

Alternative regression models to Beta distribution  
under Bayesian approach

*Rosineide Fernando da Paz*

UNIVERSIDADE FEDERAL DE SO CARLOS  
CENTRO DE CINCIAS EXATAS E TECNOLOGIA  
PROGRAMA INTERINSTITUCIONAL DE PS-GRADUAO EM ESTATSTICA UFSCar-USP

ROSINEIDE FERNANDO DA PAZ

**ALTERNATIVE REGRESSION MODELS TO BETA DISTRIBUTION UNDER  
BAYESINA APPROACH**

Thesis submitted to the Departamento de Estatística - Des/UFSCar and to the Instituto de Ciências Matemáticas e de Computação - ICMC-USP, in partial fulfillment for PhD degree in Statistics - joint Graduate Program in Statistics UFSCar-USP.

Advisor: Prof. Dr. Jorge Luís Bazán Guzmán

São Carlos  
August 2017



UNIVERSIDADE FEDERAL DE SO CARLOS  
CENTRO DE CINCIAS EXATAS E TECNOLOGIA  
PROGRAMA INTERINSTITUCIONAL DE PS-GRADUAO EM ESTATSTICA UFSCar-USP

ROSINEIDE FERNANDO DA PAZ

**MODELOS DE REGRESSÃO ALTERNATIVOS À DISTRIBUIÇÃO BETA SOB  
ABORDAGEM BAYESIANA**

Tese 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 Doutor em Estatística - Programa Interinstitucional de Ps-Graduação em Estatística UFSCar-USP.

Orientador: Prof. Dr. Jorge Luís Bazán Guzmán

São Carlos  
Agosto de 2017



## UNIVERSIDADE FEDERAL DE SÃO CARLOS

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

---

### Folha de Aprovação

---

Assinaturas dos membros da comissão examinadora que avaliou e aprovou a Defesa de Tese de Doutorado da candidata Rosineide Fernando da Paz, realizada em 25/08/2017:

---

Prof. Dr. Jorge Luis Bazán Guzmán  
USP

---

Prof. Dr. Heleno Bolfarine  
USP

---

Prof. Dr. Artur José Lemonte  
UFRN

---

Prof. Dr. Caio Lucidius Naberezny Azevedo  
UNICAMP

---

Prof. Dr. Luis Aparecido Milan  
UFSCar



*Este trabalho é dedicado às crianças adultas que,  
quando pequenas, sonharam em se tornar cientistas.*





# ACKNOWLEDGEMENTS

---

---

Os agradecimentos principais são direcionados à Deus, quando algumas vezes, sentindo-me desacreditada e perdida nos meus objetivos, ideais ou minha pessoa, me deu forças e me fez acreditar em mim mesma.

Um agradecimento especial vai para o meu marido Amilton José Monteiro, que sempre esteve lá, por mim.

Agradeço aos professores participantes da banca examinadora que dividiram comigo este momento tão importante e esperado e por terem sido mediadores do meu conhecimento e terem despertado em mim a busca contínua de desenvolvimento e por informações: Artur J. Lemonte, Caio L. N. Azevedo, Heleno Bolfarine, Luís A. Milan e Jorge Luis Bazán. E em especial ao orientador dessa tese (Jorge Luis Bazán) que acompanhou todo o desenvolvimento desse trabalho.

Agradeço também a Coordenação de Aperfeiçoamento de Pessoal de Nível Superior (CAPES), pelo suporte financeiro.

Enfim, agradeço aos amigos, familiares, professores e todos aqueles que cruzaram em minha vida, participando de alguma forma na construção e realização deste tão desejado sonho de obter o título de doutora (um dos ingredientes para minha felicidade).



*“A verdadeira viagem de descobrimento não consiste em procurar novas paisagens,  
mas em ter novos olhos. ”*  
*(Marcel Proust)*



# ABSTRACT

PAZ, R. F. **Alternative regression models to Beta distribution under Bayesian approach.** 2017. 120 p. Tese (Doutorado em Estatística – Interinstitucional de Pós-Graduação em Estatística) – Instituto de Ciências Matemáticas e de Computação, Universidade de São Paulo, São Carlos – SP, 2017.

