it seems a valuable distribution for moderate symmetric data.

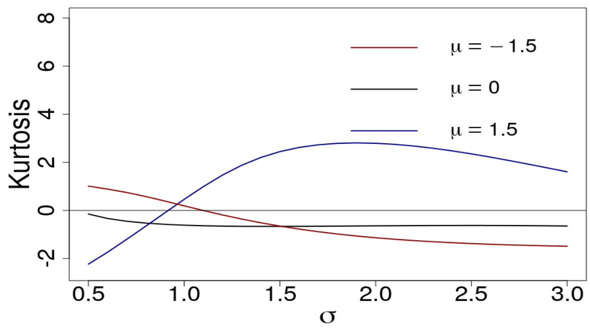
Figure 35 – Measure of Kurtosis of the logistic skew normal distribution in function of dispersion, location and shape parameter.



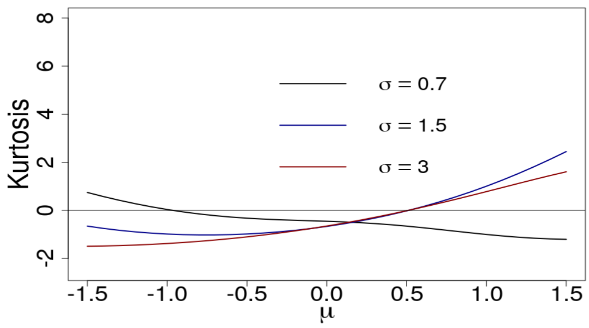
(a) Measure of Kurtosis of the logistic skew normal distribution for $\mu = 0$.



(b) Measure of Kurtosis of the logistic skew normal distribution for $\sigma = 0.7$.

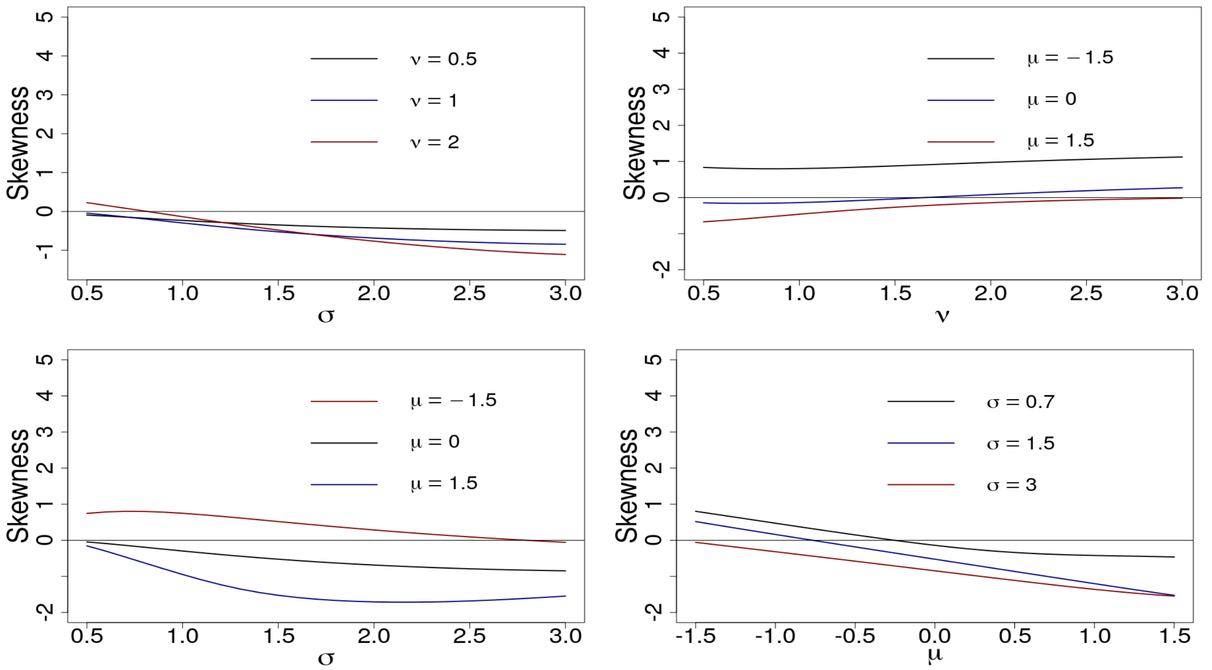


(c) Measure of Kurtosis of the logistic skew normal distribution for $\nu = 1$.



(d) Measure of Kurtosis of the logistic skew normal distribution for $\nu = 0$.

Figure 36 – Measures of Skewness of the logistic skew normal distribution in function of dispersion and location parameter.

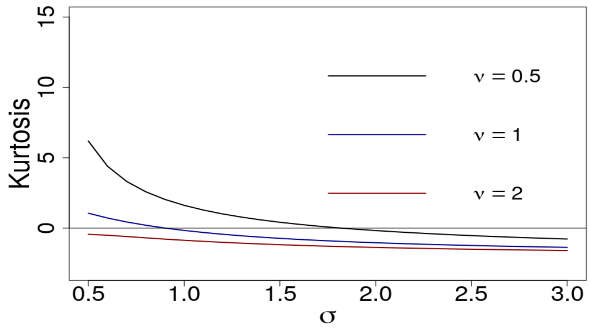


2.4.5 Logistic power exponential

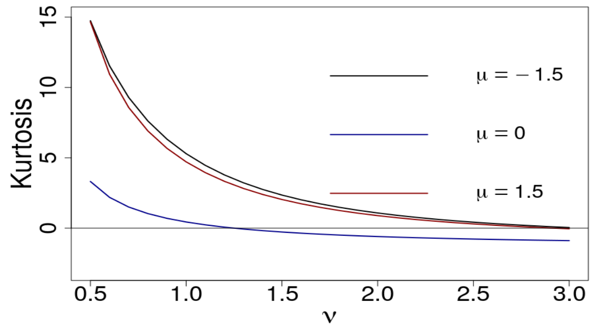
Figure 37 presents the Kurtosis for the LPE distribution. The LPE distribution is the most indicated for data with a high Kurtosis value (from the considered scenarios and distributions, it reached the high value), and just a few scenarios display negative Kurtosis. The distribution presented negative, positive, and close to 0 values of Skewness, so the LPE seems suitable for both positive and negative skew data.

Jointly with the LEG distribution, the is the most adequate for positive skew data, but it is also flexible for left skew data.

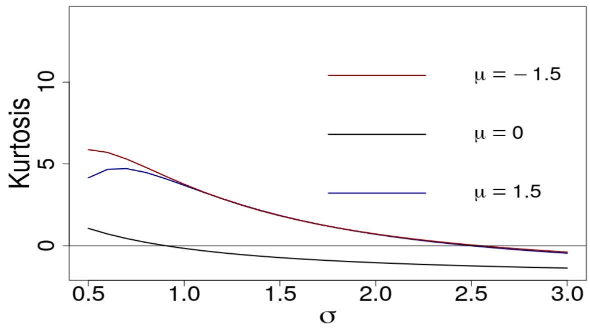
Figure 37 – Measures of Kurtosis of the logistic power exponential distribution in function of dispersion, location and shape parameter.



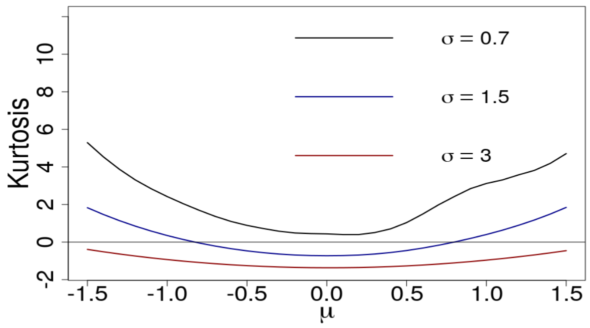
(a) Measure of Kurtosis of the logistic power exponential distribution for $\mu = 0$.



(b) Measure of Kurtosis of the logistic power exponential distribution for $\sigma = 0.7$.

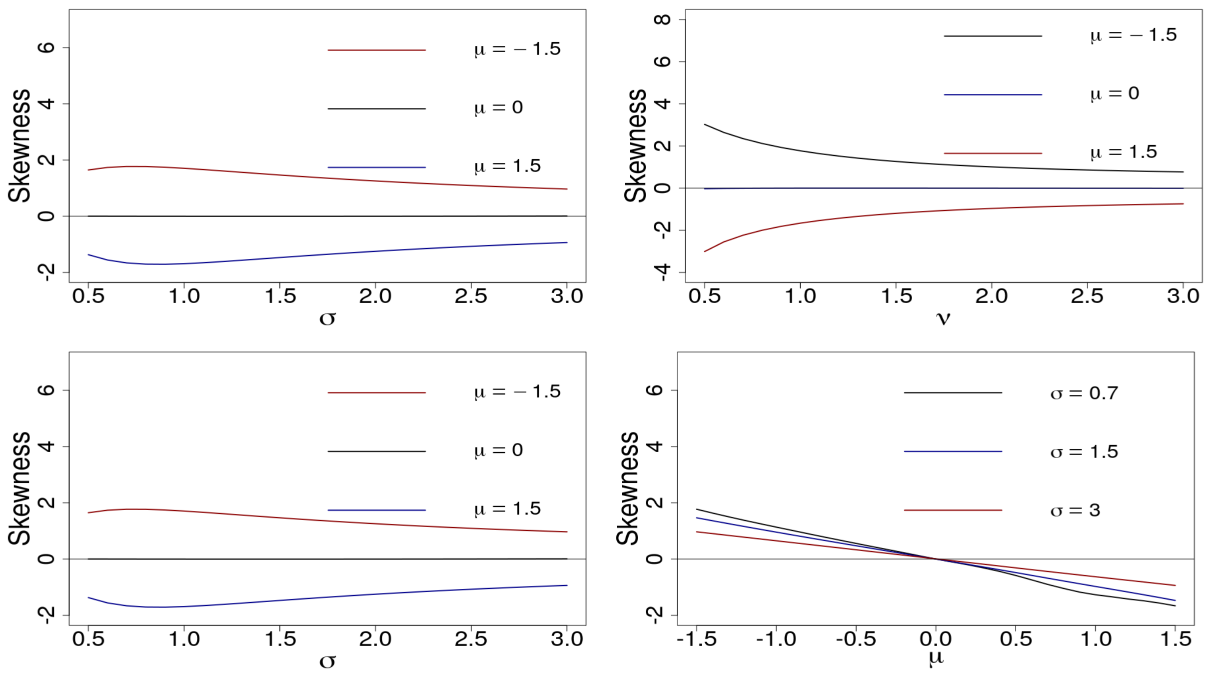


(c) Measure of Kurtosis of the logistic power exponential distribution for $\nu = 1$.



(d) Measure of Kurtosis of the logistic power exponential distribution for $\nu = 0$.

Figure 38 – Measures of Skewness of the logistic power exponential distribution in function of dispersion and location parameter.



A relevant conclusion about the distributions is: the three-parameter distributions present incredible flexibility compared to the two parameters. In general, the distributions presented a positive Skewness for positives values of μ, moderate negative Skewness for $\mu = 0$, and negative Skewness for negative values of μ. In many scenarios, the LSN, LGU, and LRG distributions, for example, show low values for the Skewness. These three distributions also presented low Kurtosis values. But we can highlight the logistic exponential Gaussian and logistic power exponential distributions for positive skew data and the logistic power exponential distributions for data with high Kurtosis value.

CHAPTER 3

THE JSB MIXED REGRESSION MODELS

This chapter presents novel regression models and mixed regression models for bounded response based on *JSB* distributions. Mixed models have been proposed for the analysis of data that are grouped or have some hierarchy. Longitudinal data have two sources of variation, within-individual, i.e., the variation in the repeated measurements within each individual; and between-individual, i.e., the variation in the data between different individuals. All distributions presented in chapter 2 do not have a closed-form for the expected value $E(Y)$. The goal is to define a parametric function of covariates to the location, dispersion, and shape parameters of the distribution.

3.1 The JSB regression models

Let $\boldsymbol{Y} = (Y_1, \ldots, Y_n)$ be a vector of independent random variables following one of the distributions described in chapter 2 with location parameter μ_i, dispersion parameter σ_i and shape parameter ν_i, and consider $\boldsymbol{x}_{1i}$, $\boldsymbol{x}_{2i}$ and $\boldsymbol{x}_{3i}$ three p, d and c dimensional vectors, respectively, containing the explanatory variables. Assuming that the random variables $\boldsymbol{Y}_i$ are mutually independent with JSB distribution, that is

$$\boldsymbol{Y}_i \sim JSB(\mu_i, \sigma_i, \nu_i), \tag{3.1}$$

with

$$s_1(\mu_i) = \boldsymbol{x}'_{1i}\boldsymbol{\beta} \qquad s_2(\sigma_i) = \boldsymbol{x}'_{2i}\boldsymbol{\gamma}, \qquad s_3(\nu_i) = \boldsymbol{x}'_{3i}\boldsymbol{\delta}, \tag{3.2}$$

where $\boldsymbol{\beta} = (\beta_0, \ldots, \beta_{p-1})$, $\boldsymbol{\gamma} = (\gamma_0, \ldots, \gamma_{d-1})$ and $\boldsymbol{\delta} = (\delta_0, \ldots, \delta_{c-1})$ represents, respectively, $p-$, $d-$ and $c-$ dimensional vectors of unknown regression coefficients.

The link function s_1 relates the location parameter with the covariates, since $\mu \in \mathbf{R}$ the identity function was taken, that is, $\boldsymbol{x}_1'\boldsymbol{\beta} \in \mathbf{R}$. Similarly, s_2 is a link relating the dispersion parameter σ_i with the covariates, as σ_i must be strictly positive we decide the log link. Finally, the link function s_3 depends on the distribution chosen, if $\nu > 0$ we adopt logarithmic and for $-\infty < \nu < \infty$ we adopt identity. All link functions are strictly monotonic and twice differentiable.

3.2 The JSB mixed regression models

Data with repeated measures have more than one observation per individual. This data brings correlation patterns about individuals. From this, special attention is necessary for the data analysis. Longitudinal data requires specific statistical methods because the set of observations on one subject tends to be intercorrelated.(RIZOPOULOS, 2012).

In longitudinal studies, we can explore individuals' patterns over time, making inferences about particular individuals. The mixed model has the characteristic to includes random effects for each individual to incorporate the correlation between repeated measurements. These random effects represent the individual influence on the repeated measurements, and their magnitude measures the variability across individuals. (WU, 2009).

The graphical representation of an LMM in Figure 39 shows the longitudinal response (y) from two individuals (1 and 2), and x represents the time the response was taken. As we can see in Figure 39, the linear regression model may be suitable to describe each individual. Assume y_{ij} as the response of individual $i, i = 1, \ldots, n$ at time $t_{ij}, j = 1, \ldots, n_i$, assuming different slopes and intercepts for each individual, a plausible model for y_{ij} is:

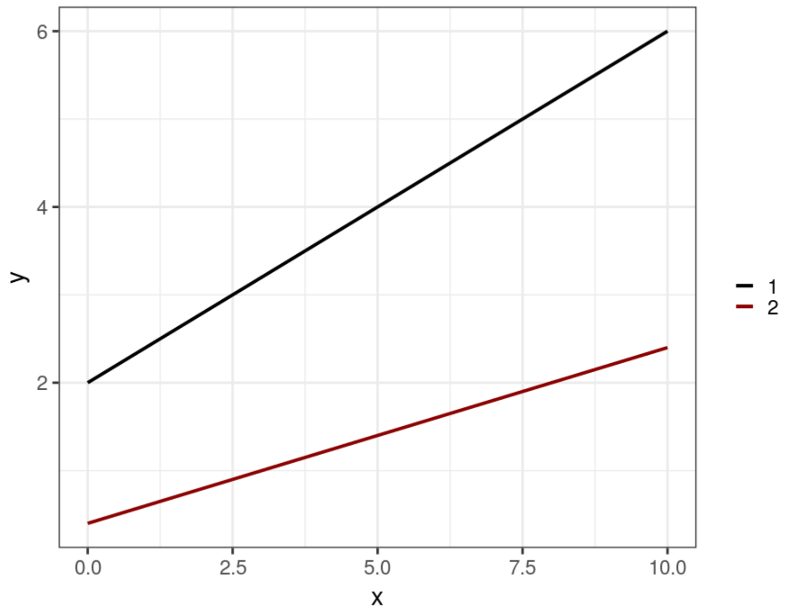
$$y_{ij} = \tilde{\beta_{i0}} + \tilde{\beta_{i1}} t_{ij} + \varepsilon_{ij}, \tag{3.3}$$

where $\tilde{\beta_{i0}}$ and $\tilde{\beta_{i1}} t_{ij}$ are subject-specific regression coefficients and $\varepsilon_{ij} \sim N(0, \sigma^2)$. Since subjects are sampled randomly from a population of individuals, it is reasonable to assume that the subject-specific regression coefficients $\tilde{\beta_{i0}}$ and $\tilde{\beta_{i1}} t_{ij}$ will be sampled randomly from the corresponding population regression coefficients. It is customary to adopt that the distribution of the regression coefficients in the population is a bivariate normal distribution with mean vector $\beta = (\beta_0, \beta_1)'$ and covariance matrix D. The model is reformulate as:

$$y_{ij} = (\beta_0 + b_{i0}) + (\beta_1 + b_{i1}) t_{ij} + \varepsilon_{ij}. \tag{3.4}$$

where $\tilde{\beta_{i0}} = \beta_0 + b_{i0}, \tilde{\beta_{i1}} = \beta_1 + b_{i1}$, and the new terms b_{i0} and b_{i1} are called random effects, having a bivariate normal distribution with mean zero and covariance matrix D. The parameters β_0 and β_1 are called fixed effects.

Figure 39 – A graphical representation from the linear mixed model, the black and red lines represent two hypothetical linear regression of two different individuals.



Source: Elaborated by the author.

The generalization of the above model is known as the linear mixed model and has the form (LAIRD; WARE, 1982) :

$$\boldsymbol{Y}_i = X_i\boldsymbol{\beta} + Z_i\boldsymbol{b}_i + \boldsymbol{\varepsilon}_i. \tag{3.5}$$

where $\boldsymbol{b}_i \sim N(0, \boldsymbol{D})$, $\boldsymbol{\varepsilon}_i \sim N(0, \sigma^2 I_{n_i})$, $\boldsymbol{Y}_i$ represents the response variable vector with dimension n_i from subject i. β is a p-dimensional vector of regression coefficients (fixed effects), $\boldsymbol{b}_i$ is a q-dimension vector of random effects, $\boldsymbol{X}_i$ is a $n_i \times p$ matrix representing the values of the covariates, and $\boldsymbol{Z}_i$ is a $n_i \times q$ design matrix for the random effects. The random effect b_i described how the subject i deviates from the population trend.

The main diagonal elements of matrix $\boldsymbol{D}$ represent random effects, $\boldsymbol{b}_i$, variances, for q random effects by subject i the matrix $\boldsymbol{D}$ can be represent as a $q \times q$ matrix:

$$\boldsymbol{D} = \begin{bmatrix} Var(b_{1i}) & cov(b_{1i}, b_{2i}) & \dots & cov(b_{1i}, b_{qi}) \\ cov(b_{1i}, b_{2i}) & Var(b_{1i}) & \dots & cov(b_{2i}, b_{qi}) \\ \vdots & \vdots & \vdots & \vdots \\ cov(b_{1i}, b_{qi}) & cov(b_{2i}, b_{qi}) & \dots & Var(b_{qi}). \end{bmatrix} \tag{3.6}$$

The $\boldsymbol{D}$ matrix can assume different structures. No one correlation structure for $\boldsymbol{D}$ was fixed, so the model does not imply restriction to the correlations. Every parameter from $\boldsymbol{D}$ is estimated freely. For more details about the correlation structure for the $\boldsymbol{D}$ matrix, see Singer and ANDRADE (1986). The inferences about mixed regression require estimates from the unknown regression coefficients and estimates from the covariance matrix and the random effects $\boldsymbol{b}_i$. The variances from the $\boldsymbol{D}$ matrix tell how the individuals are different. Large variances indicate large differences in individual profiles.

The new mixed models based on *JSB* distributions also include random effects for each individual to incorporate the correlation between repeated measurements. However, now the focus from the *JSB* mixed regression models is longitudinal data bounded in the interval $(0,1)$.

Let $\boldsymbol{Y}_i = (y_{i1}, \ldots, y_{in_i})'$ be a vector of response variables for the sample unit i such that every component y_{ij} takes values in the $(0,1)$ interval. The *JSB* mixed regression model is given by:

$$Y_{ij}|\boldsymbol{b}_i \sim JSB(\mu_{ij}, \sigma_{ij}, \nu_{ij}) \tag{3.7}$$
$$s_1(\mu_{ij}) = \boldsymbol{x}'_{1ij}\boldsymbol{\beta} + \boldsymbol{z}'_{ij}\boldsymbol{b}_i \qquad s_2(\sigma_{ij}) = \boldsymbol{x}'_{2ij}\boldsymbol{\gamma}, \qquad s_3(\nu_{ij}) = \boldsymbol{x}'_{3ij}\boldsymbol{\delta},$$
$$\boldsymbol{b}_i \sim N_q(0, \boldsymbol{D}),$$

for $j = 1, \ldots, n_i$ and $i = 1, \ldots, n$ where $\boldsymbol{\beta} = (\beta_0, \ldots, \beta_{p-1})$, $\boldsymbol{\gamma} = (\gamma_0, \ldots, \gamma_{d-1})$ and $\boldsymbol{\delta} = (\delta_0, \ldots, \delta_{c-1})$ are vectors of regression coefficients (fixed effects), $\boldsymbol{x}_{1ij} = (x_{1ij1}, \ldots, x_{1ijp})'$, $\boldsymbol{z}_{ij} = (z_{ij1}, \ldots, z_{ijr})$, $\boldsymbol{x}_{2ij} = (x_{2ij1}, \ldots, x_{2ijq})$ and $\boldsymbol{x}_{3ij} = (x_{3ij1}, \ldots, x_{3ijc})$ are vectors containing explanatory variables (overlapping or even identical). We assume that the random effects $\boldsymbol{b}_1, \ldots, \boldsymbol{b}_n$ are normally distributed, $\boldsymbol{b}_i \sim N_q(0, \boldsymbol{D})$ for $i = 1, \ldots, n$ where $\boldsymbol{D}$ is a positive definite matrix. The link function s_1, s_2 and s_3 are the same defined for the model 3.1.

CHAPTER 4

ESTIMATION

This chapter presents the estimation methods under the classical and Bayesian approaches for JSB mixed regression models. The penalized likelihood estimators are obtained by maximizing the penalized likelihood estimators and are computed by the Rigby and Stasinopoulos (RS) algorithm. Under the Bayesian approach, the posterior distribution from the interest parameters is complex, and the No-U-Turn-Sampler (NUTS) algorithm is employed to simulated values from the posterior distribution. Some computational issues from the algorithms are covered at the end of the chapter.