The Beta distribution is a bounded domain distribution which has dominated the modeling the distribution of random variable that assume value between 0 and 1. Bounded domain distributions arising in various situations such as rates, proportions and index. Motivated by an analysis of electoral votes percentages (where a distribution with support on the positive real numbers was used, although a distribution with limited support could be more suitable) we focus on alternative distributions to Beta distribution with emphasis in regression models. In this work, initially we present the Simplex mixture model as a flexible model to modeling the distribution of bounded random variable then we extend the model to the context of regression models with the inclusion of covariates. The parameters estimation is discussed for both models considering Bayesian inference. We apply these models to simulated data sets in order to investigate the performance of the estimators. The results obtained were satisfactory for all the cases investigated. Finally, we introduce a parameterization of the L-Logistic distribution to be used in the context of regression models and we extend it to a mixture of mixed models.

**Keywords:** L-Logistic distribution, Bounded response, Mixture model, Simplex distribution, Bayesian inference, Beta distribution, Human development index, Regression model.



# RESUMO

PAZ, R. F. **Modelos de regressão alternativos à distribuição Beta sob abordagem bayesiana**. 2017. 120 p. Tese (Doutorado em Estatística – Interinstitucional de Pós-Graduação em Estatística) – Instituto de Ciências Matemáticas e de Computação, Universidade de São Paulo, São Carlos – SP, 2017.

A distribuição beta é uma distribuição com suporte limitado que tem dominado a modelagem de variáveis aleatórias que assumem valores entre 0 e 1. Distribuições com suporte limitado surgem em várias situações como em taxas, proporções e índices. Motivados por uma análise de porcentagens de votos eleitorais, em que foi assumida uma distribuição com suporte nos números reais positivos quando uma distribuição com suporte limitado seira mais apropriada, focamos em modelos alternativos a distribuição beta com ênfase em modelos de regressão. Neste trabalho, apresentamos, inicialmente, um modelo de mistura de distribuições Simplex como um modelo flexível para modelar a distribuição de variáveis aleatórias que assumem valores em um intervalo limitado, em seguida estendemos o modelo para o contexto de modelos de regressão com a inclusão de covariáveis. A estimação dos parâmetros foi discutida para ambos os modelos, considerando o método bayesiano. Aplicamos os dois modelos a dados simulados para investigarmos a performance dos estimadores usados. Os resultados obtidos foram satisfatórios para todos os casos investigados. Finalmente, introduzimos a distribuição L-Logística no contexto de modelos de regressão e posteriormente estendemos este modelo para o contexto de misturas de modelos de regressão mista.

**Palavras-chave:** Distribuição L-Logística, Resposta limitada, Modelo de mistura, Distribuição Simplex, Inferência bayesiana, Distribuição Beta, Índice de desenvolvimento humano, Modelo de regressão.





# LIST OF FIGURES

---

---

Figure 1	– Histograms of the data of voting percentage obtained by PT in presidential elections, in the cities of Sergipe State, from year 1994 and 1998, when the PT lost the presidential election, to 2002, 2006 and 2010, when the PT candidate was Presidential winner, and its estimated densities based on the posterior predictive distribution for 1, 2 and 3 components. . . . .	38
Figure 2	– Histograms and posterior density function. . . . .	49
Figure 3	– Real histogram and Estimated density function for the MHDI data set. . . . .	52
Figure 4	– Classification of HDI of cities in the states São Paulo and Northeastern region of Brazil where the cities classified in the second component are in black in (A) and cities classified in the first component are in black in (B). . . . .	53
Figure 5	– Scatter plot with marginal histograms of the data. . . . .	60
Figure 6	– Scatter plot of the classified data. . . . .	61
Figure 7	– L-Logistic probability density function for scale parameter $m = 0.1, 0.5$ and $0.7$ and some values of parameter $b$ . . . . .	68
Figure 8	– L-Logistic probability density function for shape parameter $b = 0.1, 1$ and $4$ and some values of scale parameter $m$ . . . . .	68
Figure 9	– The mode, skewness ( $\gamma_M$ and $\gamma_{0.125}$ ) and kurtosis ( $k_Q$ ) of the L-Logistic distribution for some values of the parameters. . . . .	72
Figure 10	– Descriptive measures of the L-Logistic distributions for some values of the parameters . . . . .	73
Figure 11	– Estimated densities for Beta and L-Logistic models for de scenarios with $n = 100$ , $\phi = 10$ and $r = 5\%$ . . . . .	81
Figure 12	– Posterior predictive error bars with 95% confidence intervals of the generated values $y_{(i)}^{rep}$ versus ordered observed data $y_{(i)}$ for the PPOBC data, using L-Logistic and Beta models. . . . .	83
Figure 13	– Estimated density of PPOBC data. . . . .	83
Figure 14	– Scatterplot and histograms of the real data. . . . .	85
Figure 15	– Standard residual versus adjusted values for the L-Logistic and Beta models. . . . .	86
Figure 16	– L-Logistic probability density function for scale parameter $m = 0.2, 0.5$ and $0.8$ and some values of parameter $b$ . . . . .	91
Figure 17	– L-Logistic probability density function for shape parameter $b = 0.5, 1$ and $2$ and some values of scale parameter $m$ . . . . .	91

Figure 18 – Chais values for the parameters of MLLMR model considering the simulated data where the values of the parameters of component 1 are in green, values of the parameters of component 2 are in black and the values of parameters of component 3 are in red. . . . . 99

Figure 19 – Chais values for the parameters of MLLMR model considering the data of percentage of votes, where the values of the parameters of component 1 are in black and values of the parameters of component 2 are in red. . . . . 101

---

# LIST OF ALGORITHMS

---

---

Algoritmo 1 – Algorithm for simulating samples from the posterior distribution of the parameters of the mixture of Weibull . . . . .	34
Algoritmo 2 – Algorithm for simulate samples from the jointly posterior distribution of the parameters of the mixture of L-logistic regression models . . . . .	58
Algoritmo 3 – Algorithm for simulate samples from the posterior distribution of the parameters of the mixture of mixed L-Logistic regression models . . . . .	98
Algoritmo 4 – Algorithm for simulate samples from the posterior joint probability distribution of the parameters of mixture of simplex . . . . .	112



# LIST OF TABLES

---



---

Table 1	– Twice the natural logarithm of the Bayes factor of the data of voting percentage under one model resulting of mixture of Weibull distribution relative to another. . . . .	37
Table 2	– Posterior mean and HPD interval of parameters of the best Weibull mixture model chosen by Bayes factor evaluation. Data of voting percentage obtained by PT in presidential elections in the Sergipe State from year 1994 to 2010 was considered for fitting of the models. . . . .	40
Table 3	– Parameters used to simulate the data sets and the posterior relative frequency for the number of components obtained from each simulated data set of size $n$ . . . . .	49
Table 4	– Posterior mean of the parameters and empirical standard deviation (SD) for simulated data sets considering six models with $k = 2$ and $k = 3$ described in Table 3. . . . .	50
Table 5	– Relative frequency of $k$ to the MHDI data set considering alternative SM models. . . . .	51
Table 6	– Posterior estimates of the parameters and the empirical standard deviation for the MHDI data set. . . . .	51
Table 7	– Model comparison criteria to the models proposed to MHDI data. . . . .	61
Table 8	– Number of observations classified across the models and components. . . . .	61
Table 9	– Posterior mean, credibility intervals and standard empirical deviation of the estimated parameters for sub-model $M_0$ and $M_1$ . . . . .	62
Table 10	– $E_Y[Y]$ , $E_Y[Y^2]$ , and $Var_Y(X)$ of the L-Logistic distribution for some values of $b$ and $m$ . . . . .	73
Table 11	– Posterior mean with 95% HPD interval, prior distributions for parameter $b$ and true values of the parameters of L-Logistic distribution used to simulate the data sets. . . . .	78
Table 12	– Bias and root mean square error ( $\sqrt{MSE}$ ) of the Bayesian estimator of the parameters $m$ and $b$ . . . . .	80
Table 13	– Comparison of Bias, MSE and percentage of selection of the model L-Logistic versus Beta considering WAIC, EAIC, EBIC and DIC for different scenarios of contaminated Beta data (two values of $\phi$ 3% of outliers and three sample sizes) by considering 100 dataset replications in each scenario. . . . .	81
Table 14	– Estimates and 95% HPD intervals for the parameters of the L-Logistic and Beta models, and statistics for model comparison. . . . .	82
Table 15	– Model comparison criteria for model comparison. . . . .	86
Table 16	– Parameter estimates and 95% HPD intervals for the L-Logistic and Beta models. . . . .	87