4.1 Classical Estimation

For the mixed regression models, we assume random effects $\boldsymbol{b}$, providing an additional source of variation. In the model described in (3.7), the response variables $Y_1, \ldots, Y_n$ are assumed to be independently distributed with pdf $f(y_i|\boldsymbol{\Theta}, \boldsymbol{b})$ from the JSB class, where $\Theta = (\boldsymbol{\beta}^T, \boldsymbol{\gamma}^T, \boldsymbol{\delta}^T)$. The model marginal likelihood is obtained by integrating out the random effect

$$L(\boldsymbol{\Theta}|\boldsymbol{Y}) = \prod_{i=1}^{n} \int \prod_{j=1}^{n_i} f(y_{ij}|\mu_{ij}, \sigma_{ij}, v)\phi_p(\boldsymbol{b}_i|\boldsymbol{0}, \boldsymbol{D})d\boldsymbol{b}, \tag{4.1}$$

where ϕ_q is the normal distribution pdf of dimension q. The main obstacle to dealing with the marginal likelihood is the integral over the random variables. Under the assumption that both the conditional distribution $\boldsymbol{Y}|\boldsymbol{b}$ and the marginal distribution $\boldsymbol{b}$ are normal, the marginal distribution $\boldsymbol{Y}$ is normal, so the marginal likelihood has an exact solution. In our case, the conditional is not normal, although we can approximate the marginal likelihood by some numerical methods, such as Gauss-Hermite quadrature or Monte Carlo integration. However, these options become less advantageous as the dimension of the random effect $\boldsymbol{b}_i$ increases.

The penalized likelihood was chosen instead of marginal likelihood. The penalized likelihood estimators are obtained by maximizing the penalized likelihood and must be computed numerically by the RS algorithm (STASINOPOULOS *et al.*, 2017). The estimators are similar to those that were presented by Breslow and Clayton (1993) for GLMM. The penalized log-likelihood function of the model (3.7) is given by:

$$l_p = l(\boldsymbol{\Theta}) - \frac{1}{2}\boldsymbol{b}'\boldsymbol{D}\boldsymbol{b}, \tag{4.2}$$

where the log-likelihood of $L(\boldsymbol{\Theta}|\boldsymbol{b})$ is

$$\begin{aligned} l(\boldsymbol{\Theta}) &= \sum_{i=1}^{n}\sum_{j=1}^{n_i} \log\left[f(y_{ij}|\mu_{ij},\tau_{ij},\nu)\right], \\ &= \sum_{i=1}^{n}\sum_{j=1}^{n_i} \log\left[g\left(\ln\left(\frac{y_{ij}}{1-y_{ij}}\right)|\mu_{ij},\tau_{ij},\nu\right)\frac{1}{y_{ij}(1-y_{ij})}\right], \\ &= \sum_{i=1}^{n}\sum_{j=1}^{n_i} \log\left[g\left(\ln\left(\frac{y_{ij}}{1-y_{ij}}\right)|\mu_{ij},\tau_{ij},\nu\right)\right] - \log\left(y_{ij}(1-y_{ij})\right), \end{aligned} \tag{4.3}$$

with $\boldsymbol{b}$ being a vector compound of $\boldsymbol{b}_i$ random effects. The $\boldsymbol{D}$ matrix is now a diagonal block matrix with $\boldsymbol{D}$ matrix at each of n diagonal positions. We repeat the $\boldsymbol{D}$ notation just for convenience. Let $\boldsymbol{\Omega} = (\boldsymbol{\Theta}^T, \boldsymbol{b}^T)$ the vector of all parameters from 3.7. The score function is given by vector

$$\boldsymbol{U}(\boldsymbol{\Omega}) = \frac{\partial l_p}{\partial \boldsymbol{\Omega}}. \tag{4.4}$$

Additionally, consider the following components of the score function $\boldsymbol{u}_1 = \partial l_p/\partial\boldsymbol{\beta}$, $\boldsymbol{u}_2 = \partial l_p/\partial\boldsymbol{b}$, $\boldsymbol{u}_3 = \partial l_p/\partial\boldsymbol{\gamma}$ and $\boldsymbol{u}_4 = \partial l_p/\partial\boldsymbol{\delta}$ so we can write (4.4) as:

$$\boldsymbol{U} = \begin{pmatrix} \boldsymbol{X}_1'\boldsymbol{s}_1 \\ \boldsymbol{Z}'\boldsymbol{s}_1 - \boldsymbol{D}\boldsymbol{b} \\ \boldsymbol{X}_2'\boldsymbol{s}_2 \\ \boldsymbol{X}_3'\boldsymbol{s}_3 \end{pmatrix}, \tag{4.5}$$

where $\boldsymbol{s}_k = \partial l(\boldsymbol{\Theta})/\partial\boldsymbol{\eta}_k$ is an $N-$dimensional vector ($N = \sum_{i=1}^{n} n_i$) with elements $\partial l(\boldsymbol{\Theta})/\partial\eta_{kl}$ for $l = 1,\dots,N$ and $k = 1,2,3$ represents the number of linear predictors. The penalized maximum likelihood estimators $\hat{\boldsymbol{\beta}}, \hat{\boldsymbol{b}}, \hat{\boldsymbol{\gamma}}$ and $\hat{\boldsymbol{\delta}}$ can be obtained by solving the equation system $\boldsymbol{U}(\boldsymbol{\Omega}) = \boldsymbol{0}$.

However, the equation system $\boldsymbol{U}(\boldsymbol{\Omega}) = \boldsymbol{0}$ does not have an explicit solution demanding a numerical approximation by an algorithm. (RIGBY; STASINOPOULOS, 2005) revealed that the RS algorithm, for known components of matrix $\boldsymbol{D}$, gives the maximum penalized log-likelihood estimates. The RS algorithm have three steps: the outer iteration, the inner iteration and the modified backfitting. These steps are nested, so the outer iteration repeatedly calls the inner iteration, which in turn repeatedly calls the modified backfitting algorithm.

The outer iteration estimates the parameters separately, first estimating $\boldsymbol{\beta}$ and $\boldsymbol{b}$ given the latest estimates of $\hat{\boldsymbol{\gamma}}$ and $\hat{\boldsymbol{\delta}}$, then estimating $\boldsymbol{\gamma}$ given $\hat{\boldsymbol{\beta}}$, $\hat{\boldsymbol{b}}$ and $\hat{\boldsymbol{\delta}}$, and finally estimating $\boldsymbol{\delta}$ given $\hat{\boldsymbol{\beta}}$, $\hat{\boldsymbol{b}}$ and $\hat{\boldsymbol{\delta}}$.

To fit a particular parameter, $\boldsymbol{\beta}$ for example, the outer step calls the inner step. The inner step is a local scoring algorithm that fits a weighted least square (WLS) to a modified response variable and obtains revised estimates of $\hat{\boldsymbol{b}}^{(r)}$ and $\hat{\boldsymbol{\beta}}^{(r)}$ at the r-th iteration of the step. The inner step updates the weights, the score function, and the modified response variable at each iteration, and is equivalent to the Fisher scoring procedure (HASTIE; TIBSHIRANI, 1990). The expected Fisher information matrix is given by:

$$\boldsymbol{A} = E\left(-\frac{\partial^2 l_p}{\partial\boldsymbol{\Omega}\partial\boldsymbol{\Omega}'}\right) = \begin{pmatrix} \boldsymbol{A}_{11} & \boldsymbol{A}_{12} & \boldsymbol{A}_{13} & \boldsymbol{A}_{14} \\ \boldsymbol{A}_{21} & \boldsymbol{A}_{22} & \boldsymbol{A}_{23} & \boldsymbol{A}_{14} \\ \boldsymbol{A}_{31} & \boldsymbol{A}_{32} & \boldsymbol{A}_{33} & \boldsymbol{A}_{14} \\ \boldsymbol{A}_{41} & \boldsymbol{A}_{42} & \boldsymbol{A}_{43} & \boldsymbol{A}_{44} \end{pmatrix}, \tag{4.6}$$

where:

$$\begin{pmatrix} \boldsymbol{A}_{11} & \boldsymbol{A}_{12} \\ \boldsymbol{A}_{21} & \boldsymbol{A}_{22} \end{pmatrix} = \begin{pmatrix} \boldsymbol{X}_1'\boldsymbol{W}_{11}\boldsymbol{X}_1 & \boldsymbol{X}_1'\boldsymbol{W}_{11}\boldsymbol{Z} \\ \boldsymbol{Z}'\boldsymbol{W}_{11}\boldsymbol{X}_1 & \boldsymbol{Z}'\boldsymbol{W}_{11}\boldsymbol{Z}+\boldsymbol{D} \end{pmatrix}, \boldsymbol{A}_{33} = (\boldsymbol{X}_2'\boldsymbol{W}_{22}\boldsymbol{X}_2), \boldsymbol{A}_{44} = (\boldsymbol{X}_3'\boldsymbol{W}_{33}\boldsymbol{X}_3), \tag{4.7}$$

and $\boldsymbol{W}_{kk} = E\left(-\partial^2 l(\boldsymbol{\Theta})/\partial\boldsymbol{\eta}_k\partial\boldsymbol{\eta}_k'\right)$ = diag $\{E(-\partial^2 l(\boldsymbol{\Theta})/\partial\eta_{lk}\partial\eta_{lk})\}$ over $l = 1,\ldots,N$, for $k = 1,2,3$ is the weight matrix. The other matrices are equal to zero. The RS algorithm does not use cross derivatives (between the linear predictors) of the log-likelihood for the estimation, so $W_{ks} = 0$, for $k \neq s$.

The fixed and random effects, $\boldsymbol{\beta}$ and $\boldsymbol{b}$, are estimated by the modified backfitting step. At each iteration m of the modified backfitting step, the algorithm computes update values of $\hat{\boldsymbol{\beta}}^{(r,m)}$ and $\hat{\boldsymbol{b}}^{(r,m)}$. The expressions for $\hat{\boldsymbol{\beta}}, \hat{\boldsymbol{b}}, \hat{\boldsymbol{\gamma}}$ and $\hat{\boldsymbol{\delta}}$ for the RS algorithm are:

$$\hat{\boldsymbol{\beta}}^{(m+1,r)} = (\boldsymbol{X}_1'\boldsymbol{W}_{11}^{(r)}\boldsymbol{X}_1)^{-1}\boldsymbol{X}_1'\boldsymbol{W}_{11}^{(r)}(\boldsymbol{\eta}_1^{(r)} + (\boldsymbol{W}_{11}^{-1})^{(r)}\boldsymbol{u}_1^{(r)} - \boldsymbol{Z}\hat{\boldsymbol{b}}^{(m,r)}). \tag{4.8}$$

$$\hat{\boldsymbol{b}}^{(m+1,r)} = (\boldsymbol{Z}'\boldsymbol{W}_{11}^{(r)}\boldsymbol{Z}+\boldsymbol{D})^{-1}\boldsymbol{Z}'\boldsymbol{W}_{11}^{(r)}(\boldsymbol{\eta}_1^{(r)} + (\boldsymbol{W}_{11}^{-1})^{(r)}\boldsymbol{u}_2^{(r)} - \boldsymbol{X}_1\hat{\boldsymbol{\beta}}^{(m+1,r)}). \tag{4.9}$$

$$\hat{\boldsymbol{\gamma}}^{(m+1,r)} = (\boldsymbol{X}_2'\boldsymbol{W}_{22}^{(r)}\boldsymbol{X}_2)^{-1}\boldsymbol{X}_2'\boldsymbol{W}_{22}^{(r)}(\boldsymbol{\eta}_2^{(r)} + (\boldsymbol{W}_{22}^{-1})^{(r)}\boldsymbol{u}_3^{(r)}). \tag{4.10}$$

$$\hat{\boldsymbol{\delta}}^{(m+1,r)} = (\boldsymbol{X}_3'\boldsymbol{W}_{33}^{(r)}\boldsymbol{X}_3)^{-1}\boldsymbol{X}_3'\boldsymbol{W}_{33}^{(r)}(\boldsymbol{\eta}_3^{(r)} + (\boldsymbol{W}_{33}^{-1})^{(r)}\boldsymbol{u}_4^{(r)}). \tag{4.11}$$

We uptade m until the numerical values of $\hat{\boldsymbol{\beta}}$ and $\hat{\boldsymbol{b}}$ do not change.

With convergence from the backfitting part, the algorithm checks the global deviance (equal to minus twice the current fitted log- likelihood). If it has not converged yet, we update r and obtain new values of the weights $\boldsymbol{W}_{11}$, the score vectors $\boldsymbol{u}_1$, $\boldsymbol{u}_2$ and the modified responses z_1, z_2:

$$z_1 = \boldsymbol{\eta}_1 + \boldsymbol{W}_{11}^{-1}\boldsymbol{u}_1 \text{ and } z_2 = \boldsymbol{\eta}_1 + \boldsymbol{W}_{11}^{-1}\boldsymbol{u}_2, \tag{4.12}$$

and return to the modified backfitting part. We continue updating r until the global deviance converges. After the convergence of $\boldsymbol{\beta}$, we get the next parameter, $\boldsymbol{\gamma}$. The inner and modified backfitting steps are used now to estimate $\hat{\boldsymbol{\gamma}}$ with the weight matrix $\boldsymbol{W}_{22}$, score vector $\boldsymbol{u}_3$ and modified response z_3.

We continue updating r until the global deviance converge. After the convergence of $\boldsymbol{\beta}$ we get the next parameter, $\boldsymbol{\gamma}$. The inner and modified backfitting steps are used now to estimate $\hat{\boldsymbol{\gamma}}$ with the weight matrix $\boldsymbol{W}_{22}$, score vector $\boldsymbol{u}_3$ and modified response z_3. With the convergence of $\boldsymbol{\gamma}$ the inner and modified backfitting steps are used now to estimate $\hat{\boldsymbol{\delta}}$.

Finally after estimating $\boldsymbol{\beta}$, $\boldsymbol{\gamma}$, and $\boldsymbol{\delta}$, the global deviance is checked again. With the convergence of the global deviance, we stop the algorithm and obtain the last values of $\hat{\boldsymbol{\beta}}, \hat{\boldsymbol{b}}, \hat{\boldsymbol{\gamma}}$ and $\hat{\boldsymbol{\delta}}$. Otherwise, we continue to update the parameters.

The described process works with known components from the $\boldsymbol{D}$ matrix. (RIGBY; STASINOPOULOS, 2014) suggested maximum likelihood estimators from an internal model approximated at the end of each m iterations from the backfitting step. The internal mixed model assumed at m iteration is given by:

$$\boldsymbol{e} = \boldsymbol{Z}\boldsymbol{b} + \boldsymbol{\varepsilon}; \quad \boldsymbol{\varepsilon} \sim N(0, \Sigma), \tag{4.13}$$

where $\Sigma = \sigma_e^2 W_{11}^{-1}$ and $\boldsymbol{e}$ are the current partial residuals (within backfitting, $\boldsymbol{e} = z_2 - \boldsymbol{X}_1\hat{\boldsymbol{\beta}}$). We can write (4.13) as:

$$z_2 = \boldsymbol{X}_1\hat{\boldsymbol{\beta}} + \boldsymbol{Z}\hat{\boldsymbol{b}} + \boldsymbol{\varepsilon}. \tag{4.14}$$

Consider (4.14) the linear model with response variable z_i for an individual i, Σ the covariance matrix of errors and $\boldsymbol{V}_i(\phi) = \boldsymbol{Z}_i\boldsymbol{D}\boldsymbol{Z}_i' + \boldsymbol{\Sigma}_{ii}$ the marginal variance for individual i with ϕ the components of the matrix $\boldsymbol{V}_i$. We can obtain the Restricted maximum likelihood (REML) estimator for $\boldsymbol{D}$ from the linear model in 4.14 reached by maximizing the restrict maximum likelihood (HARVILLE, 1976).

Laird and Ware (1982) uses the Expectation-Maximization (EM) algorithm to obtain the REML numerical estimates. The EM algorithm is a iterative algorithm for likelihood estimation in models with incomplete data (DEMPSTER; LAIRD; RUBIN, 1977). The REML estimator for $\hat{\boldsymbol{D}}$ are described in details in Laird and Ware (1982). With the REML estimates from $\hat{\boldsymbol{D}}$ we back to backfitting part to update $\hat{\boldsymbol{\beta}}$ and $\hat{\boldsymbol{b}}$.

4.2 Bayesian Estimation

For the specification of the model under a Bayesian approach we need the specification of complete likelihood

$$L(\boldsymbol{\Theta},\boldsymbol{b}|\boldsymbol{Y}) = \prod_{i=1}^{n}\prod_{j=1}^{n_i}\left[g\left(\ln\left(\frac{y_{ij}}{1-y_{ij}}\right)|\mu_{ij},\tau_{ij},\nu_{ij}\right)\frac{1}{y_{ij}(1-y_{ij})}\right]\phi_p(\boldsymbol{b}_i|\boldsymbol{0},\boldsymbol{D}), \qquad (4.15)$$

where $g(.)$ is the baseline distribution defined in Chapter 3 and ϕ_p is the density of a normal multivariate with vector of means 0 and covariance matrix $\boldsymbol{D}$ and in Bayesian approach we take the precision parameter, $\tau = 1/\sigma$.

In the Bayesian context we assume distributions for the parameters, called prior distributions, that express our uncertainty about the parameters. For the fixed effects model defined in (3.1) we must assume prior distributions for the parameter vector $\boldsymbol{\Theta} = (\boldsymbol{\beta},\boldsymbol{\gamma},\boldsymbol{\delta})$. In the mixed models we assume random effects $\boldsymbol{b}_i \sim N(\boldsymbol{0},\boldsymbol{D})$ and look at matrix $\boldsymbol{D}$ as a parameter. Assuming independence between parameters the joint prior distribution $p(\boldsymbol{\Theta})$ is given by:

$$p(\boldsymbol{\Theta}) = p(\boldsymbol{\beta})p(\boldsymbol{\gamma})p(\boldsymbol{\delta}) \qquad (4.16)$$

where $p(\boldsymbol{\beta}), p(\boldsymbol{\gamma})$ and $p(\boldsymbol{\delta})$ are the prior density distributions. The normal distribution is fixed for the parameters $\beta_j \sim N(\mu_\beta,\sigma_\beta)$, $\gamma_j \sim N(\mu_\gamma,\sigma_\gamma)$, $\delta_j \sim N(\mu_\delta,\sigma_\delta)$. The inverse Wishart distribution is the prior for the covariance matrices of the random effects $\boldsymbol{D} \sim IW(c,\boldsymbol{R})$. With these all specification the model can be write under a bayesian approach:

$$\begin{aligned}
&Y_{ij}|\beta,\gamma,\delta \sim JSB(\mu_{ij},\tau_{ij},\nu_{ij})\\
&g_1(\mu_{ij}) = \boldsymbol{x}'_{1ij}\boldsymbol{\beta} + z_{ij}\boldsymbol{b}_i\\
&g_2(\tau_{ij}) = \boldsymbol{x}_{2ij}\boldsymbol{\gamma}\\
&g_3(\nu_{ij}) = \boldsymbol{x}_{3ij}\boldsymbol{\delta}\\
&\boldsymbol{b}_i|\boldsymbol{D} \sim N(\boldsymbol{0},\boldsymbol{D})\\
&\boldsymbol{\Theta} \sim N(\boldsymbol{0},\boldsymbol{\Sigma})\\
&\boldsymbol{D} \sim IW(c,\boldsymbol{R})
\end{aligned} \qquad (4.17)$$

where $\boldsymbol{\Theta} = (\boldsymbol{\beta},\boldsymbol{\gamma},\boldsymbol{\delta})$ and $\boldsymbol{\Sigma}$ is a diagonal matrices with elements $\sigma_\beta,\sigma_\gamma,\sigma_\delta$. Combining the priors with likelihood function defined in (4.15), the posterior distribution for is given by:

$$p(\boldsymbol{\Theta},\boldsymbol{D}|Y) \propto L(\boldsymbol{\Theta},\boldsymbol{b}|\boldsymbol{Y})p(\boldsymbol{\beta})p(\boldsymbol{\gamma})p(\boldsymbol{\delta})p(\boldsymbol{D}). \qquad (4.18)$$

The posterior distribution does not have a closed-form since it is analytically intractable. The simulation of values direct from it is very laborious or impracticable. An alternative way is to use simulation methods based on Markov chains called MCMC methods. The MCMC methods are useful to draw samples from the parameters densities.

The MCMC methods are based on the construction of Markov Chains for posterior distribution (stationary distribution). The most popular algorithms of the MCMC are Metropolis-Hastings and Gibbs Sampling. These methods are considered straightforward and efficient. Although for complex models with many parameters, they may require a long time for convergence. The No-U-Turn-Sampler (NUTS) algorithm developed by Hoffman and Gelman (2014) is employed in this work.

The method is based on the Hamiltonian dynamics, which describes how objects move through a system. The dynamic is defined in terms of position x and momentum p of an object in the function of time t. This object has potential energy $U(x)$ and a Kinetic energy $K(p)$ associated with the momentum and position. The total energy of the system is called Hamiltonian $H(x,p)$. The Hamiltonian equations determine how the position x and momentum p change over time t.

The Hamiltonian Monce Carlo (HMC) transforms the idea of the Hamiltonian dynamic into a probabilistic context. Specifically, for the Bayesian context which we need to simulate values of our posterior distribution. Assume that $\boldsymbol{\Psi} = (\boldsymbol{\Theta}, \boldsymbol{b}, \boldsymbol{D})$ and $L(\boldsymbol{\Psi}|\boldsymbol{Y})$ are our parameter vector and log posterior distribution respectively. The HMC introduces an auxiliary momentum variable $\boldsymbol{r}_d$ for each variable Ψ_d at vector $\boldsymbol{\Psi}$. This momentum variable should be drawn independently from the standard normal distribution, yielding the joint density given by:

$$p(\boldsymbol{\Psi}, \boldsymbol{r}) \propto \exp\{L(\boldsymbol{\Psi}|\boldsymbol{Y}) - (1/2)\boldsymbol{r}\cdot\boldsymbol{r}\}. \tag{4.19}$$

The Hamiltonian equations define the position and momentum of the object in time t. In the Bayesian context, we simulate values from our posterior distribution if we solve the Hamiltonian equations. The Hamiltonian equations can be approximate through the Leapfrog integrator. The Leapfrog algorithm takes discrete steps of some small-time interval ε and runs L iterations of the algorithm to find a reasonable approximation of Hamiltonian equations:

$$r^{t+\varepsilon/2} = r^t + (\varepsilon/2)\nabla_{\boldsymbol{\Psi}} U(\boldsymbol{\Psi}^t);\ \ \boldsymbol{\Psi}^{t+\varepsilon} = \boldsymbol{\Psi}^t + \varepsilon r^{t+\varepsilon/2};\ \ r^{t+\varepsilon} = r^{t+\varepsilon/2} + (\varepsilon/2)\nabla_{\boldsymbol{\Psi}} U(\boldsymbol{\Psi}^{t+\varepsilon}) \tag{4.20}$$

where r^t and $\boldsymbol{\Psi}^t$ denote the values of momentum and position variable at time t and $\nabla_{\boldsymbol{\Psi}}$ denotes the gradient with respect to $\boldsymbol{\Psi}^t$ given by:

$$\frac{\partial U(\boldsymbol{\Psi})}{\partial \boldsymbol{\Psi}} = \nabla_{\boldsymbol{\Psi}} U(\boldsymbol{\Psi}) = \left(\frac{\partial L(\boldsymbol{\Psi})}{\partial \boldsymbol{\beta}}, \frac{\partial L(\boldsymbol{\Psi})}{\partial \boldsymbol{\gamma}}, \frac{\partial L(\boldsymbol{\Psi})}{\partial \boldsymbol{\delta}}, \frac{\partial L(\boldsymbol{\Psi})}{\partial \boldsymbol{b}}, \frac{\partial L(\boldsymbol{\Psi})}{\partial \boldsymbol{D}}\right) \tag{4.21}$$

To draw a value of posterior at iteration m we started by simulating r from a $N(0,\boldsymbol{I})$, run the Leapfrogs equations (4.20) L times, and get the last value for $\boldsymbol{\Psi}$. The propose value is accepted according to a Metropolis acceptance.