Table 17 – Posterior mean and credibility intervals of the estimated parameters for the MHDI data, and model comparison between the LLR and LLMR models. . . . .	95
Table 18 – Posterior mean and 95% HPD intervals for the parameters of MLLMR model applied to simulated data. . . . .	99
Table 19 – Posterior mean and 95% HPD intervals for the parameters of MLLMR model applied to data of votes percentage. . . . .	100
Table T – Landscape multiple page table . . . . .	118

# CONTENTS

---

---

1	<b>INTRODUÇÃO</b>	23
2	<b>A MOTIVATION: STUDY OF THE VOTES OF A BRAZILIAN POLITICAL PARTY</b>	27
2.1	Introduction	28
2.2	The votes of a political party	29
2.3	The general mixture model	30
2.4	The Weibull mixture model	32
2.5	Choosing the number of components in the mixture model	34
2.6	Results	37
2.7	Discussion and further development	39
3	<b>MIXTURE OF SIMPLEX DISTRIBUTIONS WITH UNKNOWN NUMBER OF COMPONENTS</b>	41
3.1	Introduction	42
3.2	Simplex Mixture Distribution	43
3.3	Inference	44
3.4	Analysis of simulated data sets	48
3.5	Analysis of a municipal HDI data set in Brazil	51
3.6	Final comments	52
4	<b>MODELING MHDI WITH A FINITE MIXTURE OF SIMPLEX REGRESSION MODELS</b>	55
4.1	Introduction	56
4.2	Model specification	56
4.3	Bayesian inference	57
4.4	Data analysis	58
4.5	Conclusion	62
5	<b>L-LOGISTIC REGRESSION MODELS: PRIOR SENSITIVITY ANALYSIS, ROBUSTNESS TO OUTLIERS AND APPLICATIONS</b>	65
5.1	Introduction	66
5.2	The L-Logistic Distribution	67
5.3	Properties of the L-Logistic distribution	69

5.4	Bayesian inference . . . . .	73
5.5	Simulation studies . . . . .	76
5.6	Applications to a real data set . . . . .	81
5.7	Final remarks . . . . .	87
6	<b>FINITE MIXTURE OF MIXED L-LOGISTIC REGRESSION: A BAY- ESIAN APPROACH . . . . .</b>	<b>89</b>
6.1	Introduction . . . . .	89
6.2	L-Logistic distribution . . . . .	90
6.3	L-Logistic median regression model . . . . .	91
6.4	L-Logistic mixed median regression (LLMR) model . . . . .	92
6.5	Mixture of L-Logistic mixed-effect models . . . . .	96
6.6	Remarks . . . . .	100
7	<b>CONTRIBUTIONS AND FUTURE DEVELOPMENTS . . . . .</b>	<b>103</b>
7.1	Contributions . . . . .	103
7.2	Future development . . . . .	104
	<b>BIBLIOGRAPHY . . . . .</b>	<b>105</b>
APPENDIX A	<b>PROCEDURE FOR SIMULATE SAMPLE FROM A MIXTURE OF SIMPLEX DISTRIBUTIONS WITH UNKNOWN NUMBER OF COMPONENT . . . . .</b>	<b>111</b>
APPENDIX B	<b>PROOFS OF PROPERTIES OF THE L-LOGISTIC AND RESULTS FOR PRIOR SENSITIVITY ANALY- SIS . . . . .</b>	<b>115</b>