The problem with HMC is the decision of parameters L and ε. The performance of the algorithm has a strong connection with their choice. If the parameter ε is large, the simulation may be inaccurate and yield low acceptance rates. If it is small, the algorithm will have an unnecessary computational waste. Concerning the L parameter, if its value is modest, the successive samples will be closer to each other resulting in undesirable random walk behavior and slow mixing. For larges values for the L parameter, the trajectory begins to loop back and retrace their steps.

The algorithm NUTS automatically selects an appropriate value for L in each iteration to avoid this undesirable random walk. Firstly, as the parameter will no longer be fixed, it is necessary to take the criteria to advise us when the proposed value $\hat{\boldsymbol{\Psi}}$ starts to move back towards $\boldsymbol{\Psi}$ and then stop the leapfrog steps (HOFFMAN; GELMAN, 2014). The criterion is given by: $(\hat{\boldsymbol{\Psi}}-\boldsymbol{\Psi})\cdot\hat{\boldsymbol{r}}$, where $\hat{\boldsymbol{\Psi}}$ and $\hat{\boldsymbol{r}}$ are the proposed values and $\boldsymbol{\Psi}$ the initial value of the iteration. This criterion suggests an algorithm that will continue to runs Leapfrog steps until this criterion becomes less than 0. However, this approach does not guarantee time reversibility then it does not guarantee the convergence to the correct distribution. NUTS overcome this problem by employing a recursive algorithm that preserves reversibility by running the Hamiltonian simulation both forward and backward in time and introducing a slice variable u to follow slice sampling. The joint probability of $(\boldsymbol{\Psi},\boldsymbol{r})$ and u is given by:

$$p(\boldsymbol{\Psi},\boldsymbol{r},u) \propto \mathbb{I}[u \in [0, \exp\{L(\boldsymbol{\Psi}) - \frac{1}{2}\boldsymbol{r}\cdot\boldsymbol{r}\}]], \tag{4.22}$$

where $\mathbb{I}(\cdot)$ is 1 if the expression in brackets is true and 0 if it is false. The marginal probability of ($\boldsymbol{\Psi}$ and $\boldsymbol{r}$) is

$$p(\boldsymbol{\Psi},\boldsymbol{r}) \propto \exp\{L(\boldsymbol{\Psi}) - \frac{1}{2}\boldsymbol{r}\cdot\boldsymbol{r}\}, \tag{4.23}$$

which is independent of u.

Sampling directly from (4.22) may not be easy. An alternative is to use the Gibbs sampling idea, so we sample from the conditional distribution $p(u|\Psi,r)$ and the conditional distribution for $(\boldsymbol{\Psi},\boldsymbol{r})$ given u, which is uniform over the region $u \leq \{L(\boldsymbol{\Psi}) - \frac{1}{2}\boldsymbol{r}\cdot\boldsymbol{r}\}$.

The algorithm NUTS used to sample from $\boldsymbol{\Psi},\boldsymbol{r}|u$ proceeds by: 1. chooses M as the chain size where $m = 1,\ldots M$; 2. sample from $\boldsymbol{r} \sim N(\boldsymbol{0},\boldsymbol{I})$ and initializes $\boldsymbol{\Psi}^{m-1}$, $\zeta = \{(\boldsymbol{\Psi}^{m-1},\boldsymbol{r})\}, j = 0$; 3. sample $u \sim U([0,\exp\{L(\boldsymbol{\Psi}^{m-1}) - (1/2)\boldsymbol{r}\cdot\boldsymbol{r}\}])$; 4. chooses a direction v_j (i.e, forwards in time if v_j = 1 and backwards in fictional time if v_j = -1); 5. runs the Leapfrog defined in 4.20 with $\boldsymbol{\Psi}^{m-1}, \boldsymbol{r}$ and $\varepsilon = v_j\varepsilon$ to obtain the new values $\hat{\boldsymbol{\Psi}},\hat{\boldsymbol{r}}$; 6. checks $u \leqslant \exp\{L(\hat{\boldsymbol{\Psi}}) - (1/2)\hat{\boldsymbol{r}}\cdot\hat{\boldsymbol{r}}\}$ and $\mathbb{I}[L(\hat{\boldsymbol{\Psi}}) - (1/2)\hat{\boldsymbol{r}}\cdot\hat{\boldsymbol{r}} > \log u - 1000]$; 7. if step 6 is true the algorithm updates $\zeta = \zeta \cup \{\hat{\boldsymbol{\Psi}},\hat{\boldsymbol{r}}\}$;

8. checks the criterion $(\hat{\boldsymbol{\Psi}} - \boldsymbol{\Psi}^{m-1}) \cdot \hat{\boldsymbol{r}}$ and updates j; 9. if step 8 is true the algorithm sample $\boldsymbol{\Psi}^m$ from ζ and update m and back to step 2, otherwise back to the step 3.

We define as ζ a subset of all states $(\boldsymbol{\Psi}, \boldsymbol{r})$ trace out during a given NUTS iteration by Leapfrog. Any state in ζ be in the slice define by u and have equal conditional probability $p(\hat{\boldsymbol{\Psi}}, \hat{\boldsymbol{r}}|u)$. The joint distribution of $\hat{\boldsymbol{\Psi}}$ and $\hat{\boldsymbol{r}}$ given u, ζ, ε is uniform on the elements of ζ, allowing us to at the end of each NUTS iteration sample $\boldsymbol{\Psi}$ and $\boldsymbol{r}$ from ζ. This process is useful only for the parameter L. More details about the use of stochastic optimization, specifically the primal-dual algorithm of Nesterov (2009) to find an optimal value for ε, can be find in Hoffman and Gelman (2014).

4.3 Computational Issues

We cite the computational questions about the estimation methods for the Bayesian and classical approaches. All methods are applicable in R software through the `gamlss` (RIGBY *et al.*, 2019). (STASINOPOULOS; RIGBY, 2007) (RIGBY; STASINOPOULOS, 2005) package for the classical approach and `rstan` (Stan Development Team, 2020) for the Bayesian approach.

The `gamlss` package grants the fitting of Generalized additive model for location, scale, and shape (GAMLSS) class of models. In the GAMLSS, the response variable distribution is relaxed to allow distributions apart from the exponential family. The systematic part of the model is expanded to modeling the mean (or location) and other parameters as a linear function of explanatory variables and/or random effects. The package likewise covers the transformation given in Chapter 2, granting us to adjust the *JSB* class. The estimation process is made by the RS algorithm (default of package). The package offers the CG algorithm. However the CG algorithm requires information about the cross derivatives of the log-likelihood function with respect to the distribution parameters. The package offers a significant flexibilization to user preferences. It is possible to adjust the iterations number of the algorithm, define initial values, change the number of iterations, and define the numerical convergence criteria.

For our particular situation, the mixed models is fitted by the aid of **re** function which accomplishes an interface with **lme** function from the `nlme` package (PINHEIRO *et al.*, 2020). The function employs a hybrid algorithm between the Expectation-Maximization (EM) algorithm and Newton-Raphson (NR) method to maximize the restricted log-likelihood. NR iterations are more computationally intensive than the EM iterations and can be quite unstable when far from the optimum and converge quickly close to the optimum. The hybrid algorithm performed a moderate number of EM iterations, then switching to NR iterations. An example of the code using the `gamlss` package to fit a *JSB* mixed regression model is included in the Appendix B.

In the Bayesian context, Stan provides a flexible probabilistic programming language for statistical modeling (Stan Development Team, 2020). Stan supports an efficient program to run the NUTS algorithm. We can employ regression models and mixed models presented here using the language. A code for the *JSB* mixed regression model, specifically the LPE mixed regression model, using the Stan language is included in Appendix B. The `rstan` package offers a significant flexibilization to user preferences, for example, access to the log density, to the depth of tree explored by the NUTS, and to the number of leapfrog iterations (CARPENTER *et al.*, 2017). The default convergence diagnostic is the potential scaled reduction statistic $\hat{R}$, its value should be closed to 1 when the chains have converged, see Vehtari *et al.* (2020) more for details about $\hat{R}$.

For the priors, a Normal prior distribution centered at zero with standard error equal to 100 was chosen for the regression coefficient, which is sufficiently noninformative (PAZ; BALAKRISHNAN; BAZÁN, 2019). The covariance matrix has an associated correlation matrix, which can be factored into a square root matrix $\boldsymbol{L}_b$. A prior was placed on the $\boldsymbol{L}_b$ matrix following the Sorensen and Vasishth (2015) approach. The posterior is given by:

$$p(\boldsymbol{\Theta},\boldsymbol{b},\boldsymbol{D}|\boldsymbol{Y},\boldsymbol{x}) = L(\boldsymbol{\theta}\boldsymbol{Y},\boldsymbol{x})p(\boldsymbol{\beta})p(\boldsymbol{\gamma})p(\boldsymbol{\delta})p(\boldsymbol{\sigma}_b)p(\boldsymbol{L}_b)\phi_b(\boldsymbol{Z}_{\boldsymbol{b}}) \tag{4.24}$$

where ϕ_b is a normal multivariate with dimension q mean $\mathbf{0}$ and covariance matrix I. $\boldsymbol{\sigma}_b$ is a $q-$ dimensional vector containing the variance of every q random effect. Independents gamma-inverse priors are selected for every σ_b. The $\boldsymbol{L}_b$ matrix prior is a Cholesky LKJ Correlation distribution with parameter $v = 2$. The parameter $v = 2$ implies that the correlation on the off-diagonal are near zero, reflecting the fact that we have no prior information about the correlation between random effects.

4.4 Model Comparison Criteria

Model selection is a crucial step to statistical inference. As the main goal of this work is to introduce alternative models, we will contrast different models. The criteria do not check the model, but they compare them and explore directions for improvement.

Some criteria are presented, like the Akaike information criterion (AIC) and the Watanabe-Akaike information criterion (WAIC) (WATANABE; OPPER, 2010). The first for classical context and the second for Bayesian context. These criteria have an attractive aspect, being alternative adjustment measures for cross-validation (STONE, 1977).

4.4.1 Classical criteria

Akaike (1981) proposed the use of the Kullback Leibrle (KL) distance as the fundamental basis for model selection. The KL information between models f and g with parameter θ is defined for continuous function as the integral:

$$I(f,g) = \int f(x)\log(f(x))dx - \int f(x)\log(g(x|\theta))dx. \tag{4.25}$$

We seek a model that minimizes $I(f,g)$ over g, where f is known, and this represents a model that loses as little information as possible. The first term on the right of (4.25), the expectation $E_f[\log(f(x))]$, is a constant that depends only on the unknown true distribution f, so (4.25) can be written as: $I(f,g) = C - E_f[\log(g(x|\theta))]$. However, the KL distance can not be computed without full knowledge of both f and g. So as the true distribution f is unknown in practical problems, $E_f[\log(g(x|\theta))]$ becomes the quantity of interest. The AIC criterion adapted the KL distance and gives an estimated expected relative distance between the true and estimated model (ANDERSON; BURNHAM, 2004). The AIC criterion is given by:

$$\text{AIC} = -2\log(L(\hat{\boldsymbol{\Theta}}|\boldsymbol{y})) + 2k. \tag{4.26}$$

The BIC (SCHWARZ *et al.*, 1978) is closely related to AIC criterion with a diffrent penalty term. The BIC criterion is given by:

$$\text{BIC} = -2\log(L(\hat{\boldsymbol{\Theta}}|\boldsymbol{y})) + k\log(N), \tag{4.27}$$

where N is total of observations.

4.4.2 Bayesian criteria

In the Bayesian approach, a recent and important criterion is the WAIC criterion (WATANABE; OPPER, 2010). This criterion aims to estimate the predictive performance of the model in a new data set. Let us consider f with parameter θ as the true model and y as a realization of it. But now consider $\tilde{y}$ as future data, an out-of-sample predictive fit for a new data point $\tilde{y}_l$ is

In the Bayesian approach, a recent and important criterion is the WAIC criterion (WATANABE; OPPER, 2010). This criterion aims to estimate the predictive performance of the model in a new data set. Consider f with parameter θ the true model, y a realization of it and $\tilde{y}$ a future data. The log predictive density of the new data point $\tilde{y}_l$ is given by:

$$\log P(\tilde{y}_l) = \log E_\theta(p(\tilde{y}_l|\theta)) = \log \int p(\tilde{y}_l|\theta)P(\theta)d\theta, \tag{4.28}$$

only here we denote $P(\tilde{y}_l)$ as the predictive density and $P(\theta)$ the posterior function. The notation P is used just for simplification of the expressions that will be presented here. However, the future data are themselves unknown, and then we work with the expected log predictive density, $E(\log P(\tilde{y}_l)) = \int \log P(\tilde{y}_l) f(\tilde{y}_l) d\tilde{y}$. This expression is similar to that find on the right side of (4.25). To evaluate $E(\log P(\tilde{y}_l))$ we work with posterior distribution $P(\theta|\boldsymbol{y})$ and summarize the predictive accuracy of the fitted model to data by: (GELMAN; HWANG; VEHTARI, 2014)

$$\log \prod_{l=1}^{N} P(\boldsymbol{\theta}|y_l) = \sum_{l=1}^{N} \log \int p(y_l|\theta)P(\theta)d\theta \tag{4.29}$$

To compute (4.29) in practice we use draws from $P(\theta)$, the usual posterior simulations, to calculate the expectation $\sum_{l=1}^{N} \log\left(\frac{1}{S}\sum_{s=1}^{S} p(y_l|\theta^s)\right)$, where $\theta^s, s = 1,\ldots,S$ are the posterior simulations.

WAIC is a fully Bayesian criterion, starting with the expression in (4.29) and then adding a correction for overfitting based on the effective number of parameters. This correction (denoted by p_{WAIC}) uses the variance of individual terms in the log predictive density summed over the n data points. To calculate p_{WAIC} we compute the posterior variance for each data point y_l, that is, $V_{s=1}^{S} \log p(y_l|\theta^s)$, where $V_{s=1}^{S} a_s = \frac{1}{S-1}\sum_{s=1}^{S}(a_s - \hat{a})^2$. Summing over all the data points y_l (using the notation defined in 4.18) gives:

$$p_{WAIC} = \sum_{l=1}^{N} V_{s=1}^{S}(\log p(y_l|\boldsymbol{\Theta}^s, \boldsymbol{b}^s)). \tag{4.30}$$

So the WAIC criterion is given by:

$$\text{WAIC} = -2\left(\sum_{l=1}^{N} \log\left(\frac{1}{S}\sum_{s=1}^{S} p(y_l|\boldsymbol{\Theta}^s, \boldsymbol{b}^s)\right) - p_{WAIC}\right). \tag{4.31}$$

where $p(y|\boldsymbol{\Theta}, \boldsymbol{b})$ is the density of the *JSB* distribution. The EAIC and EBIC criteria were also considered (GELMAN; HWANG; VEHTARI, 2014):

$$\begin{aligned} \text{EAIC} &= \bar{D} + 2 \times p \\ \text{EBIC} &= \bar{D} + 10 \times \log(n), \end{aligned} \tag{4.32}$$

where

$$\bar{D} = S^{-1}\sum_{s=1}^{S}\left(-2\sum_{l=1}^{N} \log(p(y_l|\boldsymbol{\Theta}^s, \boldsymbol{b}^s))\right). \tag{4.33}$$

4.5 Residual analysis

The residual analysis in some linear regression models can be straightforward, since the residuals are normally distributed and can be standardized to have equal variances. Any discrepancy about this should be taken with caution when validating the model. However, in non-normal regression models, the residuals are not normally distributed. In the GLM there are many residuals to check the validity of the model (PAULA, 2004). For our models, the randomized quantile residuals (DUNN; SMYTH, 1996) was considered. These residuals should be used to perform diagnostic analysis in complex regression models (PEREIRA, 2019). The randomized quantile residual is normally distributed if the fit of the model is adequate.

Let $F(y;\boldsymbol{\theta})$ be the cumulative function of Y. If F is continuous, then $F(y;\boldsymbol{\theta})$ is uniformly distributed in the unit interval, and the randomized quantile residual for the observation l is defined by:

$$r_{q,l} = \Phi^{-1}\{F(y_l;\boldsymbol{\Omega}_l)\}, \tag{4.34}$$

where Φ is the cumulative distribution function of the standard normal distribution. For a graphical analysis of the residuals to assess the fit of the proposed model, we check the normality of the residuals by considering a QQplot. Additionally, we show the envelope graph for the randomized quantile residuals as proposed by Atkinson (1983) to measure the quality of the model estimation. The envelope was built following the steps: 1. Simulate distribution values using model estimates; 2. Fit y simulated versus explanatory variables; 3. Take the randomized quantile residuals from the model. We repeat these steps 19 times, so we have r_{lk} $(l = 1,\ldots,N)$ and $k = 1,\ldots,19$. The next step is to arrange each group of N residuals in ascending order and take the $r_{(l),0.025}$ and $r_{(l),0.975}$ quantiles.

CHAPTER 5

SIMULATION STUDIES

The first simulation study presents the performance of the proposed distributions in comparison with the Beta distribution. This simulation presents a study to evaluate the robustness of JSB distribution to outliers. The second simulation study performed out parameters recovery to evaluate the performance of the estimation method considered here. Specifically, the second simulation presents a zero-and-one distribution of the JSB class, and it checks out the performance of the estimation method for this particular model.

5.1 First Simulation Study

The first simulation presents the robustness to outliers of the distributions from Chapter 2 in comparison with the Beta distribution. That is, the simulation study observe the performance of the JSB and Beta distributions regarding contaminated data coming from a Beta distribution with outliers, following (PAZ; BALAKRISHNAN; BAZÁN, 2019). The parameterization of the Beta distribution employed is (FERRARI; CRIBARI-NETO, 2004):

$$f(y;\mu,\phi) = \frac{\Gamma(\phi)}{\Gamma(\mu\phi)\Gamma((1-\mu)\phi)} y^{\mu\phi-1}(1-y)^{(1-\mu)\phi-1}, \tag{5.1}$$

where $\mu = E(Y)$, the dispersion of the distribution, for fixed μ, decreases as ϕ increases, Var(y) = $\frac{V(\mu)}{1+\phi}$.

The scenarios considered are samples from a Beta distribution for three different values of the mean μ: 0.2, 0.5 and 0.8, two different values of ϕ, 10, and 20 and three sample sizes, $75, 150$ and 200. These specific values of μ and σ were choose to get a variety of samples, that is, samples with low values, $E(Y) = 0.2$ and high values $E(Y)$ = 0.8. The samples were contaminated by outliers simulated from a uniform distribution with parameters 0.9 and 1. The proportions of outliers considered were 0.02, 0.05, and 0.1 for each dataset. $r = 2\%, 5\%, 10\%$

were randomly replaced from each sample with outliers from a uniform distribution. Every scenario were replicated 1000 times.

The models were fitted by the RS algorithm to the contaminated samples. The comparison criteria AIC and BIC from JSB distributions and Beta distribution for each contaminated sample were used. The average and the percentage of cases where each distribution displayed the lowest value of these criteria were determined. The purpose of this analysis is to report if these criteria are adequate for choosing alternative models in the *JSB* class in comparison with the Beta distribution for the Beta contaminated data. Table 12 presents the percentages for Beta, logistic Gumbel (LGU), logistic reverse Gumbel (LRG), logistic exponential Gaussian (LEG), logistic skew normal type I (LSN1), logistic power exponential (LPE) for the scenario with $\mu = 0.2$ and all values of r. Tables 26, 28, 27 and 29 from Appendix A present the scenarios with $\mu = 0.5$, $\mu = 0.8$.

The percentages of cases where the JSB distribution outperformed the Beta are relevant. According to Table 12, for $\mu = 0.2$, the Beta distribution did not perform better than the JSB distributions in any of the scenarios considered. The LEG and LPE distributions displayed the most significant results with $\mu = 0.2$. The results of other μ values were similar except for few scenarios. According to Table 3.1.1 of the Supplementary Material, for $\mu = 0.5$, the LEG distribution presented good performance. Nevertheless, in some scenarios with small sample sizes, the LRG distribution showed important results, outperforming the LEG in some cases. Finally, in tables 27 and 29 from Appendix B shows, for $\mu = 0.8$, that the Beta distribution had the best fit in scenarios with $\phi = 10$ and large sample sizes. But with $n = 75$ and $r = 10$, the LRG distribution presented good results, indicating that the distribution fits small sample sizes well.

Table 12 – Percentage of cases that each distribution displayed the lowest value of AIC and BIC for different scenarios with contaminated samples from a Beta distribution with $\mu = 0.2$ (1000 replications of each scenario).

			AIC						BIC					
ϕ	% of outiler	n	Beta	LGU	LRG	LEG	LSN1	LPE	Beta	LGU	LRG	LEG	LSN1	LPE
10	2	75	0	0	4.7	30.8	0	64.5	0	0	15.2	20.3	0	64.5
		150	0	0	0.1	29.3	0	70.6	0	0	0.7	28.7	0	70.6
		200	0	0	0	26.9	0	73.1	0	0	0	26.9	0	73.1
	5	75	0	0	6.9	36.6	0	56.5	0	0	31.0	31.7	0	55.3
		150	0	0	0.1	33.4	0	66.5	0	0	2.9	30.6	0	66.5
		200	0	0	0	34.5	0	65.5	0	0	0.6	33.9	0	65.5
	10	75	0	0	2.0	56.2	0	41.8	0	0	22.9	36.7	0	40.4
		150	0	0	0	54.7	0	45.3	0	0	0.4	54.3	0	45.3
		200	0	0	0	58.5	0	41.5	0	0	0	58.5	0	41.5
20	2	75	0	0	8.0	39.4	0	52.6	0	0	15.3	0.6	0	84.1
		150	0	0	0.3	42.9	0	56.8	0	0	3.1	6.6	0	90.3
		200	0	0	0	38.1	0	61.9	0	0	0	13.5	0	86.5
	5	75	0	0	0.4	42.3	0	57.3	0	0	13.9	29.4	0	56.7
		150	0	0	0	40.7	0	59.3	0	0	0.5	40.2	0	59.3
		200	0	0	0	38.6	0	61.4	0	0	0	13.3	0	86.7
	10	75	0	0	0	55.2	0	44.8	0	0	0	55.2	0	44.8
		150	0	0	0	55.3	0	44.7	0	0	0	55.3	0	44.7
		200	0	0	0	58.1	0	41.9	0	0	0	58.1	0	41.9

Table 13 – Average of AIC and BIC regarding replicas for different scenarios with contaminated samples from a Beta distribution with $\mu = 0.2$ (1000 replications of each scenario).

	% of		AIC						BIC					
ϕ	outiler	n	Beta	LGU	LRG	LEG	LSN1	LPE	Beta	LGU	LRG	LEG	LSN1	LPE
10	2	75	-82.39	-40.20	-81.54	-90.94	-82.23	**-93.58**	-87.03	-44.84	-86.17	-97.89	-89.18	**-100.53**
		150	-156.16	-64.27	-165.81	-189.69	-165.49	**-195.03**	-150.14	-58.25	-159.79	-180.66	-156.46	**-186.00**
		200	-209.01	-73.57	-219.53	-253.75	-221.54	**-261.36**	-202.41	-66.97	-212.94	-243.85	-211.64	**-251.47**
	5	75	-49.88	-6.26	-71.77	-75.49	-53.80	**-78.11**	-54.52	-10.90	-76.41	-82.44	-60.75	**-85.07**
		150	-101.85	-13.51	-147.67	-162.01	-116.62	**-166.65**	-95.82	-7.49	-141.65	-152.98	-107.58	**-157.62**
		200	-130.89	-12.52	-193.09	-213.35	-151.22	**-219.94**	-124.29	-5.92	-186.50	-203.45	-141.33	**-210.04**
	10	75	-21.59	14.47	-55.56	-57.09	-26.14	**-57.25**	-26.23	9.83	-60.20	-64.04	-33.09	**-64.21**
		150	-51.88	20.40	-118.30	**-128.05**	-67.29	-126.67	-45.86	26.42	-112.28	**-119.02**	-58.26	-117.64
		200	-70.43	28.47	-157.61	**-172.07**	-91.15	-171.07	-63.84	35.06	-151.01	**-162.18**	-81.25	-161.17
20	2	75	-100.81	-38.01	-122.68	-127.38	-104.11	**-128.14**	-105.45	-42.64	-127.32	-134.33	-111.06	**-135.09**
		150	-181.55	-50.59	-245.36	-259.26	-196.76	**-261.04**	-175.53	-44.57	-239.34	-250.23	-187.73	**-252.00**
		200	-242.47	-38.45	-326.53	-346.68	-263.17	**-349.45**	-235.88	-31.86	-319.94	-336.78	-253.28	**-339.55**
	5	75	-54.80	1.77	-104.53	-106.26	-60.41	**-108.29**	-59.43	-2.87	-109.16	-113.21	-67.36	**-115.25**
		150	-107.98	5.03	-211.36	-221.99	-125.46	**-225.07**	-101.96	11.05	-205.34	-212.96	-116.42	**-216.04**
		200	-137.26	13.15	-276.57	-292.00	-160.78	**-296.25**	-130.67	19.74	-269.98	-282.11	-150.88	**-286.35**
	10	75	-19.36	25.09	-76.13	-81.71	-23.72	**-82.13**	-23.99	20.45	-80.76	-88.66	-30.67	**-89.08**
		150	-46.64	41.80	-158.15	**-176.28**	-61.48	-174.75	-40.62	47.82	-152.12	**-167.24**	-52.45	-165.72
		200	-63.30	57.08	-211.41	**-236.05**	-83.26	-234.20	-56.70	63.68	-204.82	**-226.16**	-73.37	-224.30

The conclusion is that the distributions from the JSB class, like the LEG and LPE, have better performance for samples with outliers in relation to the Beta model, independent of sample size and dispersion. A different result comes from $\mu = 0.8$, where the Beta distribution shows better performance for high dispersion. However, with lower values of the dispersion parameter and a more relevant presence of outliers, the LPE distribution again shows consistent results. Additionally, Table 13 presents the average of AIC and BIC regarding the replicas. The results from Table 12 and Table 13 lead to the same conclusions. In summary, the study have shown that bounded data exist that do not follow the Beta distribution, but it can be accommodated by some distributions in the *JSB* class.

5.2 Second Simulation Study

In the second simulation, an extension of the JSB distributions is introduced. A mixed distribution that is continuous within the interval $(0,1)$ with additional probabilities at either 0 and 1 is presented. It is called the zero-and-one inflated distribution. Only an inflated version of the LPE distribution, denoted by $f_Y(y|\boldsymbol{\theta})$, is generated due to the fact that the LPE distribution had the best performance among all other models in the previous simulation study. The zero-and-one-inflated LPE (ZOILPE) distribution has support in $[0,1]$ and pdf given by:

$$f_W(y|\boldsymbol{\theta}) = \begin{cases} p_0, & \text{if } y = 0 \\ (1 - p_0 - p_1) f_Y(y|\boldsymbol{\theta}), & \text{if } 0 < y < 1 \,, \\ p_1, & \text{if } y = 1 \end{cases} \tag{5.2}$$

for $0 \leq y \leq 1$, $0 < p_0 < 1$, $0 < p_1 < 1$ such that $\xi_0 = p_0/p_2$, $\xi_1 = p_1/p_2$ and $p_2 = 1 - p_0 - p_1$. $\xi_0 > 0$ and $\xi_1 > 0$ with

$$p_0 = \frac{\xi_0}{1+\xi_0+\xi_1} \text{ and } p_1 = \frac{\xi_1}{1+\xi_0+\xi_1} . \tag{5.3}$$

The default link functions for ξ_0 and ξ_1 are the log functions. The samples were simulated using the distribution defined in (5.2). The model is fitted as two separate models, both parts using the RS algorithm. $f_Y(y|\boldsymbol{\theta})$ is fitted for $(0,1)$ response and the probabilities are obtained by the multinomial distribution for the zeros, ones, non-zeros, and non-ones. Ospina and Ferrari (2012) described the procedure with details. The `gamlss.inf` (ENEA *et al.*, 2019) package fits these distributions by the RS algorithm (HOSSAIN *et al.*, 2016).

A parameter recovery study was developed to illustrate the performance of the classical estimates of the ZOILPE models. The goal of this simulation study is to show the behavior of the estimates based on RMSE and the frequentist mean.

The datasets were simulated according to the ZOILPE model. We did not take any covariate and, we chose different values for $\mu = (-1.5, 0, 1.5)$, $\sigma = (0.7, 2)$, $p_0 = (0.2, 0.3)$ and $p_1 = (0.2, 0.3)$. The sample sizes are taken as $n = 700, 800$ and 900. Because the ZOILPE is more complex than the models from the previously simulation, we considered only 100 replicates.

The average and the root mean square error (RMSE) from the estimates of each parameter were calculated. The results are all presented in Table 14. The RMSE decreases as the sample size increases. Also, the difference between the mean and the true parameter values is small. We can conclude that the ZOILPE model presented in 5.2 with the estimation method presented by (ENEA *et al.*, 2019) and available at `gamlss.inf` package is a alternative model to fit data bounded on the $(0,1)$ interval and with exacts values 0 and 1.

Table 14 – Classical estimates of the parameters in the ZOILPE model and the root mean square error (RMSE), based on 100 simulated datasets.

Parameter		μ	σ	p_0	p_1	μ	σ	p_0	p_1
	True	-1.5	0.7	0.2	0.2	-1.5	0.7	0.3	0.3
n = 700	Mean	-1.429	0.814	0.203	0.198	-1.442	0.460	0.305	0.298
	RMSE	0.081	0.278	0.020	0.020	0.121	0.269	0.029	0.027
n = 800	Mean	-1.339	0.720	0.204	0.198	-1.390	0.489	0.304	0.298
	RMSE	0.172	0.187	0.019	0.018	0.140	0.236	0.027	0.026
n = 900	Mean	-1.443	0.661	0.204	0.198	-1.497	0.506	0.304	0.298
	RMSE	0.067	0.137	0.018	0.017	0.103	0.220	0.025	0.024
	True	0	0.7	0.2	0.2	0	0.7	0.3	0.3
n = 700	Mean	0.036	0.824	0.203	0.197	0.049	0.372	0.300	0.299
	RMSE	0.048	0.285	0.020	0.020	0.051	0.470	0.027	0.026
n = 800	Mean	0.137	0.715	0.204	0.198	-0.323	0.463	0.295	0.306
	RMSE	0.147	0.182	0.019	0.018	0.348	0.251	0.020	0.022
n = 900	Mean	0.054	0.661	0.204	0.198	0.015	0.508	0.304	0.298
	RMSE	0.065	0.138	0.018	0.017	0.104	0.218	0.025	0.024
	True	1.5	0.7	0.2	0.2	1.5	0.7	0.3	0.3
n = 700	Mean	1.526	0.812	0.203	0.197	1.566	0.460	0.305	0.298
	RMSE	0.044	0.271	0.020	0.020	0.125	0.269	0.029	0.027
n = 800	Mean	1.648	0.714	0.204	0.198	1.611	0.485	0.304	0.298
	RMSE	0.158	0.182	0.019	0.018	0.140	0.241	0.027	0.026
n = 900	Mean	1.547	0.661	0.204	0.198	1.501	0.506	0.304	0.298
	RMSE	0.059	0.138	0.018	0.017	0.103	0.220	0.025	0.024
	True	-1.5	2	0.2	0.2	-1.5	2	0.3	0.3
n = 700	Mean	-1.337	2.232	0.203	0.198	-1.398	0.650	0.303	0.296
	RMSE	0.194	0.678	0.020	0.020	0.118	1.461	0.029	0.022
n = 800	Mean	-1.023	2.056	0.204	0.198	-1.494	3.183	0.296	0.298
	RMSE	0.510	0.529	0.019	0.018	0.355	1.834	0.026	0.028
n = 900	Mean	-1.328	1.886	0.204	0.198	-1.495	1.446	0.304	0.298
	RMSE	0.200	0.395	0.018	0.017	0.294	0.627	0.025	0.024
	True	0	2	0.2	0.2	0	2	0.3	0.3
n = 700	Mean	0.104	2.223	0.203	0.197	0.135	1.066	0.301	0.298
	RMSE	0.141	0.669	0.020	0.020	0.140	1.339	0.028	0.023
n = 800	Mean	0.323	2.034	0.204	0.198	0.735	1.677	0.298	0.304
	RMSE	0.351	0.517	0.019	0.018	0.861	0.591	0.024	0.029
n = 900	Mean	0.142	1.891	0.204	0.198	0.010	1.451	0.304	0.298
	RMSE	0.174	0.392	0.018	0.017	0.293	0.622	0.025	0.024
	True	1.5	2	0.2	0.2	1.5	2	0.3	0.3
n = 700	Mean	1.591	2.233	0.203	0.197	1.841	0.763	0.294	0.299
	RMSE	0.133	0.676	0.020	0.020	0.494	1.413	0.018	0.036
n = 800	Mean	1.904	2.058	0.204	0.198	0.574	1.329	0.295	0.306
	RMSE	0.434	0.532	0.019	0.018	0.998	0.708	0.020	0.022
n = 900	Mean	1.628	1.888	0.204	0.198	1.502	1.447	0.304	0.298
	RMSE	0.163	0.395	0.018	0.017	0.294	0.627	0.025	0.024

CHAPTER

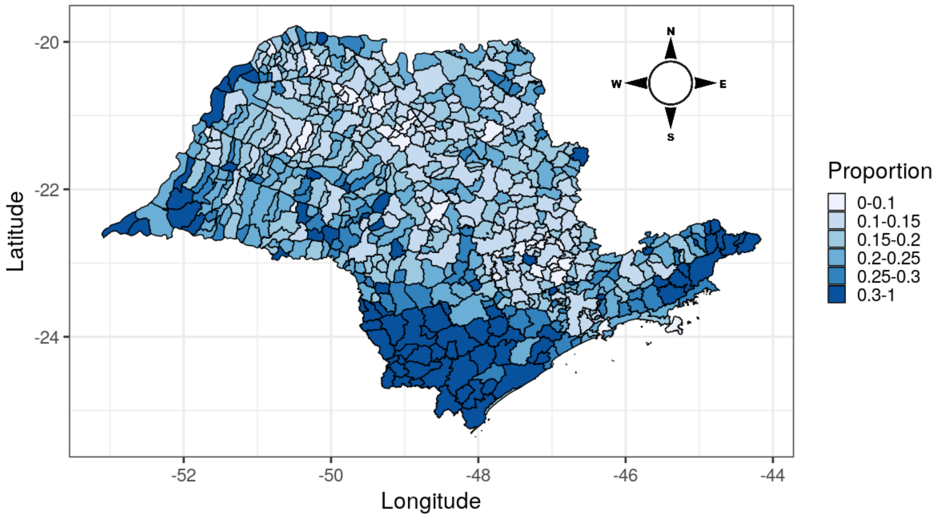
6

APPLICATIONS

6.1 Proportion of individuals vulnerable to poverty

The first application illustrated the use of the *JSB* distributions. The dataset contains the proportion of individuals vulnerable to poverty from the 645 municipalities of the State of São Paulo in Brazil (BRAZIL, 2021a). The map of São Paulo state in Figure 40 represents the proportion of people vulnerable to poverty from each city.

Figure 40 – Map of the São Paulo state, made with `rgdal` and `ggplot2` packages in `R` language, for the real proportion of people vulnerable to poverty in each city.



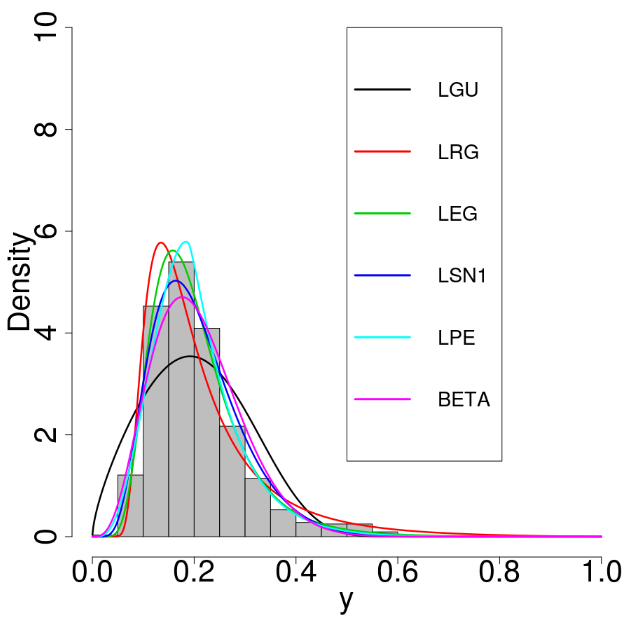
Source: Elaborated by the author.

An individual is classified as vulnerable to poverty if their per capita household income is up to BRL 255. The application goal is to fit the *JSB* distributions and the Beta distribution to the data. The Kurtosis and Skewness computed for the sample were respectively 5.797 and 1.433. These values show that the data has a high value of Kurtosis and moderately positive Skewness. The conclusions from Chapter 2 show that the LGU, LRG, and LSN distributions do not accurately accommodate high Kurtosis data. On the other hand, the LEG and LPE distributions fit high Kurtosis and moderately positive data, so to confirm it, the distributions are compared by the AIC and BIC criteria. This application does not consider any covariate. Six distributions are fitted to the data, the logistic Gumbel, logistic reverse Gumbel, logistic exponential Gaussian, logistic skew normal type I, logistic power exponential, and Beta distributions. The RS algorithm fitted all distributions. The algorithm finds the maximum likelihood estimators for the parameters of the distributions. Table 15 shows the values of the AIC and BIC criteria. Figure 41 exhibits the histogram of the data with the density of the estimated distributions superimposed.

Table 15 – Model comparison criteria.

Model	AIC	BIC
Logistic Gumbel	-1210.5	-1201.5
Logistic reverse Gumbel	-1368.7	-1359.8
Logistic exponential Gaussian	**-1437.8**	**-1424.4**
Logistic skew normal type I	-1415.4	-1402.0
Logistic power exponential	-1429.7	-1416.3
Beta	-1380.2	1371.3

Figure 41 – Histogram of the data with the estimated distribution density superimposed.



Source: Elaborated by the author.

The logistic exponential Gaussian distribution produces the best fitting to the random variable according to the criteria. It is possible to see from 41 that the *JSB* distributions presented long tails and high kurtosis. These characteristics make the *JSB* more flexible to the data. Another point is the advantage that distribution with tree parameters concern with distributions with two parameters.

Table 16 presents the location parameter μ, dispersion parameter σ, the expected value and variance of the random variable, and the probability that the random variable assumes higher values than 0.3256 (Brazil mean of proportion individuals vulnerable to poverty).

Table 16 – Maximum Likelihood estimates, E(Y), Var(Y) and P(Y $\geq$ 0.3256) from Logistc Exponential Gaussian.

μ	σ	ν	E(Y)	Var(Y)	P(Y $\geq$ 0.3256)
-1.7757	0.4201	0.3431	0.2058	0.0083	0.0143

The results from 16 show that the expected value from the cities of São Paulo state for the proportion of individuals vulnerable to poverty are smaller than Brazil mean of proportion individuals vulnerable to poverty. Since the calculated μ was low, -1.77, $E(Y)$ also displayed a low value because of the relation between the location parameter and $E(Y)$. The probability of a city presented a value higher than the Brazil mean is 0.0143. The results show that the São Paulo cities have a lower index of poverty compared to Brazilian cities.

6.2 The votes of a brazilian political party

The data contains the proportion of votes obtained by a political party in the Brazilian presidential elections. The response variable is the proportion of votes get by the Workers Party (PT in Portuguese) in five presidential elections, every four years, from 1994 to 2010, in the 75 cities of the Sergipe state from Brazil. This data was introduced by Paz, Ehlers and Bazán (2015). The aim is to identify the time effect at the proportion of votes obtained by the PT. Additionally, the MHDI (Municipal Human Development Index) was taken as a covariate. The Brazilian MHDI is comprised of the same three dimensions as the Global HDI – longevity, education, and income –, but goes beyond them; it adapts the global methodology to the Brazilian context and the availability of national indicators (BRAZIL, 2021b). The MHDI is divide into categories presented in (BRAZIL, 2021a): between 0.5 and 0.599: low $MHDI_1$, between 0.6 and 0.699: medium $MHDI_2$. The only city with high MHDI was removed from the fitting of the model, but it stays in the exploratory analysis. Table 17 display the data set for six different cities. Figure 42 shows the map of the Sergipe state. The map represents the proportion of votes every year.

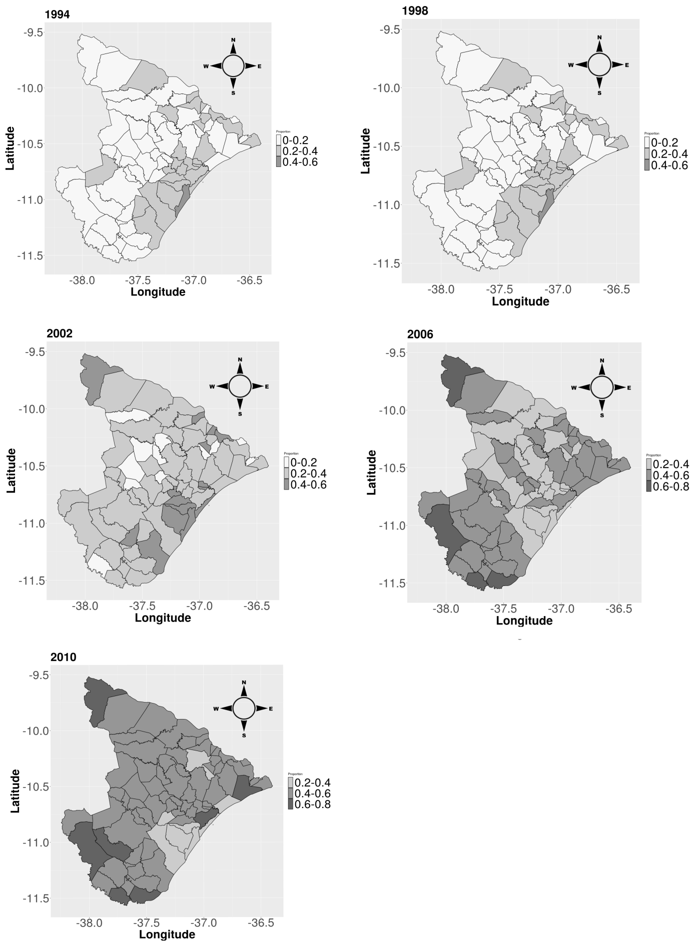
Table 17 – Proportion of votes for six cities from Sergipe state in five presidential election.

Id	1994	1998	2002	2006	2010	MHDI
1	0.23	0.29	0.31	0.49	0.57	0.61
2	0.17	0.19	0.27	0.37	0.37	0.58
3	0.43	0.43	0.52	0.37	0.32	0.77
4	0.10	0.11	0.27	0.43	0.51	0.59
5	0.16	0.15	0.22	0.37	0.37	0.58
6	0.33	0.39	0.46	0.32	0.36	0.65

The Kurtosis and Skewness computed for the sample were respectively 2.014 and 0.13. These values show that the data has a low value of Kurtosis and almost 0 for Skewness. Because of the low value from the Kurtosis, we can consider the distributions LGU, LRG, and LSN. Especially the LSN best accommodates data with low values of Skewness. The distributions are compared by the AIC, BIC, EAIC, EBIC, and WAIC criteria to confirm this conclusion.

A descriptive analysis, described in Table 18, verifies how the proportion of votes behaves over time and concerning the covariate MHDI. The data summarized in Table 18 reveals that the MHDI covariate appears to be relevant because of the difference displayed between the groups. Figure 43 shows that the proportion of votes increases with time, especially in the low MHDI group.

Figure 42 – Map of Sergipe state, make with `rgdal` and `ggplot2`, for the proportion of votes over time.



Source: Elaborated by the author.

For modeling, consider y_{ij}: proportion of votes obtained by the Workers Party in the city $i = 1, \ldots, 74$ at the presidential elections $j = 1, ..5$. The model proposed is given below:

$$Y_{ij} \sim JSB(\mu_{ij}, \tau_{ij}, \nu_{ij})$$

Four different models were build. The models 2, 3 and 4 are mixed models. The model 4 has the Time as covariate in the precision parameter besides the random effect.

Model 1:

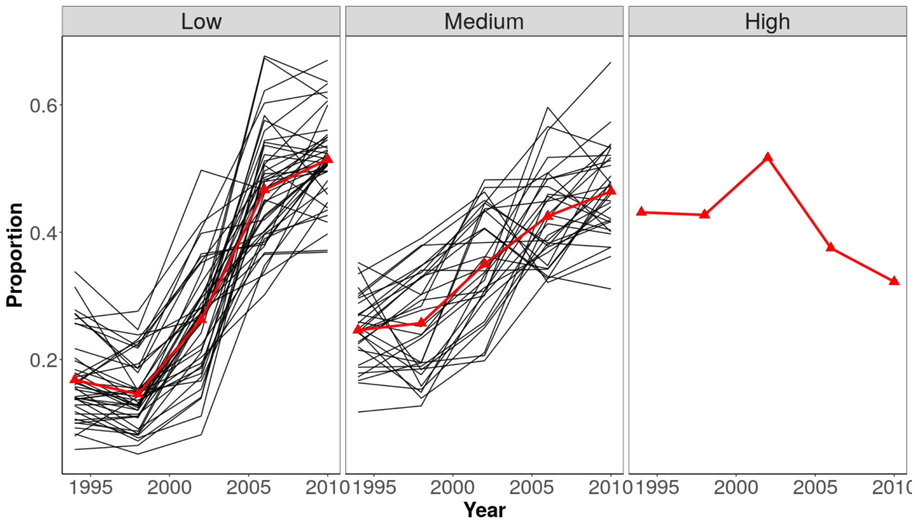
$$\mu_{ij} = \beta_0 + \beta_1 \times (\text{MHDI}_{i2}) + \beta_2 \times (\text{Time}_{ij}) + \beta_3 \times (\text{MHDI}_{i2}) \times (\text{Time}_{ij}) + b_i$$

$$s_2(\tau_{ij}) = \gamma_0 \qquad s_3(\nu_{ij}) = \delta_0.$$

Table 18 – Summary of the descriptive measures separated by sex group.