---

# INTRODUÇÃO

---

Random variables with support on a bounded subset of the real line are common in practical problems and are frequently analyzed by researchers, for instance, Impartial Anonymous Culture (STENSHOLT, 1999) and the Human Development Index (HDI) (MCDONALD; RANSOM, 2008; CIFUENTES *et al.*, 2008). Examples of bounded variables which are often analyzed are rates and proportions bounded in  $(0,1)$  interval. In these case, we usually transform the variables by the logit transformation in order to deal with an unbounded variables However, some problems arise with this approach. One problem is that rates and proportions display more variation around the mean and this variation decrease in the neighborhood of the lower and upper limits of the standard unit interval. In addition an antisymmetric distributions can be more appropriate for modeling this kind of data. For modeling this kind of data, different models have been proposed in the past. For example, among others, Buckley (2003), Ferrari and Cribari-Neto (2004), Lemonte and Bazán (2016), Gómez-Déniz, Sordo and Calderín-Ojeda (2014), Bayes, Bazán and Castro (2017) and Jones (2009). However, there are still continuous distributions with bounded support that need further study.

In the mixture model context, there are some studies which consider mixtures of Beta distributions (BOUGUILA; ZIOU; MONGA, 2006; BOUGUILA; ELGUEBALY, 2012), but other probability distributions with support in the  $(0,1)$  interval, the Simplex distribution for an example, have not yet been completely analyzed or studied. The Simplex distribution was proposed by Barndorff-Nielsen and Jorgensen (1991) and has recently been considered as a complementary and alternative regression model to the beta regression model (LÓPEZ, 2013; SONG; TAN, 2000). A simple advantage of the Simplex distribution is that both, mean and dispersion parameter, are shown explicitly in its probability density function. The distribution can have one or two modes, and cannot emulates a flat distribution as the uniform distribution on the interval  $(0, 1)$ .

In the context of regression model, a distribution not studied yet is the Logit-Logistic

(called here L-Logistic), originally proposed by [Tadikamalla and Johnson \(1982\)](#). This distribution was studied by, among others, [Tadikamalla and Johnson \(1990\)](#) and [Johnson and Tadikamalla \(1991\)](#), who proposed the method of moments and the percentile points method to fit this distribution. However, regression models were not studied by considering this distribution.

In this thesis, we consider the Simplex distribution and the L-Logistic distribution as alternatives to the Beta distribution in some different situations as: mixture model, regression model, among others to model data bounded in  $(0,1)$  interval. The work was motivated by the data of percentages of votes obtained by a political party in elections in different cities of a region of the Brazil seen in [Paz, Bazán and Elher \(2015\)](#), and described here in the Chapter 2. These data sets have some characteristics of Benford's Law (see for exemple [Berdufi \(2014\)](#) and [Cuff, Lewis and Miller \(2014b\)](#)) and recently [Cuff, Lewis and Miller \(2014a\)](#) have established the relation between the Weibull distribution and Benford's Law. Thus, the percentages of votes to each city are assumed here to follow a Weibull distribution. In this work, mixture models are also used in the analysis of the percentages of votes in order to give more flexibility for the model. These data sets are an example of data which need a more flexible distribution to be suitable modeled

The Chapters 3, 4, 5 e 6 of this thesis are based on manuscripts written to present models alternatives to Beta distribution. In the third chapter a mixture of Simplex distributions for modeling proportional data is presented. The Simplex distribution is a distribution recently studied as alternative to Beta distribution ([LÓPEZ, 2013](#); [SONG; TAN, 2000](#)). However, since the data present multimodality we propose a mixture of Simplex distributions for the modeling process. A full Bayesian approach is considered in the inference process in mixture of Simplex distributions and the method adopted is the *Reversible-jump* Markov Chain Monte Carlo. The usefulness of the proposed approach is confirmed by use of the simulated mixture data from several different scenarios and through an application of the methodology to analyze municipal Human Development Index data of the cities of the Northeast region and São Paulo state in Brazil. The work presented in this chapter is a manuscript published in the Journal of Applied Statistics ([PAZ; BAZÁN; MILAN, 2015](#)).