MHDI	Time	Average	Standard Deviation	Minimum	Median	Maximum	n
High	1	0.43120		0.43120	0.43120	0.43120	1
High	2	0.42720		0.42720	0.42720	0.42720	1
High	3	0.51700		0.51700	0.51700	0.51700	1
High	4	0.37490		0.37490	0.37490	0.37490	1
High	5	0.32220		0.32220	0.32220	0.32220	1
Low	1	0.16801	0.06520	0.05870	0.15720	0.33820	43
Low	2	0.14559	0.05482	0.05150	0.13600	0.27600	43
Low	3	0.26279	0.08682	0.08190	0.26760	0.49740	43
Low	4	0.46673	0.08990	0.30150	0.46670	0.67680	43
Low	5	0.51428	0.06807	0.36880	0.51120	0.67010	43
Medium	1	0.24642	0.05608	0.11760	0.24480	0.35240	31
Medium	2	0.25725	0.08088	0.12730	0.25180	0.39040	31
Medium	3	0.35032	0.08299	0.19810	0.33950	0.48220	31
Medium	4	0.42512	0.07521	0.32100	0.42160	0.59620	31
Medium	5	0.46420	0.07203	0.31080	0.46470	0.66710	31

Figure 43 – The Sergipe city profiles grouped by MHDI classes. The red line represents the average.



Source: Elaborated by the author.

Model 2:

$$\mu_{ij} = \beta_0 + \beta_1 \times (\text{MHDI}_{i2}) + \beta_2 \times (\text{Time}_{ij}) + b_i$$

$$s_2(\tau_{ij}) = \gamma_0 \qquad s_3(\nu_{ij}) = \delta_0.$$

$$b_i \sim N(0, \sigma_b).$$

Model 3:

$$\mu_{ij} = \beta_0 + \beta_1 \times (\text{MHDI}_{i2}) + \beta_2 \times (\text{Time}_{ij}) + \beta_3 \times (\text{MHDI}_{i2}) \times (\text{Time}_{ij}) + b_i$$

$$s_2(\tau_{ij}) = \gamma_0 \qquad s_3(\nu_{ij}) = \delta_0$$

$$b_i \sim N(0, \sigma_b).$$

Model 4:

$$\mu_{ij} = \beta_0 + \beta_1 \times (\text{MHDI}_{i2}) + \beta_2 \times (\text{Time}_{ij}) + \beta_3 \times (\text{MHDI}_{i2}) \times (\text{Time}_{ij}) + b_i$$

$$s_2(\tau_{ij}) = \gamma_0 + \gamma_1 \times (\text{Time}_{ij}) \qquad s_3(\nu_{ij}) = \delta_0$$

$$b_i \sim N(0, \sigma_b).$$

All models are fitted using the RS algorithm within package `gamlss` and NUTS algorithm within `rstan`. Table 19 shows the AIC, BIC, EAIC, EBIC, and WAIC criteria. The LSN model with covariate in parameter τ ($\sigma = 1/\tau$ in the classical approach) and random intercept presented the smallest AIC. However, the Beta model 4 pointed to the best fitting since it presented the smallest BIC, EAIC, EBIC and WAIC. The conclusion from the criteria pointed to the Beta model as the better choice. But the LSN model 4 is also presented to show that the *JSB* mixed regression model can be an alternative even in this kind of situation. The aim is not to compare the model but to show that the LSN model is an alternative. The Beta estimates and residuals are presented under the Bayesian approach, while the LSN estimates are presented under the classical approach.

Table 20 shows the classical and Bayesian estimates, the standard deviation (SD), the confidence intervals (IC) for classical estimates, and the high posterior density interval (HPD) for Bayesian estimates. We can conclude from the estimates is that time has a positive impact on the proportion of votes, revealing an increase in the proportion of votes through the years. However, the growth trend is bigger in cities with low MHDI than in cities with medium MHDI since the interaction between MHDI and time was negative. We can see that time was relevant to consider as a covariate to parameter σ since there is a decrease in AIC. It is important to report that the estimates from the Beta model and the LSN model are similar.

In summary, the random intercept helped to improve the adjustment of the model to data. The QQplot of the randomized quantile residuals was checked to see if the distributions are adequate to the data. The envelope for the randomized quantile residuals is used to verify if the residuals are consistent with the fitted model. Considering the QQplot in the Figures 44a and 44c, the residuals show normality. Envelopes from 44b and 44d show that models are well fitted to the data.

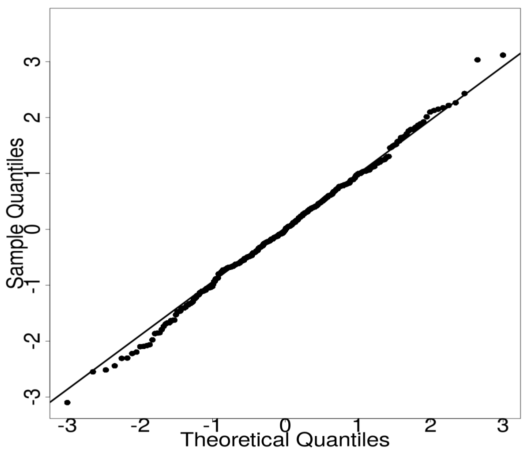
Table 19 – The AIC, BIC, WAIC, EBIC and EAIC criterion for all models fitted under the classical and Bayesian approach.

		Distribution					
Model		LG	LRG	LEG	LSN1	LPE	Beta
1	EAIC	-742.8	-699.7	-777.8	-775.6	-781.9	-802.3
	EBIC	-693.7	-650.6	728.7	-728.5	-734.8	-753.2
	WAIC	-746.0	-702.3	-783.0	782.7	-788.1	-807.6
	AIC	-742.9	-704.7	-781.0	-781.1	-787.1	-807.4
	BIC	-728.3	-685.1	-757.5	-757.6	-764.4	-787.9
2	EAIC	-766.78	-763.2	-762.3	-763.6	-757.1	-772.04
	EBIC	-717.65	-714.1	-713.1	-714.47	-709.9	-722.9
	WAIC	-731.1	-715.8	-743.0	744.3	-740.8	-757.4
	AIC	-734.0	-724.2	-753.7	-753.6	-745.6	-759.9
	BIC	-625.1	-536.5	-624.6	-624.7	-630.3	-667.25
3	EAIC	-815.89	-830.9	-830.3	-828.8	-825.68	-846.44
	EBIC	-768.75	-783.7	-783.14	-781.65	-780.5	-799.31
	WAIC	-775.0	-783.8	-802.0	-800.9	-802.3	**-823.3**
	AIC	-775.9	-792.4	-813.6	-813.7	-802.9	-826.0
	BIC	-775.0	-784.6	-801.5	-802.0	-802.3	-823.2
4	EAIC	-834.29	-836.0	-840.69	840.0	-823.7	**-846.9**
	EBIC	-789.15	-790.9	-795.54	-794.9	-780.5	**-801.7**
	WAIC	-791.0	-793.7	-817.3	-816.4	-815.8	823.2
	AIC	-767.7	-796.4	-829.1	**-829.5**	-817.9	-826.2
	BIC	-791.0	-793.4	-817.2	-816.6	-816.2	**-823.2**

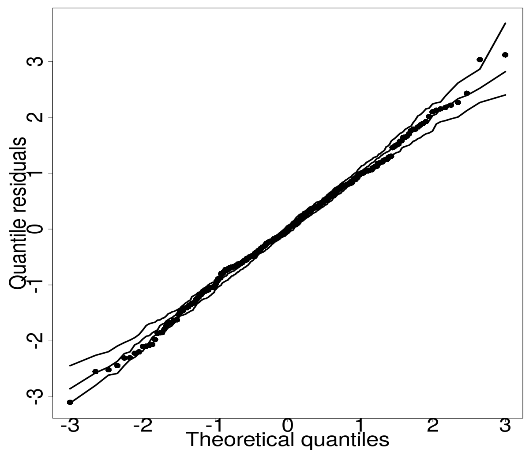
Table 20 – Estimates, standard deviation (SD) and confidence intervals (IC) of the LSN model 4. Estimates, standard deviation (SD) and high posterior density (HPD) of the fixed of the Beta model 4.

	Penalized Likelihood estimates (LSN)			Bayesian posterior estimates (Beta)		
Parameter	Estimate	SD	95% IC	Estimate	SD	95% HPD
β_0	-2.04	0.63	(-2.11;-198)	-1.89	0.06	(-2.01;-1.78)
Time	0.78	0.08	(0.78;0.80)	0.69	0.08	(0.54;0.86)
$MHDI_2$	0.13	0.02	(0.13;0.14)	0.12	0.01	(0.12;0.14)
Time $\times$ $MHDI_2$	-0.06	0.01	(-0.06;-0.06)	-0.06	0.01	(-0.08;-0.04)
γ_0	-0.62	0.10	(-0.64;-0.62)	-3.26	0.14	(-3.89;-3.16)
γ_1	-0.04	0.01	(-0.04;-0.04)	-0.02	0.01	(-0.05;0.02)
δ_0	0.0767	1.53	(-0.08;0.23)			
σ_b	0.34			0.26		

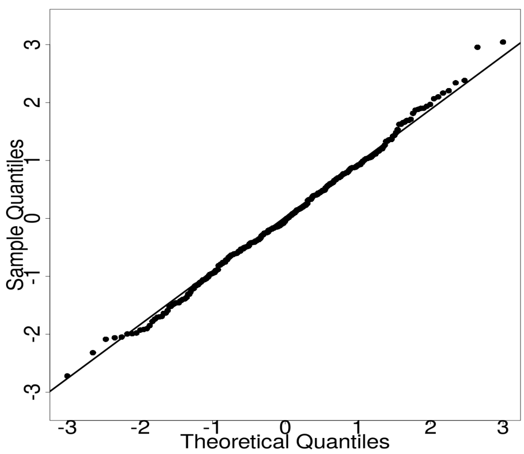
Figure 44 – QQplot and envelope of the randomized quantiles residuals of Model 4 with LSN and Beta distributions.



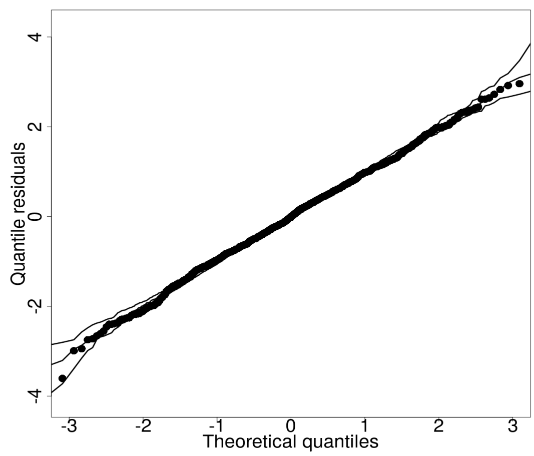
(a) QQplot for the randomized quantile residuals of Model 4 with LSN distribution.



(b) Envelope for the randomized quantile residuals of Model 4 with LSN distribution.



(c) QQplot for the quantile residuals of Model 4 with Beta distribution.



(d) Envelope for the quantile residuals of Model with Beta distribution.

6.3 Mortality rate from bronchial and lung cancer in Brazil over time

6.3.1 Description of the data

In Brazil, lung cancer is the second most prevalent form of cancer in men and women (following non-melanoma skin cancer) and the first worldwide since 1985, both in incidence and mortality. About 85% of diagnosed cases of lung cancer are associated with smoking tobacco(INSTITUTE, 2021a).

The mortality rate measures the number of deaths associated with a factor or cause (disease, violence, epidemic) in a population. The rate is the number of deaths per 100,000 people (the total number of a group can vary between studies). The mathematical expression of mortality rate is (number of deaths/ size of population) $\times$ 100000. A mortality rate of 5 in a city of 1,000,000 inhabitants means 50 deaths in this city.

The dataset of our study comes from the public health records of the Brazilian 27 states (counting the Federal District as a state) (INSTITUTE, 2021b). Specifically, we have the mortality rate of bronchial and lung cancer from the 27 Brazilian states over the last 30 years. The aim is to identify factors that could influence the mortality rate from bronchial and lung cancer over time.

Time can be an important factor. In Brazil, the national public health system was created in 1988, called the Unified Health System (SUS in Portuguese). Different policies have been implemented through this system by each government. Considering this fact, we divide our longitudinal analysis into seven government periods (91-94, 95-98, 99-2002, 2003-2006, 2007-2010, 2011-2014 and 2015-2018), so $j = 1, 2 \ldots, 7$, and the response variable is the average mortality rate for each period.

In this study, following covariates are consider additionally, sex (male or female) as a factor and four age groups (the original data came with these groups), between 50 and 59 years: Age_1, between 60 and 69 years: Age_2, between 70 and 79 years: Age_3, and over 80 years Age_4. The MHDI was also used. The Brazilian MHDI is comprised of the same three dimensions as the Global HDI – longevity, education, and income –, but goes beyond them; it adapts the global method to the Brazilian context and the availability of national indicators. More specifically, one dimension of the MHDI was used, the MHDI Longevity (BRAZIL, 2021c), available at (BRAZIL, 2021b). The MHDI Longevity is more suited to the problem and is classified as a factor where between 0.6 and 0.699 denotes Medium MHDI_1, between 0.7 and 0.799 denotes High MHDI_2, and between 0.8 and 1 is Very High MHDI_3 (BRAZIL, 2021a).

Thus, the dataset of this study corresponds to 27 states, and in each state, there are two groups based on sex and four age groups for seven times intervals. Table 21 displays the dataset for interval 7 (2015-2018 period). Additionally, Figure 45 displays the mortality rate of every age group over time, showing to what extent the time interval is important to explain the mortality rate in the different age groups.

Table 21 – Bronchial and Lung cancer mortality rate by sex, age and state, Brazil in interval 7 (2015-2018).

	Male				Female			
	50-59	60-69	70-79	80-	50-59	60-69	70-79	80-
Acre	26.73	92.27	179.16	249.83	23.11	96.21	132.65	166.18
Alagoas	17.72	48.53	95.48	176.11	16.00	37.79	56.77	73.24
Amapá	21.35	48.70	136.53	259.70	12.44	27.73	56.16	106.55
Amazonas	19.99	76.42	205.12	274.70	13.09	50.50	91.49	135.09
Bahia	14.65	45.95	87.38	131.55	12.35	30.19	39.08	60.98
Ceará	22.26	71.31	155.74	241.11	21.40	55.45	105.41	123.28
Distrito Federal	15.70	66.11	154.84	273.09	12.22	44.79	80.87	134.54
Espírito Santo	22.29	67.34	157.68	215.35	15.81	38.67	61.22	104.35
Goiás	24.43	77.52	181.26	234.52	19.38	48.63	101.87	140.62
Maranhão	14.94	53.77	123.31	151.03	12.21	34.95	61.22	69.03
Mato Grosso	21.28	66.49	131.70	158.34	20.33	36.70	75.32	86.09
Mato Grosso do Sul	27.50	84.38	170.46	248.03	21.72	44.19	77.95	130.20
Minas Gerais	19.96	64.82	127.76	180.92	15.19	36.40	56.96	88.36
Paraná	26.76	83.48	174.08	219.29	20.64	51.33	94.73	122.22
Paraíba	16.62	57.24	108.06	173.41	18.74	38.50	65.58	118.72
Pará	14.77	53.64	115.64	181.91	11.81	31.41	61.38	88.48
Pernambuco	21.75	65.67	119.45	179.30	17.03	39.43	68.28	111.88
Piauí	16.64	63.82	134.48	225.80	14.65	42.11	68.24	106.69
Rio Grande do Norte	22.24	59.01	136.37	178.07	20.58	48.43	67.66	98.69
Rio Grande do Sul	46.58	145.77	300.05	392.74	34.56	72.53	129.14	134.34
Rio de Janeiro	25.56	82.97	170.36	238.41	22.50	48.91	76.28	90.68
Rondônia	16.30	71.20	125.98	212.13	13.17	30.28	86.39	154.17
Roraima	23.51	110.53	253.66	250.00	15.66	56.30	126.90	100.00
Santa Catarina	37.50	130.49	251.31	309.77	28.62	58.52	96.04	109.26
Sergipe	19.46	58.95	114.23	180.76	21.65	31.63	46.28	92.93
São Paulo	25.25	81.35	167.95	237.07	20.73	49.11	76.19	104.66
Tocantins	15.39	60.19	99.07	208.36	18.96	28.53	81.07	122.28

6.3.2 Descriptive analysis

The response variable y is the mortality rate divided by 500 to fit it into the models. The mortality rate is an bounded response, then we can divide it to fit into our models. Tables 22 shows a summary of the response variable with mean, standard deviation, minimum, median, maximum and n representing the number of observations per group. Table 22 reveals that the average of the male response variable is higher than female response variable, but a growth pattern exists for both sexes. Figure 45 shows that the response variable is higher among older groups and also has a higher growth rate among older people.

Figure 45 – The mortality rate plotted for different age groups over time.

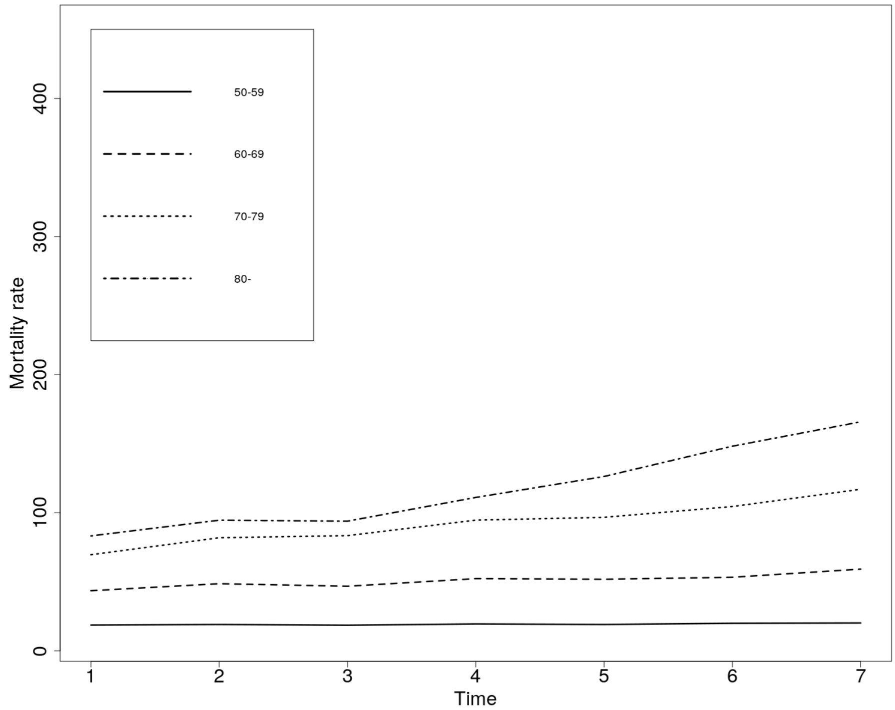
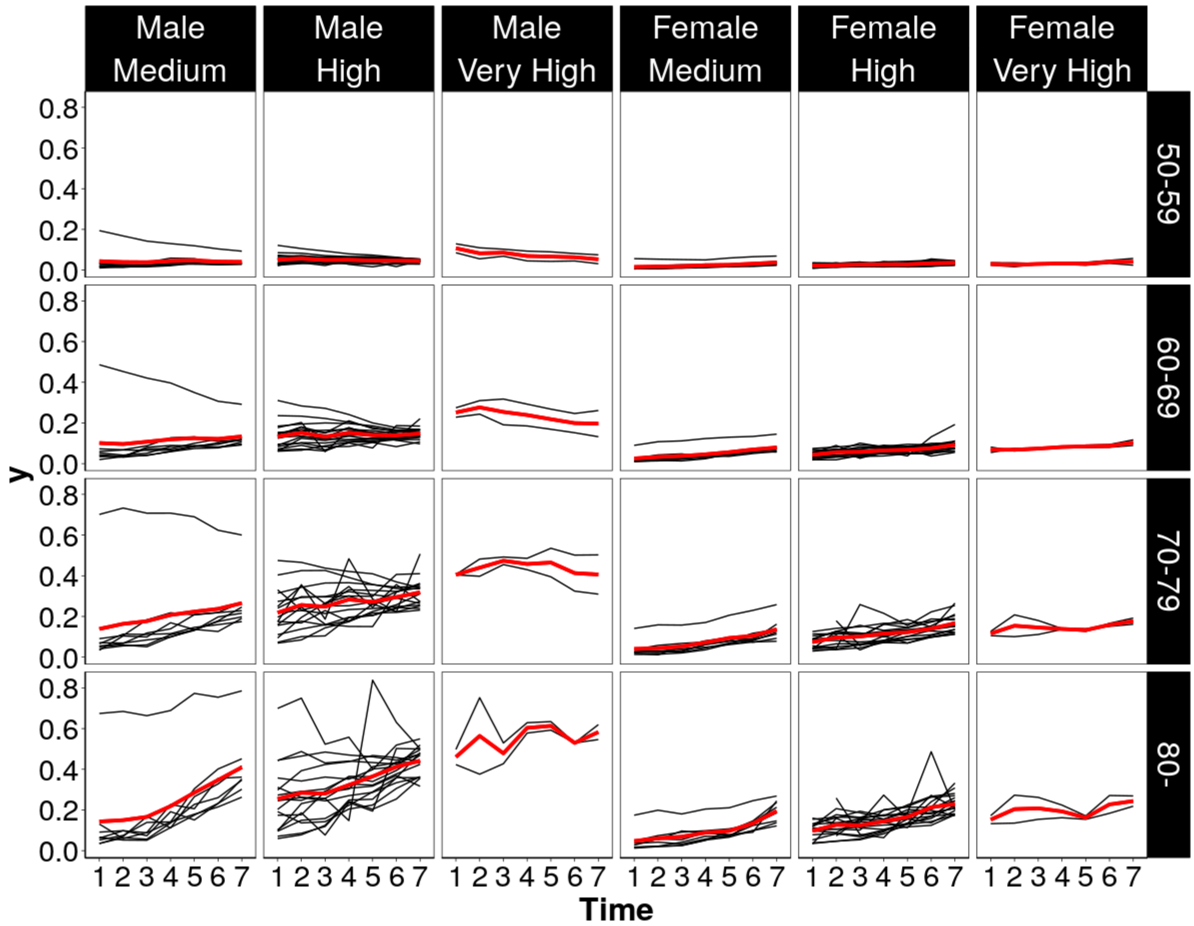


Table 22 – Summary of the descriptive measures separated by sex.

Sex	Time	Average	Standard Deviation	Minimum	Median	Maximum	n
Male	1	0.15787	0.15330	0.01066	0.09271	0.70142	108
Male	2	0.17638	0.16446	0.01366	0.10814	0.75180	108
Male	3	0.17211	0.15001	0.01606	0.11952	0.70760	108
Male	4	0.19647	0.15542	0.02188	0.14941	0.70774	108
Male	5	0.20547	0.16443	0.01624	0.16587	0.83776	108
Male	6	0.21829	0.16124	0.01726	0.17755	0.75396	108
Male	7	0.23599	0.17162	0.02930	0.20713	0.78548	108
Female	1	0.05313	0.04278	0.00582	0.03668	0.17434	102
Female	2	0.06821	0.05712	0.00636	0.05098	0.27286	107
emale	3	0.07064	0.05477	0.00968	0.05378	0.26206	108
Female	4	0.08110	0.05540	0.01166	0.06971	0.27300	108
Female	5	0.08849	0.05597	0.01586	0.07703	0.22492	108
Female	6	0.10770	0.07576	0.01790	0.09159	0.48618	108
Female	7	0.12627	0.07938	0.02362	0.11307	0.33236	108

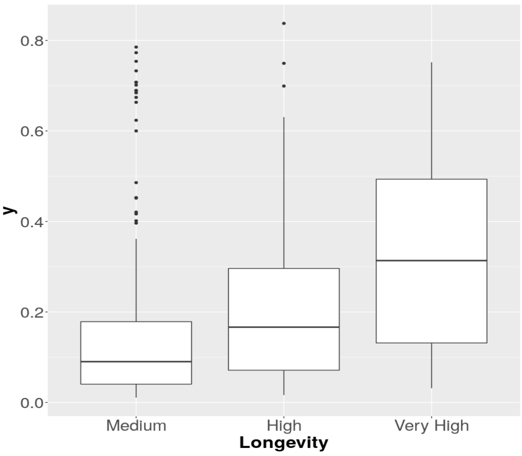
For a graphical display of the response variable's evolution, Figure 46 presents the state profile over time. In summary, Figure 46 does not show evidence of growth in the overall response variable. We note that for some states with medium MHDI, the response variable increases for older groups, although high MHDI states present stagnation.

Figure 46 – Individual and mean profiles of the response variable for states with medium, high, and very high MHDI over time, separated by sex. The red line represents the mean.

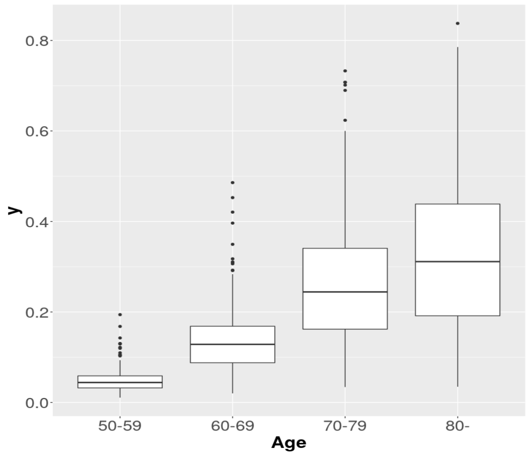


The four box-plots presented in Figure 47 depict the groups concerning the response variable. The top-left box-plot exhibits the men's response variable at every level of MHDI, while the top-right box-plot exhibits the men's response variable at every age level. The left and right bottom box-plots display the women's response variable for every level of MHDI and age, respectively.

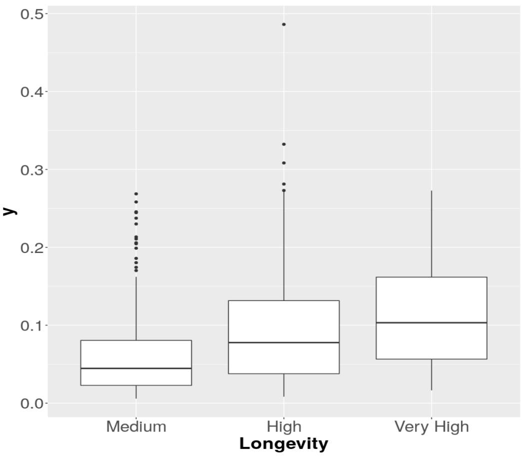
Figure 47 – Box-plots of mortality rate for women, first row, and men, second row, separated by MHDI and age.



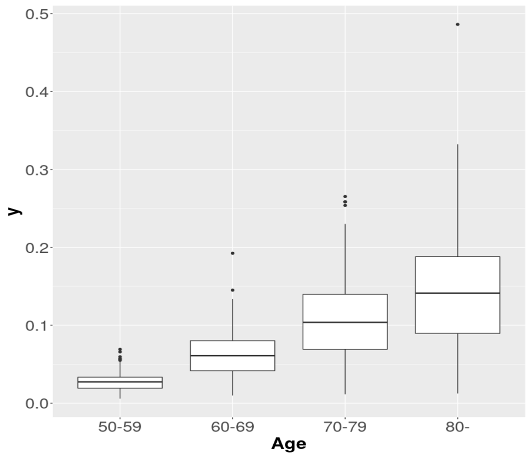
(a) Box-plot of response variable for men in states with medium, high and very high MHDI.



(b) Box-plot of response variable for men in each age group.



(c) Box-plot of response variable for women in states with medium, high and very high MHDI.



(d) Box-plot of response variable for women in each age group.

6.3.3 Model

For modeling, we consider y_{ij} is the mortality rate divided by 500 of state i, $i = 1, \ldots, 27$ in interval $j = 1, ..7$. The model proposed is given below:

$$Y_{ij} \sim JSB(\mu_{ij}, \sigma_{ij}, \nu_{ij})$$

The model's complexity was built in growing order. The first model is the simplest. It assumes a random intercept, and covariates only for the location parameter. Model 4 considers some second-order interactions and the time effect on the dispersion parameter. Model 5 considers a random slope, improving the flexibility and inferences of the model. The models are:

Model 1:

$$\hat{\mu}_{ij} = \hat{\beta}_0 + \hat{\beta}_1 \times \text{Time}_{ij} + \hat{\beta}_2 \times \text{Female}_i + \hat{\beta}_3 \times \text{Age}_{i2} + \hat{\beta}_4 \times \text{Age}_{i3} +$$

$$\hat{\beta}_5 \times \text{Age}_{i4} + \hat{\beta}_6 \times \text{HDI}_{i2} + \hat{\beta}_7 \times \text{HDI}_{i3} + b_i$$

$$s_2(\hat{\sigma}_{ij}) = \hat{\gamma}_0 \qquad s_3(\hat{v}_{ij}) = \hat{\delta}_0.$$

$$b_i \sim N(0, \hat{\sigma}_b).$$

Model 2:

$$\hat{\mu}_{ij} = \hat{\beta}_0 + \hat{\beta}_1 \times \text{Time}_{ij} + \hat{\beta}_2 \times \text{Female}_i + \hat{\beta}_3 \times \text{Age}_{i2} + \hat{\beta}_4 \times \text{Age}_{i3} +$$

$$\hat{\beta}_5 \times \text{Age}_{i4} + \hat{\beta}_6 \times \text{HDI}_{i2} + \hat{\beta}_7 \times \text{HDI}_{i3} + \hat{\beta}_8 \times \text{Female}_i \times \text{Time}_{ij} + b_i$$

$$s_2(\hat{\sigma}_{ij}) = \hat{\gamma}_0 \qquad s_3(\hat{v}_{ij}) = \hat{\delta}_0.$$

$$b_i \sim N(0, \hat{\sigma}_b).$$

Model 3:

$$\hat{\mu}_{ij} = \hat{\beta}_0 + \hat{\beta}_1 \times \text{Time}_{ij} + \hat{\beta}_2 \times \text{Female}_i + \hat{\beta}_3 \times \text{Age}_{i2} + \hat{\beta}_4 \times \text{Age}_{i3} +$$

$$\hat{\beta}_5 \times \text{Age}_{i4} + \hat{\beta}_6 \times \text{HDI}_{i2} + \hat{\beta}_7 \times \text{HDI}_{i3} + b_i$$

$$s_2(\hat{\sigma}_{ij}) = \hat{\gamma}_0 + \hat{\gamma}_1 \times \text{Time}_{ij} \qquad s_3(\hat{v}_{ij}) = \hat{\delta}_0.$$

$$b_i \sim N(0, \hat{\sigma}_b).$$

Model 4:

$$\hat{\mu}_{ij} = \hat{\beta}_0 + \hat{\beta}_1 \times \text{Time}_{ij} + \hat{\beta}_2 \times \text{Female}_i + \hat{\beta}_3 \times \text{Age}_{i2} + \hat{\beta}_4 \times \text{Age}_{i3} + \hat{\beta}_5 \times \text{Age}_{i4} + \hat{\beta}_6 \times \text{HDI}_{i2} +$$

$$\hat{\beta}_7 \times \text{HDI}_{i3} + \hat{\beta}_8 \times \text{Female}_i \times \text{Time}_{ij} + \hat{\beta}_9 \times \text{Age}_{i2} \times \text{Time}_{ij} + \hat{\beta}_{10} \times \text{Age}_{i3} \times \text{Time}_{ij} +$$

$$\hat{\beta}_{11} \times \text{Age}_{i4} \times \text{Time}_{ij} + \hat{\beta}_{12} \times \text{HDI}_{i2} \times \text{Time}_{ij} + \hat{\beta}_{13} \times \text{HDI}_{i3} \times \text{Time}_{ij} + b_i$$

$$s_2(\hat{\sigma}_{ij}) = \hat{\gamma}_0 + \hat{\gamma}_1 \times \text{Time}_{ij} \qquad s_3(\hat{v}_{ij}) = \hat{\delta}_0.$$

$$b_i \sim N(0, \hat{\sigma}_b).$$

Model 5:

$$\hat{\mu}_{ij} = \hat{\beta}_0 + \hat{\beta}_1 \times \text{Time}_{ij} + \hat{\beta}_2 \times \text{Female}_i + \hat{\beta}_3 \times \text{Age}_{i2} + \hat{\beta}_4 \times \text{Age}_{i3} + \hat{\beta}_5 \times \text{Age}_{i4} + \hat{\beta}_6 \times \text{HDI}_{i2} +$$
$$\hat{\beta}_7 \times \text{HDI}_{i3} + \hat{\beta}_8 \times \text{Female}_i \times \text{Time}_{ij} + \hat{\beta}_9 \times \text{Age}_{i2} \times \text{Time}_{ij} + \hat{\beta}_{10} \times \text{Age}_{i3} \times \text{Time}_{ij} +$$
$$\hat{\beta}_{11} \times \text{Age}_{i4} \times \text{Time}_{ij} + \hat{\beta}_{12} \times \text{HDI}_{i2} \times \text{Time}_{ij} + \hat{\beta}_{13} \times \text{HDI}_{i3} \times \text{Time}_{ij} + b_{1i} + b_{2i} \times \text{Time}_{ij}$$
$$s_2(\hat{\sigma}_{ij}) = \hat{\gamma}_0 + \hat{\gamma}_1 \times \text{Time}_{ij} \qquad s_3(\hat{v}_{ij}) = \hat{\delta}_0.$$
$$\boldsymbol{b}_i \sim N(0, \hat{\boldsymbol{D}}).$$

The estimation of parameters was performed according to the classical approach, maximization of penalized likelihood with the RS algorithm available in the `gamlss` package of R. Note that in the classical approach the dispersion parameter, $\sigma = 1/\tau$, was taken. Model 5, with logistic exponential Gaussian distribution, presented problem in the estimation, so we removed it. Under the Bayesian approach, the NUTS algorithms is employed to simulated values from our posterior distribution and get the posterior mean estimators. The posterior sample size was 4000 , with 1000 as burn-in. These sample sizes bring R statistics close to one for the chains. Table 23 shows the following model comparison criteria AIC, BIC, WAIC, EBIC and EAIC for all fitted models.

Model 5 with the logistic power exponential distribution presented the smallest criterion value, and in all of our models displayed smaller criteria than Beta.

Table 24 shows the classic and Bayesian estimates. Men's mortality rate is higher than women's. Age is significant, revealing an increase in the rate as age increases. The MHDI values show that the rate is higher for states with higher MHDI. The interactions considered here are all significant. Mortality is higher among men. However, interaction between sex and time shows that growth is higher among women. Mortality is also higher in states with higher longevity indices, and the interaction shows a decrease between states with MHDI high and very high compared to states with medium MHDI. For the σ coefficients, the effect of time is negative, showing that the variance decreases with time.

The matrix of correlation (C) and covariances ($\boldsymbol{D}$) for random effects from the classical approach:

$$C = \begin{pmatrix} 1 & -0.92 \\ -0.92 & 1 \end{pmatrix}, \quad \boldsymbol{D} = \begin{pmatrix} 0.64 & -0.05 \\ -0.05 & 0.08 \end{pmatrix}, \tag{6.1}$$

and from the Bayesian approach:

$$C = \begin{pmatrix} 1 & -0.90 \\ -0.90 & 1 \end{pmatrix}, \quad \boldsymbol{D} = \begin{pmatrix} 0.61 & -0.04 \\ -0.04 & 0.08 \end{pmatrix}. \tag{6.2}$$

There is a negative correlation between the factors, so the states that present the highest values for the intercepts display the lowest values for the slope. States that have higher mortality

Table 23 – The AIC, BIC, WAIC and EAIC values of all models fitted according to the classical and Bayesian approaches.

		Distribution					
Model		LGU	LRG	LEG	LSN	LPE	Beta
1	EAIC	-5733.45	-5833.45	-6029.91	-6006.07	-6101.72	-5460.22
	EBIC	-5680.25	-5780.28	-5978.74	-5954.90	-6050.55	-5407.06
	WAIC	-5677.30	-5798.70	-6016.0	-5990.30	-6082.80	-5433.0
	AIC	-5714.30	-5814.94	-6012.79	-5938.89	-6086.08	-5442.11
	BIC	-5527.64	-5627.71	-5819.99	-5791.55	-5891.64	-5253.62
2	EAIC	-5814.61	-5872.59	-6064.21	-6042.84	-6127.30	-5477.98
	EBIC	-5761.44	-5819.42	6013.04	-5991.67	-6076.14	-5424.82
	WAIC	-5758.40	-5838.70	- 6047.7	-6024.30	-6106.70	-5447.70
	AIC	-5794.56	-5853.71	-6046.36	-6022.66	-6109.55	-5458.81
	BIC	-5602.34	-5660.18	-5848.28	-5824.93	-5909.86	-5264.97
3	EAIC	-5858.24	-5879.77	-6083.84	-6078.71	-6140.95	-5481.75
	EBIC	-5805.08	-5826.61	-6032.67	-6027.54	-6089.79	-5428.58
	WAIC	-5800.0	-5841.70	-6063.50	-6052.30	-6118.0	-5447.60
	AIC	-5836.99	-5858.50	-6065.41	-6040.80	-6122.22	-5461.50
	BIC	-5639.50	-5661.01	-5862.41	-5838.32	-5917.45	-5262.49
4	EAIC	-6013.32	-6333.52	-6448.69	-6448.31	-6496.75	-5737.60
	EBIC	-5960.15	-6280.35	-6397.52	-6397.15	-6445.58	-5684.12
	WAIC	-5938.20	-6290.01	-6423.8	-6422.60	-6464.70	-5697.0
	AIC	-5987.01	-6307.38	-6424.70	-6332.25	-6502.09	-5711.29
	BIC	-5762.71	-6082.48	-6194.12	-6102.75	-6259.20	-5485.23
5	EAIC	6376.77	-6732.55	-6845.47	-6844.97	**-6866.40**	-6176.80
	EBIC	-6323.61	-6679.39	-6794.30	-6793.81	**-6835.23**	-6123.63
	WAIC	-6268.30	-6658.50	-6797.30	-6796.80	**-6821.90**	-6107.80
	AIC	-6312.48	-6681.20	-	-6731.59	**-6839.93**	-6127.82
	BIC	-5978.97	-6342.44	-	-6388.66	**-6479.87**	-5780.93

rated mostly decline over time, and those with the lowest rates show a growth trend. The random effects reveal the essential role of time in the mortality rate, since the influence of time is not significant in Table 24. Random intercepts and slopes are presented in Table 25. In summary, the results show a pattern of deaths from bronchial and lung cancer in Brazil whereby states with high mortality already reached the peak, and the ensuing trend is decline in the death rate. About the states with a medium MHDI, the direction is a slight increase at a steady rate, and the effect of interaction between time and MHDI is a negative for high and very high MHDI, and a positive or 0 value for medium MHDI.

6.3.4 Residual Analysis

We computed the randomized quantile residuals for the selected model and determined the envelope for them following the method proposed in Section Chapter 4. The computed randomized quantile residuals for the classical approach were -0.004 as mean, -0.675 as the first quantile, 0.683 as the third quantile, -3.842 as the minimum and 3.001 as the maximum. The

Table 24 – Classical and Bayesian estimates, standard deviation (SD), p-value and the high posterior density (HPD) of the model 5 with logistic power exponential distribution.

	Penalized Likelihood Estimation			NUTS		
Parameter	Estimate	SD	p-value	Estimate	SD	95% HPD
Intercept	-2.46	0.06	0	-2.53	0.25	(-8.42; -7.49)
Female	-1.24	0.03	0	-1.22	0.04	(-1.17; -1.06)
Time	0.02	0.01	0.15	0.03	0.03	(-0.04; 0.08)
Age_2	0.82	0.04	0	0.81	0.05	(0.68; 0.84)
Age_3	1.21	0.04	0	1.25	0.06	(1.09; 1.26)
Age_4	1.30	0.04	0	1.29	0.06	(1.12; 1.30)
$MHDI_2$	0.95	0.03	0	0.99	0.31	(0.36; 1.46)
$MHDI_3$	1.72	0.06	0	1.78	0.57	(0.60; 2.58)
Female $\times$ Time	0.08	0.01	0	0.08	0.01	(0.07; 0.09)
Time $\times$ Age_2	0.04	0.01	0	0.04	0.01	(0.02; 0.05)
Time $\times$ Age_3	0.10	0.01	0	0.10	0.01	(0.06; 0.09)
Time $\times$ Age_4	0.16	0.01	0	0.15	0.01	(0.10; 0.14)
Time $\times$ $MHDI_2$	-0.11	0.01	0	-0.12	0.04	(-0.18; -0.05)
Time $\times$ $MHDI_3$	-0.19	0.01	0	-0.20	0.07	(-0.31; -0.05)
γ_0	-0.87	0.06	0	0.88	0.06	(0.98; 1.20)
γ_1	-0.06	0.01	0	0.06	0.01	(0.05; 0.10)
δ_0	-0.011	0.04	0.80	0.08	0.05	(0.05; 0.26)

computed randomized quantile residuals for the Bayesian model were -0.040 as average, -0.636 as the first quantile, 0.696 as the third quantile, -3.19 as the minimum, and 4.34 as the maximum.

After that, the QQplot of the randomized quantile residuals was checked. The QQplot can verify normality from the residuals and consequently check if the distribution is adequate for the data. Considering the QQplots in Figures 48a and 48b, the residuals show normal distribution, as expected.

Additionally, the envelope of the residuals was displayed. The envelope is often used to decide if the observed residuals are consistent with the fitted model. The envelopes from Figures 48c and 48d show that the model is well fitted to the data. The QQplot and envelope of model 5 with Beta distribution for the classical approach are displayed. Figures 49a and 49 show that the distribution is not adequate for the data. The `gamlss` package offers another tool to residuals diagnostic. The single worm plot suggested by the package is equivalent to the normal QQplot, detrended by subtracting the line (intercept 0 and slope 1) (STASINOPOULOS *et al.*, 2017). Figure 50 display the worm plot for model 5 with the LPE distribution.

The horizontal dotted line of the worm plot represents the expected values of the ordered residuals. The residuals are represented in the figure by the points. The two elliptic curves in the figure are the 95% confidence intervals. The worm plot in Figure 50 shows that the points are closer to the horizontal line and are outside of the two elliptic curves. The conclusions are that the residual's distribution is close to a standard normal distribution, and the specification of the model is correct.

Table 25 – Posterior summaries for model 5 with logistic power exponential of randoms effects (Mean: mean, SD: standard deviation, lower CI: lower value of 95% credibility interval, upper CI: upper value of 95% credibility interval).

	Random Intercept				Random Slope			
States	Mean	SD	lower CI	upper CI	Mean	SD	lower CI	upper CI
Ceará	-1.046	0.187	-1.425	-0.697	0.172	0.027	0.122	0.228
Rio Grande do Norte	-0.905	0.195	-1.259	-0.483	0.109	0.028	0.057	0.166
Maranhão	-0.824	0.263	-1.347	-0.335	0.087	0.036	0.017	0.159
Paraíba	-0.816	0.263	-1.319	-0.29	0.111	0.036	0.038	0.177
Acre	-0.652	0.209	-1.035	-0.224	0.129	0.031	0.068	0.191
Pernambuco	-0.536	0.185	-0.905	-0.178	0.058	0.025	0.007	0.105
Piauí	-0.482	0.259	-0.991	0.023	0.085	0.035	0.016	0.155
Pará	-0.416	0.188	-0.775	-0.045	0.002	0.027	-0.052	0.052
Tocantins	-0.267	0.283	-0.811	0.293	0.042	0.041	-0.031	0.129
Alagoas	-0.142	0.263	-0.687	0.369	-0.019	0.036	-0.091	0.052
Rondônia	-0.13	0.196	-0.503	0.259	0.01	0.029	-0.046	0.064
Santa Catarina	-0.074	0.509	-1.081	0.857	0.041	0.064	-0.086	0.161
Roraima	-0.033	0.214	-0.425	0.404	0.021	0.034	-0.042	0.092
Minas Gerais	0.004	0.19	-0.35	0.394	-0.039	0.028	-0.094	0.013
Bahia	0.014	0.254	-0.455	0.542	-0.066	0.034	-0.132	0.002
Distrito Federal	0.061	0.51	-0.92	1.091	-0.039	0.064	-0.165	0.089
Mato Grosso	0.075	0.194	-0.282	0.476	-0.023	0.027	-0.075	0.033
Goiás	0.206	0.188	-0.158	0.571	-0.006	0.027	-0.059	0.046
Espírito Santo	0.269	0.186	-0.08	0.642	-0.045	0.026	-0.092	0.007
Mato Grosso do Sul	0.302	0.19	-0.067	0.682	-0.023	0.027	-0.076	0.028
Amazonas	0.303	0.198	-0.077	0.696	-0.017	0.03	-0.075	0.038
Sergipe	0.388	0.256	-0.073	0.923	-0.062	0.035	-0.133	0.003
Amapá	0.455	0.197	0.071	0.856	-0.102	0.03	-0.163	-0.048
Sâo Paulo	0.616	0.203	0.231	1.014	-0.078	0.03	-0.136	-0.022
Paraná	0.648	0.184	0.303	1.021	-0.066	0.026	-0.117	-0.018
Rio de Janeiro	0.785	0.202	0.39	1.188	-0.094	0.029	-0.152	-0.037
Rio Grande do Sul	2.186	0.276	1.646	2.732	-0.188	0.04	-0.264	-0.109

Figure 50 – Worm plot of the model 5 with logistic power exponential distribution.