The fourth chapter is dedicated to the analysis of the Municipal Human Development Index as a function of proportion of poor people per municipality. We propose a regression model where the response follow a mixture of Simplex distribution. Estimation is performed also by a Bayesian approach making use of *Gibbs sampling* algorithm. For the choice of the number of component in the mixture, we make a comparison of the models with different components. This chapter present a work published as a expanded abstract for 60<sup>a</sup> Reunião Anual da Região Brasileira da Sociedade Internacional de Biometria (RBras), 2015, conference.

The fifth chapter deal with features of the L-Logistic distribution. As said before, this distribution was originally proposed by [Tadikamalla and Johnson \(1982\)](#) through a transformation of the standard logistic distribution. In the considered parametrization of L-Logistic distribution,

the median is an explicit parameter and we can easily write it as a function of covariates in a regression structure. If the data are highly skewed, where the median is a natural robust measure of the center, the conditional median modeling can be more useful than conditional mean modeling adopted in Beta regression models. For the model without covariate, simulation studies, considering prior sensitivity analysis and comparison with Beta distribution, give evidence that the L-logistic distribution is more robust than Beta distribution to modeling data with outliers. Applications to real and simulated data are also performed. The work presented in this chapter is under review for publication.

Finally, a mixture of L-Logistic mixed-effect models for modeling longitudinal proportion data is proposed and discussed in Chapter 6. These models are applied to simulated data providing good estimates for the parameters of the proposed models. Application to real data was also performed. In this chapter, we present a manuscripts under review for submission to an journal.



---

## A MOTIVATION: STUDY OF THE VOTES OF A BRAZILIAN POLITICAL PARTY

---

---

This chapter shows an application of a mixture model to analyze data of the percentage of votes under Bayesian approach. We give a description of the mixture model considering each component of the mixture as a Weibull distribution. In order to decide about the number of components of the mixture model, a model comparison was conducted using the Bayes factor.

---

### Abstract

Statistical modeling in Political Analysis has been used to describe electoral behavior of political party. In this paper we propose a Weibull mixture model to describe the votes obtained by a political party in Brazilian presidential elections. We considered the votes obtained by the Partido dos Trabalhadores in five presidential elections from 1994 to 2010. A Bayesian approach was considered and a random walk Metropolis algorithm within *Gibbs sampling* was implemented. Next, Bayes factor was considered to the choice of the number of components in the mixture. In addition the probability of obtain 50 percent of the votes in the first round was estimated. The results show that only few components are needed to describe the votes obtained in this election. Finally, we found that the probability of obtaining 50 percent of the votes in the first ballot is increasing along time. Future developments are discussed.

## 2.1 Introduction

Statistical modeling in Political Analysis has been used recently to describe the electoral behavior of a political party and examples of study of voting behavior are [Jones and Johnston \(1992\)](#). In Brazil the electoral behavior underwent a process of change since 1994 to the recent days. In 1994, Brazilians voted in one of the most important elections held since 1945. This was the second election held since the end of military rule from 1964 to 1985. In terms of the presidential vote, in 1994 the candidate of The Brazilian Social Democracy Party (in Portuguese: Partido da Social Democracia Brasileira, PSDB) won the majority of votes on the first ballot (54.3 per cent) and the candidate of Partido dos trabalhadores (PT) obtained 38.4 per cent of the total of votes. For more information about this election see for example [Meneguello \(1995\)](#) or the Superior Electoral Court (TSE) website <http://english.tse.jus.br>. However the PT elected its candidates for president in the last 3 elections occurring in 2002, 2006 and 2010. Results on the presidential elections in Brazil are available on the TSE website.