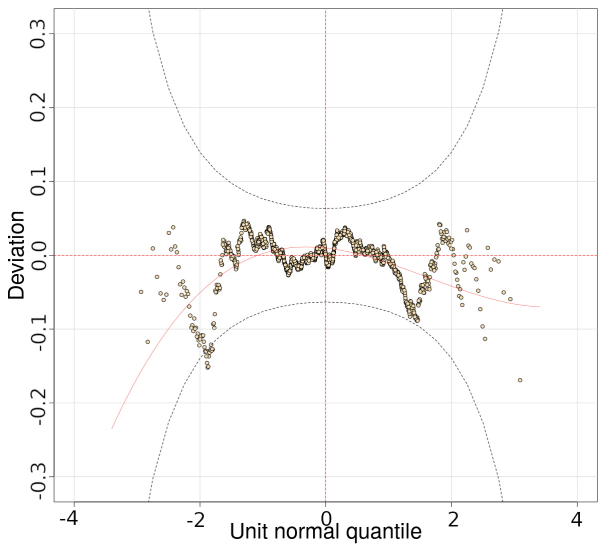
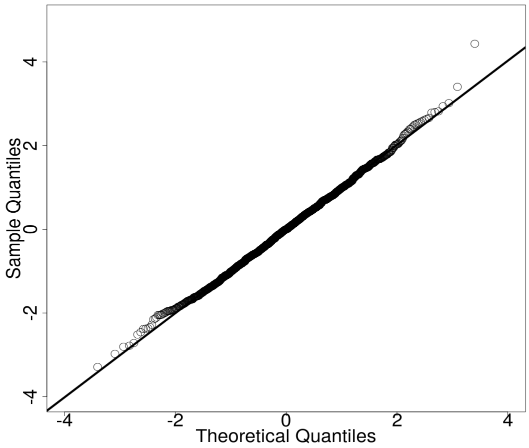
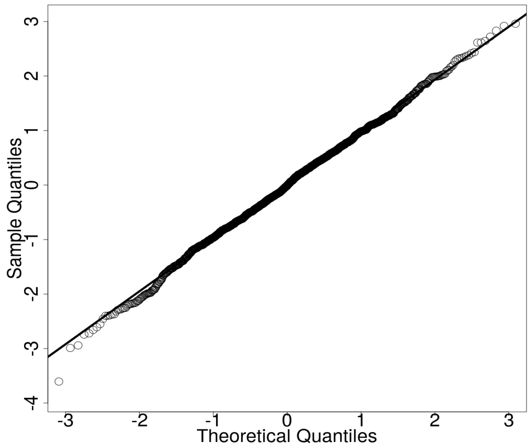


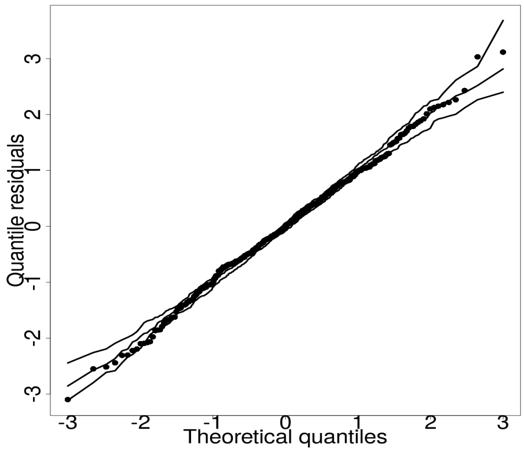
Figure 48 – QQplot and envelope of the randomized quantile residuals of model 5 with logistic power exponential distribution for classical and Bayesian approach.



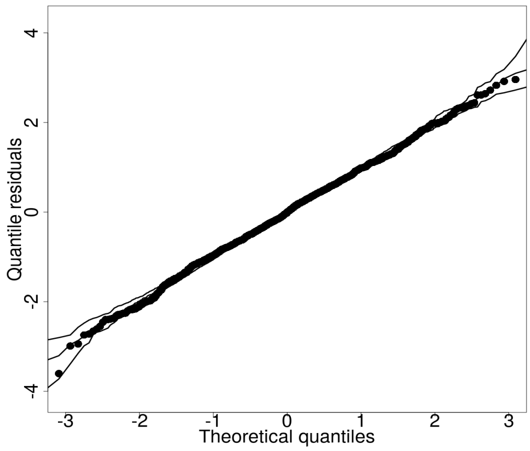
(a) QQplot of the randomized quantile residuals of model 5 from the classical approach.



(b) QQplot of the randomized quantile residuals of model 5 from the Bayesian approach.



(c) Envelope of the randomized quantile residuals of model 5 from the classical approach.



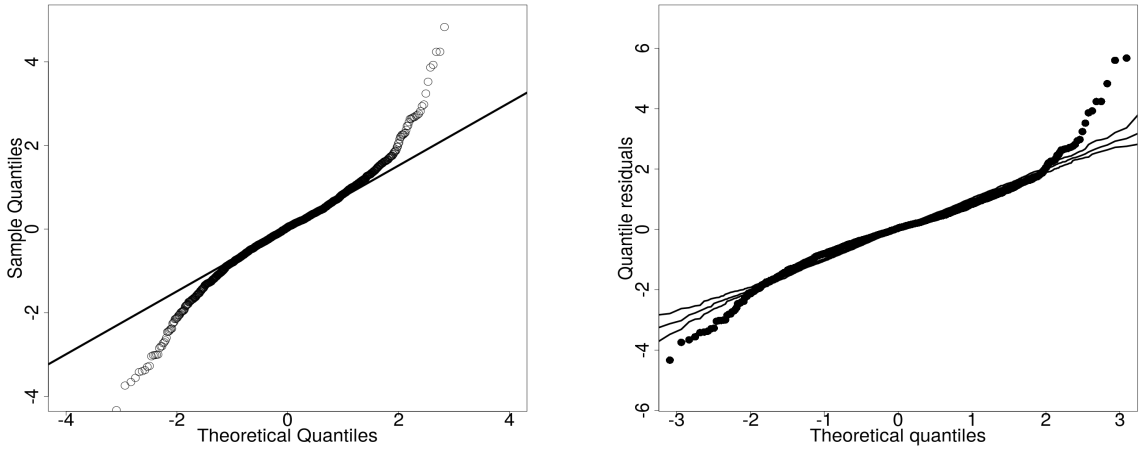
(d) Envelope of the randomized quantile residuals of model 5 from the Bayesian approach.

6.3.5 Posterior predictive distribution

In Bayesian analysis, after the data y have been observed, we can predict an unknown observable variable, $\tilde{y}$. The distribution of $\tilde{y}$ is called the posterior predictive distribution. In practice, we are most often interested in simulating draws from the posterior predictive distributions. Since we have a sample from the posterior distribution, it is typically drawn from the predictive distribution of unobserved or future data $\tilde{y}$ (GELMAN *et al.*, 2013). The application does not have future or unobserved data, so the posterior predictive distribution was evaluated for the observed data. For predictive purposes, this method can be uninformative about the model's accuracy since it predicts the data used for the fitting. However, for the model checking, it can be useful.

For each draw s, with $s = 1, \ldots, 3000$ of the posterior distributions of the parameters (Θ, $\boldsymbol{b}$, $\boldsymbol{D}$) it was draw one value y^{rep} from the logistic power exponential. The resulting vectors of

Figure 49 – QQplot and envelope of the randomized quantiles residuals of model 5 with Beta distribution for classical approach.



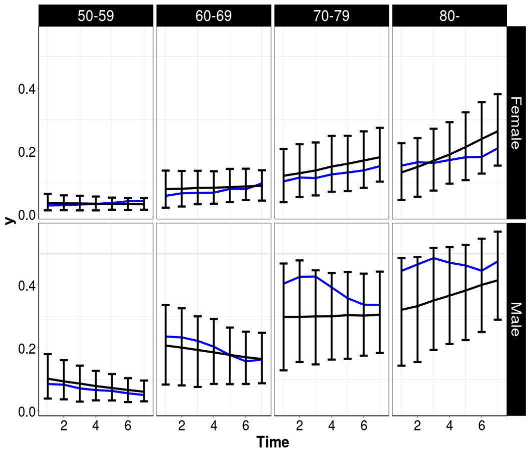
(a) QQplot of the randomized quantile residuals of model 5.

(b) Envelope of the randomized quantile residuals of model 5.

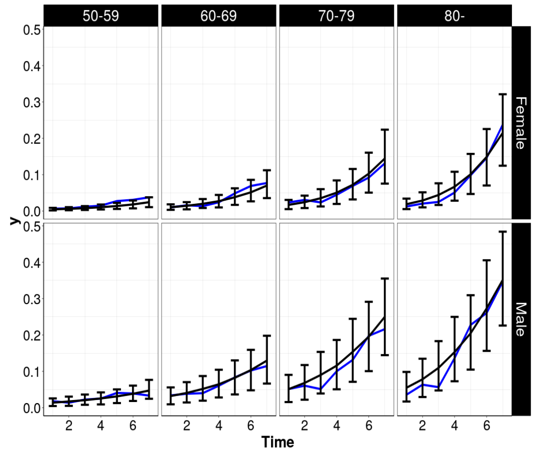
$\boldsymbol{y}^{rep}$ with size 3000 characterized the posterior predictive distribution. To obtain a point estimate of the observed data, we can take the average of each vector $\boldsymbol{y}^{rep}$.

The Figure 51 present the averages of some vectors y^{rep} and the HPD intervals computed by the package `coda`. The predict values were separated by states for better visualization. The conclusion from Figure 51 is that the HPD intervals cover all real values, and the estimates are similar to the real values.

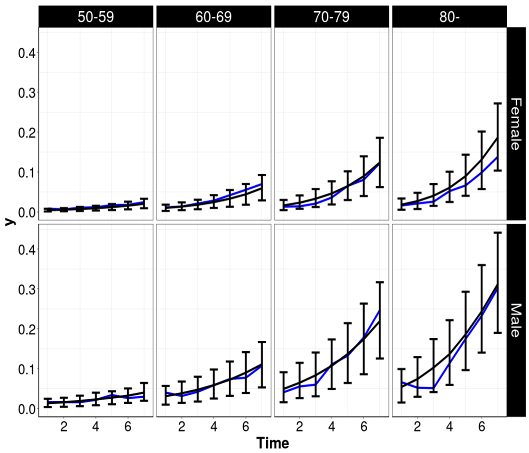
Figure 51 – Predict values with HPD intervals in black and real values in blue separately by state, sex and age.



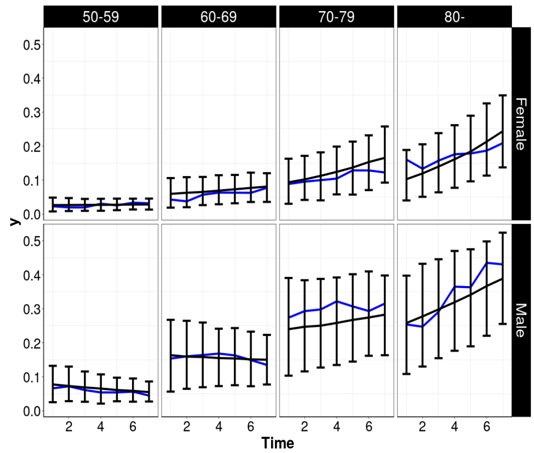
(a) Predict values with HPD intervals in black and real values in blue for state of São Paulo separately sex and age.



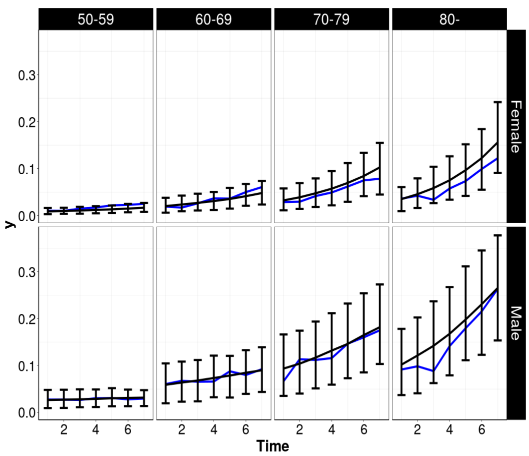
(b) Predict values with HPD interval in black and real values in blue for state of Paraíba separately sex and age.



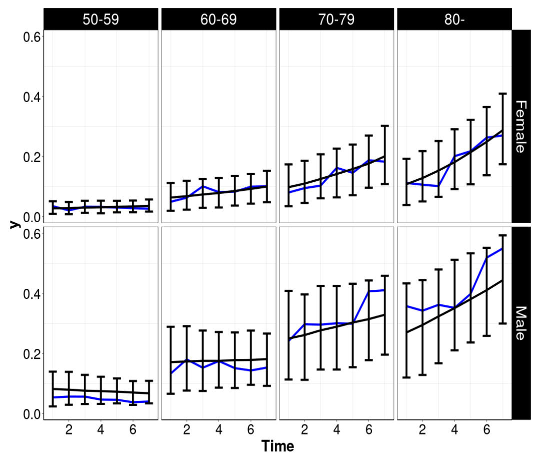
(c) Predict values with HPD interval in black and real values in blue for state of Maranhão separately sex and age.



(d) Predict values with HPD interval in black and real values in blue for state of Espirito Santo separately sex and age.



(e) Predict values with HPD interval in black and real values in blue for state of Bahia separately sex and age.



(f) Predict values with HPD interval in black and real values in blue for state of Amazonas separately sex and age.

CHAPTER

7

FINAL REMARKS

7.1 Final remarks

The JSB class of distributions introduced previously by (RODRIGUES; BAZáN; SUZUKI, 2019) is a class of bounded continuous distributions built through the convolution from a baseline cumulative density function with support on the real line and the quantiles having of the standard logistic distribution. In this paper, the focus is distributions of the JSB class. The central motivation for the new distributions studied here is a new mixed regression model based on JSB distributions, which has not been proposed in the previous literature, so it is a new alternative for the mixed regression models using the Beta distribution.

In the classical approach, the RS algorithm, introduced by Rigby and Stasinopoulos (2005) for the estimation of the GAMLSS class of model, was employed for maximization of the penalized likelihood. In the Bayesian approach, the NUTS algorithm developed by Hoffman and Gelman (2014), which can be considered an extension of the HMC, was adopted to simulate samples from the posterior.

Two simulations have been regarded in this work. The first revealed the behavior of the distributions presented in this work concerning samples with outliers. The first simulation concluded that some distributions from JSB class had better performance for samples with outliers than the Beta model. The second simulation was an extension of JSB class to support zero-and-one-inflated data (OSPINA; FERRARI, 2012). The simulation showed that JSB distribution could be a useful alternative to fit such inflated data.

Three applications were presented. Application 1 contemplate social data, where the proportion of individuals vulnerably to poverty in the municipalities of the state of São Paulo, Brazil, from 2010 is studied. This application was realized without covariates to regard the fitting of the JSB distribution in a real dataset. Application 2 studies a dataset that contains the proportion of votes obtained by a political party in the Brazilian presidential elections. The proportion of votes is modeled in the function of time and HDI. We adopt mixed regression models proposed in this work and the Beta mixed model. We detect that the Beta model seems to fit better than the distributions from JSB class. In the last application, a real dataset involves the mortality rate from bronchial and lung cancer in Brazilian states for the last 30 years. The effects considered in this study were sex, age, and state MHDI. The *JSB* mixed regression models presented here displayed lower model comparison criteria values than the Beta mixed model. The models presented here can be alternatives to the Beta model. As pointed by (CANCHO; BAZÁN; DEY, 2020), the results of this kind of study have exploratory characteristics and should be updated as new data become available. This screening can play a relevant role in new studies of the aspects of interest and support clinical studies to identify predictive biomarkers in cancer research (PEREZ-GRACIA *et al.*, 2017).

An important thing to signalize is the parametric space of the parameters is the same as baseline distributions. However, it is possible to change to the $(0,1)$ scale as presented by Lemonte and Bazán (2016) and Bayes, Bazán and Castro (2017). A paper was submitted considering the new results obtained from this work. The paper contains the classical approach, the results from the simulation, and the third application.

7.2 Suggestions and Future work

As future proposals, to extend the models described here to the case of zero-and-one inflated data is a future work as presented in Hossain *et al.* (2016). This extension is important since many real data sets in the $(0,1)$ interval can contain exact values 0 and 1. The residual analysis using the randomized quantiles residuals is crucial for the fitting of the model. With this analysis, we can check the distribution adequacy. The models can also support spatial data. An extension to a spatial model would allow fit spatial data that contains a bounded response variable. The use of alternative Bayesian estimation procedures is an important factor. Alternative procedures would allow the reduction of the fitting process or even better estimates.

The next steps are to advance in another paper with results from Chapter 2, extending the properties and definitions from JSB class. Develop an R package that contains all models presented here with both classical and Bayesian approaches. An R package is relevant since the package is an auxiliary tool that helps the fitting of these models.

BIBLIOGRAPHY

AKAIKE, H. Likelihood of a model and information criteria. **Journal of econometrics**, Elsevier, v. 16, n. 1, p. 3–14, 1981. Citation on page 75.

ANDERSON, D.; BURNHAM, K. Model selection and multi-model inference. **Second. NY: Springer-Verlag**, v. 63, n. 2020, p. 10, 2004. Citation on page 76.

ATKINSON, A. C. Diagnostic regression analysis and shifted power transformations. **Technometrics**, v. 25, p. 23–33, 1983. Citation on page 78.

AZZALINI, A. A class of distributions which includes the normal ones. **Scandinavian journal of statistics**, p. 171–178, 1985. Citations on pages 33 and 49.

BAYES, C.; BAZÁN, J.; CASTRO, M. A quantile parametric mixed regression model for bounded response variables. **Statistics and Its Interface**, v. 10, p. 483–493, 2017. Citations on pages 21 and 108.

BRAZIL, T. A. 2021. Available: <http://www.atlasbrasil.org.br/ranking>. Accessed: 13/01/2021. Citations on pages 85, 88, and 94.

____. 2021. Available: <http://www.atlasbrasil.org.br/consulta/planilha>. Accessed: 13/01/2021. Citations on pages 88 and 94.

____. 2021. Available: <http://www.atlasbrasil.org.br/acervo/atlas>. Accessed: 13/01/2021. Citation on page 94.

BRESLOW, N. E.; CLAYTON, D. G. Approximate inference in generalized linear mixed models. **Journal of the American statistical Association**, Taylor & Francis Group, v. 88, n. 421, p. 9–25, 1993. Citation on page 68.

CANCHO, V. G.; BAZÁN, J. L.; DEY, D. K. A new class of regression model for a bounded response with application in the study of the incidence rate of colorectal cancer. **Statistical methods in medical research**, SAGE Publications Sage UK: London, England, v. 29, p. 2015–2033, 2020. Citation on page 108.

CARPENTER, B.; GELMAN, A.; HOFFMAN, M.; LEE, D.; GOODRICH, B.; BETANCOURT, M.; BRUBAKER, M.; GUO, J.; LI, P.; RIDDELL, A. Stan: A probabilistic programming language. **Journal of Statistical Software**, v. 76, n. 1, p. 1–32, 2017. Citation on page 75.

DEMPSTER, A. P.; LAIRD, N. M.; RUBIN, D. B. Maximum likelihood from incomplete data via the em algorithm. **Journal of the Royal Statistical Society: Series B (Methodological)**, Wiley Online Library, v. 39, n. 1, p. 1–22, 1977. Citation on page 71.

DUNN, P. K.; SMYTH, G. K. Randomized quantile residuals. **Journal of Computational and Graphical Statistics**, v. 5, n. 3, p. 236–244, 1996. Citation on page 77.

ENEA, M.; STASINOPOULOS, M.; RIGBY, B.; HOSSAIN, A. **gamlss.inf: Fitting Mixed (Inflated and Adjusted) Distributions**. [S.l.], 2019. R package version 1.0-1. Available: <https://CRAN.R-project.org/package=gamlss.inf>. Citation on page 82.

FERRARI, S.; CRIBARI-NETO, F. Beta regression for modelling rates and proportions. **Journal of Applied Statistics**, v. 31, n. 7, p. 799–815, 2004. Citations on pages 21 and 79.

FIGUEROA-ZÚÑIGA, J. I.; ARELLANO-VALLE, R. B.; FERRARI, S. L. Mixed beta regression: A bayesian perspective. **Computational Statistics & Data Analysis**, v. 61, p. 137–147, 01 2013. Citation on page 22.

GELMAN, A.; CARLIN, J. B.; STERN, H. S.; DUNSON, D. B.; VEHTARI, A.; RUBIN, D. B. **Bayesian data analysis**. [S.l.]: CRC press, 2013. Citation on page 104.

GELMAN, A.; HWANG, J.; VEHTARI, A. Understanding predictive information criteria for bayesian models. **Statistics and computing**, v. 24, n. 6, p. 997–1016, 2014. Citations on pages 76 and 77.

HARVILLE, D. Extension of the gauss-markov theorem to include the estimation of random effects. **The Annals of Statistics**, JSTOR, v. 4, n. 2, p. 384–395, 1976. Citation on page 70.

HASTIE, T. J.; TIBSHIRANI, R. J. **Generalized additive models**. [S.l.]: CRC press, 1990. Citation on page 69.

HOFFMAN, M. D.; GELMAN, A. The no-u-turn sampler: adaptively setting path lengths in hamiltonian monte carlo. **Journal of Machine Learning Research**, v. 15, n. 1, p. 1593–1623, 2014. Citations on pages 72, 73, 74, and 107.

HOSSAIN, A.; RIGBY, R.; STASINOPOULOS, M.; ENEA, M. Centile estimation for a proportion response variable. **Statistics in Medicine**, v. 35, n. 6, p. 895–904, 2016. Citations on pages 82 and 108.

INSTITUTE, N. C. 2021. Available: <https://www.inca.gov.br/tipos-de-cancer/cancer-de-pulmao>. Accessed: 13/01/2021. Citation on page 94.

____. 2021. Available: <https://mortalidade.inca.gov.br/MortalidadeWeb/>. Accessed: 13/01/2021. Citation on page 94.

JOHNSON, N. L. Systems of frequency curves generated by methods of translation. **Biometrika**, v. 36, n. 1/2, p. 149–176, 1949. Citations on pages 21, 23, and 47.

JOHNSON, N. L.; KOTZ, S.; BALAKRISHNAN, N. **Continuous univariate distributions, volume 2**. [S.l.]: John wiley & sons, 1995. Citation on page 49.

JR, G. C. D.; KUTNER, M. H. The lagged normal family of probability density functions applied to indicator-dilution curves. **Biometrics**, JSTOR, p. 669–675, 1976. Citation on page 49.

LAIRD, N. M.; WARE, J. H. Random-effects models for longitudinal data. **Biometrics**, JSTOR, p. 963–974, 1982. Citations on pages 65, 70, and 71.

LEMONTE, A. J.; BAZÁN, J. L. New class of johnson distributions and its associated regression model for rates and proportions. **Biometrical Journal**, v. 58, n. 4, p. 727–746, 2016. Citations on pages 21, 23, and 108.

LOVISON, G.; SCHINDLER, C. Separate regression modelling of the gaussian and exponential components of an emg response from respiratory physiology. In: **Proceedings of the 29th International Workshop on Statistical Modelling**. [S.l.: s.n.], 2014. p. 189–194. Citation on page 29.

MIGLIORATI, S.; BRISCO, A. M. D.; ONGARO, A. A new regression model for bounded responses. **Bayesian Analysis**, v. 13, n. 3, p. 845–872, 2018. Citation on page 21.

NELDER, J. A.; MEAD, R. A simplex method for function minimization. **Comput. J.**, v. 7, p. 308–313, 1965. Citation on page 49.

NELSON, D. B. Conditional heteroskedasticity in asset returns: A new approach. **Econometrica: Journal of the Econometric Society**, p. 347–370, 1991. Citations on pages 37 and 49.

NESTEROV, Y. Primal-dual subgradient methods for convex problems. **Mathematical programming**, Springer, v. 120, n. 1, p. 221–259, 2009. Citation on page 74.

OLIVEIRA, L. P. d. **Modelos lineares mistos em estudos toxicológicos longitudinais**. Phd Thesis (PhD Thesis) — Universidade de São Paulo, 2015. Citation on page 21.

OSPINA, R.; FERRARI, S. L. A general class of zero-or-one inflated beta regression models. **Computational Statistics & Data Analysis**, Elsevier, v. 56, n. 6, p. 1609–1623, 2012. Citations on pages 82 and 107.

PAULA, G. A. **Modelos de regressão: com apoio computacional**. [S.l.]: IME-USP São Paulo, 2004. Citation on page 77.

PAZ, R. F. da; BALAKRISHNAN, N.; BAZÁN, J. L. L-logistic regression models: Prior sensitivity analysis, robustness to outliers and applications. **Brazilian Journal of Probability and Statistics**, Brazilian Statistical Association, v. 33, n. 3, p. 455–479, 2019. Citations on pages 75 and 79.

PAZ, R. F. da; EHLERS, R. S.; BAZÁN, J. L. A weibull mixture model for the votes of a brazilian political party. In: **Interdisciplinary Bayesian Statistics**. [S.l.]: Springer, 2015. p. 229–241. Citation on page 88.

PEREIRA, G. H. On quantile residuals in beta regression. **Communications in Statistics-Simulation and Computation**, v. 48, n. 1, p. 302–316, 2019. Citation on page 77.

PEREZ-GRACIA, J. L.; SANMAMED, M. F.; BOSCH, A.; PATIÑO-GARCIA, A.; SCHALPER, K. A.; SEGURA, V.; BELLMUNT, J.; TABERNERO, J.; SWEENEY, C. J.; CHOUEIRI, T. K. *et al.* Strategies to design clinical studies to identify predictive biomarkers in cancer research. **Cancer treatment reviews**, Elsevier, v. 53, p. 79–97, 2017. Citation on page 108.

PINHEIRO, J.; BATES, D.; DEBROY, S.; SARKAR, D.; R Core Team. **nlme: Linear and Nonlinear Mixed Effects Models**. [S.l.], 2020. R package version 3.1-149. Available: <https://CRAN.R-project.org/package=nlme>. Citation on page 74.

R Core Team. **R: A Language and Environment for Statistical Computing**. Vienna, Austria, 2020. Available: <https://www.R-project.org/>. Citation on page 49.

RIGBY, R.; STASINOPOULOS, D.; HELLER, G.; BASTIANI, F. D. **Distributions for Modeling Location, Scale, and Shape: Using GAMLSS in R**. [S.l.]: CRC press, 2019. Citations on pages 26 and 74.

RIGBY, R. A.; STASINOPOULOS, D. M. Generalized additive models for location, scale and shape. **Journal of the Royal Statistical Society: Series C (Applied Statistics)**, v. 54, n. 3, p. 507–554, 2005. Citations on pages 68, 74, and 107.

____. Automatic smoothing parameter selection in gamlss with an application to centile estimation. **Statistical methods in medical research**, Sage Publications Sage UK: London, England, v. 23, n. 4, p. 318–332, 2014. Citation on page 70.

RIZOPOULOS, D. **Joint models for longitudinal and time-to-event data: With applications in R**. [S.l.]: CRC press, 2012. Citation on page 64.

RODRIGUES, J.; BAZáN, J.; SUZUKI, A. A flexible procedure for formulating probability distributions on the unit interval with applications. **Communications in Statistics - Theory and Methods**, p. 1–17, 2019. Citations on pages 22, 23, 26, and 107.

SCHWARZ, G. *et al.* Estimating the dimension of a model. **The annals of statistics**, Institute of Mathematical Statistics, v. 6, n. 2, p. 461–464, 1978. Citation on page 76.

SEARLE, S. R.; MCCULLOCH, C. E. **Generalized, linear and mixed models**. [S.l.]: Wiley, 2001. Citation on page 22.

SINGER, J. M.; ANDRADE, D. d. Análise de dados longitudinais. **SIMPÓSIO NACIONAL DE PROBABILIDADE E ESTATÍSTICA**, Embrapa São paulo, v. 7, 1986. Citation on page 65.

SORENSEN, T.; VASISHTH, S. Bayesian linear mixed models using stan: A tutorial for psychologists, linguists, and cognitive scientists. **Tutorials in Quantitative Methods for Psychology**, v. 12, 06 2015. Citation on page 75.

Stan Development Team. **RStan: the R interface to Stan**. 2020. R package version 2.19.3. Available: <http://mc-stan.org/>. Citation on page 74.

STASINOPOULOS, D.; RIGBY, R. Generalized additive models for location scale and shape (gamlss) in r. **Journal of Statistical Software**, v. 23, n. 7, p. 1–46, 12 2007. Citation on page 74.

STASINOPOULOS, M. D.; RIGBY, R. A.; HELLER, G. Z.; VOUDOURIS, V.; BASTIANI, F. D. **Flexible regression and smoothing: using GAMLSS in R**. [S.l.]: Chapman and Hall/CRC, 2017. Citations on pages 68 and 102.

STONE, M. An asymptotic equivalence of choice of model by cross-validation and akaike's criterion. **Journal of the Royal Statistical Society: Series B (Methodological)**, v. 39, n. 1, p. 44–47, 1977. Citation on page 75.

VEHTARI, A.; GELMAN, A.; SIMPSON, D.; CARPENTER, B.; BÜRKNER, P.-C. *et al.* Rank-normalization, folding, and localization: An improved $\widehat{R}$ for assessing convergence of mcmc. **Bayesian Analysis**, International Society for Bayesian Analysis, 2020. Citation on page 75.

VERKUILEN, J.; SMITHSON, M. Mixed and mixture regression models for continuous bounded responses using the beta distribution. **Journal of Educational and Behavioral Statistics**, v. 37, p. 82–113, 02 2012. Citation on page 22.

WATANABE, S.; OPPER, M. Asymptotic equivalence of bayes cross validation and widely applicable information criterion in singular learning theory. **Journal of machine learning research**, v. 11, n. 12, p. 3571–3594, 2010. Citations on pages 75 and 76.

WU, L. **Mixed effects models for complex data**. [S.l.]: CRC Press, 2009. 1-414 p. Citation on page 64.

APPENDIX

A

TABLES OF RESULTS OF THE SIMULATION STUDY IN THE CHAPTER 5.

Table 26 – Percentage of cases that each distribution displayed a lower value for AIC and BIC for different scenarios with contaminated samples from a Beta distribution with $\mu = 0.5$. 1000 replications of each scenario.

			AIC						BIC					
ϕ	% of outiler	n	Beta	LGU	LRG	LEG	LSN1	LPE	Beta	LGU	LRG	LEG	LSN1	LPE
10	2	75	5.1	0	24.6	45.2	0	25.1	16.4	0	47.7	15.6	0	20.3
		150	4.5	0	2.9	75.3	0	20.8	10.4	0	13.9	57.9	0	17.8
		200	0.9	0	1.4	81.4	0	16.3	4.5	0	7	73.3	0	15.2
	5	75	0.1	0	28.3	49.6	0	21.7	1.5	0	64.6	14	0	19.9
		150	0	0	4.5	80.4	0	15.1	0.1	0	30.1	54.7	0	15.1
		200	0	0	2.2	85.9	0	11.9	0	0	14.4	73.7	0	11.9
	10	75	0	0	11.1	78.9	0	10	0	0	52.3	38	0	9.7
		150	0	0	1	93.6	0	5.4	0	0	15.6	79	0	5.4
		200	0	0	0.1	97.6	0	2.3	0	0	5.6	92.1	0	2.3
20	2	75	0	0	24.5	50.8	0	24.7	0	0	36.6	0.8	0	62.6
		150	0	0	3.3	80.3	0	16.4	0	0	19.9	18.4	0	61.7
		200	0	0	1.6	84.7	0	13.7	0	0	9.4	38.7	0	51.9
	5	75	0	0	6	69.4	0	24.6	0	0	38.9	37.3	0	23.8
		150	0	0	0.2	84.2	0	15.6	0	0	8.9	75.5	0	15.6
		200	0	0	0	84	0	16	0	0	0.8	57.2	0	42
	10	75	0	0	0	83.1	0	16.9	0	0	2.0	81.1	0	16.9
		150	0	0	0	89.8	0	10.2	0	0	0	89.8	0	10.2
		200	0	0	0	92.9	0	7.1	0	0	0	92.9	0	7.1

Table 27 – Percentage of cases that each distribution displayed a lower value for AIC and BIC for different scenarios with contaminated samples from a Beta distribution with $\mu = 0.8$. 1000 replications of each scenario.

			AIC						BIC					
ϕ	% of outiler	n	Beta	LGU	LRG	LEG	LSN1	LPE	Beta	LGU	LRG	LEG	LSN1	LPE
10	2	75	57.5	0.2	26.5	2.1	0	13.7	67.4	0.3	27.8	0.2	0	4.3
		150	69.1	0	15.8	8.4	0	6.7	79.7	0	17.5	0.2	0	2.6
		200	71.5	0	13.3	12.0	0	3.3	83.2	0	15.5	0.3	0	1.0
	5	75	53.1	0.2	32.4	3.4	0	10.9	61.2	0.2	34.0	0.1	0	4.5
		150	64.1	0	22.3	10.1	0	3.5	73.7	0	24.9	0.5	0	0.9
		200	67.1	0	15.1	15.6	0	2.2	79.3	0	18.6	1.0	0	1.1
	10	75	51.3	0	35.5	3.1	0	10.1	57.4	0	38.0	0.1	0	4.5
		150	61.0	0	26.1	10.2	0	2.7	68.9	0	29.7	0.3	0	1.1
		200	62.2	0	22.9	13.3	0	1.6	72.1	0	26.9	0.5	0	0.5
20	2	75	44.7	0.1	31.7	13.3	0	10.2	10.0	0	33.5	0.2	0	56.3
		150	46.0	0	15.1	34.0	0	4.9	27.9	0	27.3	3.0	0	41.8
		200	47.2	0	9.2	41.5	0	2.1	41.5	0	18.5	7.8	0	32.2
	5	75	32.7	0	42.4	18.5	0	6.4	39.6	0	55.7	1.8	0	2.9
		150	24.4	0	29.5	44.8	0	1.3	37.9	0	49.5	12.0	0	0.6
		200	15.3	0	21.6	62.6	0	0.5	24.7	0	45.6	17.3	0	12.4
	10	75	13.8	0	57.2	26.0	0	3.0	18.0	0	75.3	5.8	0	0.9
		150	8.3	0	41.5	50.0	0	0.2	14.4	0	71.7	13.7	0	0.2
		200	5.9	0	33.4	60.6	0	0.1	11.3	0	68.6	20.0	0	0.1

Table 28 – Average of AIC and BIC for different scenarios with contaminated samples from a Beta distribution with $\mu = 0.5$. 1000 replications of each scenario.

	% of		AIC						BIC					
ϕ	outiler	n	Beta	LGU	LRG	LEG	LSN1	LPE	Beta	LGU	LRG	LEG	LSN1	LPE
10	2	75	-53.20	-9.40	-50.39	-53.26	-44.65	-51.86	**-57.83**	-14.03	-55.03	**-60.22**	-51.61	-58.81
		150	-106.44	2.27	-105.79	**-116.64**	-92.30	-112.44	-100.42	8.29	-99.77	**-107.61**	-83.27	-103.41
		200	-142.81	72.60	-141.51	**-156.86**	-123.99	-151.08	-136.22	79.19	-134.91	**-146.97**	-114.09	-141.18
	5	75	-35.43	22.37	**-43.07**	-42.97	-22.28	-39.59	-40.07	17.73	-47.71	**-49.93**	-29.24	-46.55
		150	-75.53	49.49	-92.52	**-98.70**	-53.85	-90.82	-69.51	55.51	-86.50	**-89.67**	-44.82	-81.79
		200	-98.03	81.34	-122.17	**-130.86**	-68.59	-119.63	-91.43	87.94	-115.58	**-120.96**	-58.69	-109.74
	10	75	-17.40	40.88	**-32.44**	-32.16	-1.26	-25.15	-22.03	36.24	-37.07	**-39.12**	-8.21	-32.10
		150	-43.95	76.69	-73.30	**-79.20**	-17.63	-63.91	-37.92	82.71	-67.28	**-70.17**	-8.60	-54.88
		200	-60.02	103.48	-98.70	**-107.04**	-25.41	-86.63	-53.42	110.08	-92.11	**-97.14**	-15.52	-76.74
20	2	75	-85.97	-14.39	-93.90	**-95.12**	-73.24	-92.53	-90.61	-19.03	-98.54	**-102.07**	-80.19	-99.48
		150	-163.95	-1.03	-189.52	**-197.23**	-140.07	-190.98	-157.93	4.99	-183.50	**-188.20**	-131.03	-181.95
		200	-220.16	158.97	-253.44	**-264.60**	-188.57	-256.05	-213.56	165.57	-246.84	**-254.71**	-178.67	-246.15
	5	75	-55.22	23.86	-79.17	**-79.26**	-36.72	-75.85	-59.86	19.23	-83.80	**-86.21**	-43.67	-82.80
		150	-111.77	54.73	-162.39	**-169.51**	-78.87	-161.56	-105.75	60.75	-156.37	**-160.48**	-69.84	-152.53
		200	-144.42	106.15	-213.44	**-224.00**	-100.68	-212.69	-137.82	112.75	-206.84	**-214.10**	-90.78	-202.80
	10	75	-27.10	45.40	-57.27	**-61.81**	-5.89	-55.86	-31.74	40.77	-61.90	**-68.76**	-12.85	-62.82
		150	-62.56	84.35	-121.62	**-137.38**	-26.52	-123.51	-56.54	90.38	-115.60	**-128.35**	-17.48	-114.47
		200	-84.84	124.85	-163.36	**-184.59**	-36.79	-165.95	-78.24	131.45	-156.76	**-174.69**	-26.89	-156.05

Table 29 – Average of AIC and BIC for different scenarios with contaminated samples from a Beta distribution with $\mu = 0.8$. 1000 replications of each scenario.

	% of		AIC						BIC					
ϕ	outiler	n	Beta	LGU	LRG	LEG	LSN1	LPE	Beta	LGU	LRG	LEG	LSN1	LPE
10	2	75	-112.55	-84.31	-109.93	-109.16	-104.40	-106.65	-117.19	-88.94	-114.56	-116.11	-111.35	-113.60
		150	-236.06	-172.43	-230.75	-234.78	-225.40	-228.68	-230.03	-166.40	-224.73	-225.75	-216.37	-219.65
		200	-315.46	-225.39	-308.34	-314.24	-301.08	-305.45	-308.86	-218.79	-301.74	-304.35	-291.18	-295.55
	5	75	-112.23	-80.13	-110.30	-108.98	-102.78	-105.65	-116.86	-84.76	-114.94	-115.93	-109.73	-112.60
		150	-235.59	-164.36	-231.68	-234.52	-222.54	-226.77	-229.57	-158.34	-225.66	-225.49	-213.51	-217.74
		200	-315.30	-205.73	-310.08	-314.33	-297.55	-303.27	-308.71	-199.14	-303.48	-304.43	-287.66	-293.37
	10	75	-112.94	-77.18	-111.85	-109.78	-102.04	-105.40	-117.58	-81.81	-116.49	-116.73	-108.99	-112.35
		150	-237.67	-159.69	-235.30	-236.67	-222.06	-226.89	-231.65	-153.67	-229.28	-227.64	-213.03	-217.86
		200	-318.34	-205.83	-315.21	-317.53	-296.64	-303.54	-311.74	-199.23	-308.62	-307.63	-286.75	-293.64
20	2	75	-149.63	-119.52	-146.67	-146.77	-141.40	-144.65	-154.27	-124.15	-151.31	-153.72	-148.35	-151.60
		150	-306.73	-225.75	-302.18	-307.52	-293.34	-300.90	-300.71	-219.73	-296.15	-298.49	-284.31	-291.87
		200	-409.47	-240.53	-403.43	-410.98	-391.87	-401.70	-402.88	-233.94	-396.83	-401.09	-381.97	-391.81
	5	75	-143.36	-99.26	-143.50	-142.12	-130.79	-137.71	-148.00	-103.89	-148.14	-149.08	-137.74	-144.66
		150	-295.38	-190.96	-296.45	-299.11	-274.49	-288.16	-289.36	-184.94	-290.43	-290.08	-265.46	-279.13
		200	-392.99	-185.19	-394.96	-398.78	-363.99	-383.36	-386.39	-178.59	-388.36	-388.88	-354.09	-373.46
	10	75	-135.44	-80.26	-138.75	-136.39	-118.15	-128.74	-140.08	-84.89	-143.38	-143.34	-125.10	-135.70
		150	-281.34	-159.67	-288.03	-288.94	-252.00	-271.75	-275.32	-153.65	-282.00	-279.90	-242.96	-262.72
		200	-375.70	-159.36	-384.63	-386.42	-335.62	-363.29	-369.10	-152.76	-378.04	-376.52	-325.73	-353.39

APPENDIX

B

R CODE FOR FIT THE MIXED REGRESSION MODEL CONSIDERING JSB DISTRIBUTIONS

GAMLSS CODE

```
library(gamlss)

y =  read.csv("MortalityData.csv",header = T,sep = ",",dec = ".")
head(y)
y = y[,-1]
y = na.exclude(y, na.action = "exclude", fill = NULL)
y$y = y$y/500
hist(y$y)




#Fiting the model
gen.Family(family = "PE",type = "logit")

m1PE<-gamlss(y~Sex*Time + Age*Time + Time +
Longevity*Time +re(random=~Time|id), sigma.formula = ~
Time,data=y,
family=logitPE, trace=TRUE,method =RS(80))

summary(m1PE)
getSmo(m1PE)
```

```
AIC(m1PE)
BIC(m1PE)

#Getting the quantile residuals
qqnorm(residuals(m1PE),xlim=c(-4,4), ylim = c(-4,5),cex.axis = 2.7,main = "",
cex.lab = 2.7,cex = 2.5)
qqline(residuals(m1PE), lwd = 5)

#Envelope
rq = residuals(m1PE)
B <- 19
n = length(y$y)
e <- matrix(0,n,B)
tr = fitted.values(m1PE)
sr = m1PE$sigma.fv
vr = m1PE$nu.fv
head(y)
for (i in 1:B){
  set.seed(123*i)
  Sim_Y <- rlogitPE(n,tr,sr,vr)
  X = data.frame(cbind(Sim_Y,y))
  head(X)
  attach(X)
  m1PE<-gamlss(Sim_Y~Sex*Time + Age*Time + Time +
  Longevity*Time +re(random=~Time|id),
  sigma.formula = ~ Time,data=X,family=logitPE, trace=TRUE, method = RS(80))
  e[,i] <- sort(residuals(m1PE))

}

          e
e1 <- numeric(n)
e2 <- numeric(n)

for (i in 1:n)
{
  eo <- sort(e[i,])
  e1[i] <- quantile(eo,0.025)
  e2[i] <- quantile(eo,0.975)
```

```
}

med <- apply(e,1,median)
td <- rq
faixa <- range(td,e1,e2)

e1g3 <- e1
e2g3 <- e2
tdg3 <- td
par(mar = c(5, 5, 1, 1)) # Set the margin on all sides to 6
qqnorm(td, xlab="Theoretical quantiles", ylab="Quantile residuals",
ylim=faixa, xlim=c(-3,3), pch=16,
cex.axis = 2.7,main = "",cex.lab = 2.7,
cex = 2.5)
par(new=TRUE)
#
qqnorm(e1,axes=F,xlab="",ylab="",type="l",ylim=faixa
, xlim=c(-3,3),lty=1, lwd = 4,main="")
par(new=TRUE)
qqnorm(e2,axes=F,xlab="",ylab="", type="l",ylim=faixa, xlim=c(-3,3),
lty=1, lwd = 4, main="")
par(new=TRUE)
qqnorm(med,axes=F,xlab="",ylab="", type="l",ylim=faixa, xlim=c(-3,3),
lty=1, lwd = 4,main="")
```

RSTAN CODE

```
model <- "
functions{
real logisticpowerexponential_lpdf(real y, real mu, real tau, real nu){
return  (log(nu) - pow((sqrt(pow((logit(y) - mu)*tau/sqrt(tgamma(1/nu)/
(tgamma(3/nu))),2))),nu)  -log(tgamma(1/nu)/(tgamma(3/nu)))/2 + log(tau)
- lgamma(1/nu) -  log(y*(1-y)));
      }
}
data {
int<lower=0> n;     //the number of observations
int<lower=1> m;     //number of subjects
```

```
int p;              //the number of columns in the model matrix X
int q;
int J;              //the number of columns in the block matrix Z
int<lower=1, upper=m> subj[n]; //subject id
real<lower = 0,upper = 1> y[n];                  //the response
matrix[n,p] X;                //the model matrix
matrix[n, q] W;
matrix[n,J] Z;                //the  matrix of ramdom effect
}

parameters{
vector[p] beta;         //the regression parameters
vector[q] gamma;
vector[1] delta;
vector<lower=0>[J] sigma_b; //subj sd
cholesky_factor_corr[J] L_b;
matrix[J,m] Z_b;
}
transformed parameters {
matrix[J,m] b;
vector[n] mu;
vector[n] sigma;
vector[1] nu;
sigma = exp(W*gamma);
nu = exp(delta);
b=diag_pre_multiply(sigma_b, L_b) * Z_b; //subj random effect
for(i in 1:n){
mu[i] <- (X[i,]*beta + Z[i,]*b[,subj[i]]);
}
}
model{
L_b ~ lkj_corr_cholesky(2.0);
to_vector(Z_b) ~ normal(0,1);
to_vector(beta) ~ normal(0,100);
to_vector(gamma) ~ normal(0,100);
to_vector(delta) ~ normal(0,100);
for(i in 1:n){
target += logisticpowerexponential_lpdf(y[i]| mu[i], sigma[i], nu[1]);
}
```

```
}
generated quantities {
vector[n] log_lik;
for(i in 1:n){
log_lik[i] = logisticpowerexponential_lpdf(y[i]| mu[i], sigma[i], nu[1]);
}
}
"
```