In order to investigate the probabilistic behavior of votes obtained by a political party in Brazilian general elections we propose a Weibull mixture model. In particular, we considered the presidential votes obtained by PT in the five elections from 1994 to 2010 using data obtained from TSE website. As seen in [Bohn \(2011\)](#), analysts have argued that the social policies that President Luis Inacio Lula da Silva implemented enabled the number of voters of the PT to expand from middle-class and highly educated people to low-income and poorly educated individuals from the Northeast of Brazil. From the 9 states in the Northeast region we chose to analyze the data from Sergipe State (SE) for illustration purposes because this is the state with the smallest number of electoral districts, being 75 municipalities.

For estimation purposes, a Bayesian approach was considered and a random walk *Metropolis-Hasting* algorithm within *Gibbs sampling* was implemented. Next, a Bayes factor

approach was considered to the choice of the number of components in the mixture. Finally the probability of obtaining 50 percent of the votes in the first ballot was estimated. The results show that only a few components are needed in the mixture to describe the votes obtained in this election. In addition we found that the probability of obtaining 50 percent of the votes in the first ballot is increasing along time.

The rest of the work is organized as follows. In Section 2.2 we give a description of the data. In Section 2.3, we describe the finite mixture model and a finite mixture of Weibull distributions is proposed to model the data of votes of a Brazilian political party. In Section 2.4 we present the main results and the future developments are discussed in Section 2.5 .

## 2.2 The votes of a political party

Percentages of votes obtained by a political party in an election in different cities of a region or country can be assumed as a random variable  $X > 0$  due to because, as suggested by Cuff, Lewis and Miller (2014b), these are some characteristics of Benford's Law which has been invoked as evidence of elections data for example by Berdufi (2014). Benford's law, also called the first-digit law, is an observation about the frequency distribution of leading digits in many real-life sets of numerical data. This law of Leading Digits proposes a distribution for the significands (or significant digits) which holds for many data sets, and states that the proportion of values beginning with digit  $d$ ,  $d \in 1, \dots, 9$ , is approximately  $\text{Prob}(d) = \log_{10} \left( \frac{d+1}{d} \right)$ . There have been numerous attempts to pass from observing the prevalence of Benford's law to explaining its occurrence in different and diverse systems. Such knowledge gives us a deeper understanding of which natural data sets should follow Benford's law. A good recent description of this approach is given in Fewster (2009). Moreover, Cuff, Lewis and Miller (2014a) have established the relation between the Weibull distribution which the support is positive real line and Benford's Law. Thus, the Weibull distribution is used here for model data of percentage of votes. Note that, since percentage data are in bounded interval, a distribution with support in bounded interval would also be suitable for modeling such data. However in this work we use the Weibull model following the literature of the area.

As usually observed in the Histogram of the percentage data, they present a positive asymmetric distribution, that is, the votes are concentrated in lower percents and occasionally are observed higher values and the mean is greater than the median. The percentages of votes to each city are assumed here to follow a Weibull distribution which is governed by two parameters, that is  $X \sim \text{Weibull}(\delta, \eta)$ . Being zero the lower end of its support. The parameter  $\delta$  is a shape parameter and  $\eta$  is a scale parameter. The parameter  $\eta$  determines the scale along its support of votes and the parameter  $\delta$ , determines the concentration of the distribution of votes. High values of  $\eta$  correspond to a high degree of concentration (low dispersion) of votes. The Weibull distribution can be seen as a generalization of the Exponential distribution and commonly descri-

bes the time we have to wait for one event to occur, if that event becomes more or less likely with time. Here the  $\eta$  parameter describes how quickly the probability ramps up (proportional to  $x^{\eta-1}$ ). For  $0 < \eta < 1$ , the density function tends to  $+\infty$  if  $x$  approaches zero from above and is strictly decreasing. For  $\eta = 1$ , the density function of votes tends to  $1/\delta$  to lower votes  $x$  approaches zero from above and is strictly decreasing. For  $\eta > 1$ , the density function of votes tends to zero as the votes  $x$  approaches zero from above, increases until its mode  $\delta \left(\frac{\eta-1}{\eta}\right)^{1/\eta}$  and decreases after it.

Additionally by observing the histogram of percent of votes by cities in Figure 1, we can see multimodality in the data. Thus, may be it is possible to identify different populations (clusters of votes), probably because there are different electoral behaviors between cities. Consequently a Weibull mixture distribution can be assumed in order to identify these sub populations. In [Tsonas \(2002\)](#) this type of distribution has been considered in different areas but similar situations.

## 2.3 The general mixture model

Finite mixture of distributions is a flexible method of data modeling. Its more direct role in data analysis and inference is to provide a convenient and flexible family of distributions to estimate or approximate distributions which are not well modeled by any standard parametric family. This type of model is useful in the modeling of data from a heterogeneous population, that is, a population which can be divided in clusters or components. In this sense, the components in the data can be modeled for uni-modal distributions. For more details about modeling and applications of finite mixture models, see for example [McLachlan and Peel \(2004\)](#).

By observation of the data in Figure 1 we propose to model these data as a  $k$ -component mixture of distributions. This approach is flexible enough to model the data that is shown in such situation in Figure 1, where we can see the multimodality phenomenon.

A random variable  $X$  is said to follow a mixture of distributions with  $k$  components if its probability density function (pdf) is given by

$$f(x|\boldsymbol{\theta}, \boldsymbol{\omega}, k) = \sum_{j=1}^k \omega_j f_j(x|\boldsymbol{\theta}_j) \quad (2.1)$$

where each  $f_j(x|\boldsymbol{\theta}_j)$  is a pdf called component density of the mixture, indexed by a parameter vector  $\boldsymbol{\theta}_j$  (here we write  $f(x|\boldsymbol{\theta}_j)$  without the index  $j$  because the component density belong to the same parametric family),  $\boldsymbol{\theta} = (\boldsymbol{\theta}_1, \dots, \boldsymbol{\theta}_k)$  is a vector containing all the parameters of the components in the mixture and the components of the vector  $\boldsymbol{\omega} = (\omega_1, \dots, \omega_k)$  are called weights of the mixture where  $0 < \omega_j < 1$  with  $\sum_{j=1}^k \omega_j = 1$ . In the equation (2.1)  $k$  is the number of components in the mixture. We call the model defined by the pdf in (2.1) mixture model which the distribution is called mixtures of distributions. For a review on existing techniques for



Bayesian modeling and inference on mixtures of distributions, see for example [Marin, Mengersen and Robert \(2005\)](#).

In order to make inference about the parameters of the mixture model, suppose  $\mathbf{X} = (X_1, \dots, X_n)$  a random sample from the distribution defined by equation (2.1). The likelihood related to a sample  $\mathbf{x} = (x_1, \dots, x_n)$ , where each  $x_i$  is a observation of  $X_i$  for  $i = 1, \dots, n$ , is given by

$$L(\boldsymbol{\theta}, \boldsymbol{\omega} | \mathbf{x}, k) = \prod_{i=1}^n \sum_{j=1}^k \omega_j f(x_i | \boldsymbol{\theta}_j).$$

A way to simplify the inference process of mixture model is to consider a unobserved random vector  $Z_i = (Z_{i1}, \dots, Z_{ik})$  such that  $Z_{ij} = 1$  if the  $i$ th observation is from the  $j$ th mixture component and  $Z_{ij} = 0$  otherwise,  $i = 1, \dots, n$ . Note that  $\sum_{j=1}^k Z_{ij} = 1$  then we suppose each random vector  $Z_1, \dots, Z_n$  is distributed according to the multinomial distribution with parameters 1 and  $\boldsymbol{\omega} = (\omega_1, \dots, \omega_k) = (P(Z_{i1} = 1 | \boldsymbol{\omega}, k), \dots, P(Z_{ik} = 1 | \boldsymbol{\omega}, k))$ , for  $i = 1, \dots, n$ . Then

$$P(Z_{ij} = 1 | x_i, \boldsymbol{\theta}, \boldsymbol{\omega}, k) \propto P(Z_{ij} = 1 | \boldsymbol{\omega}, k) f(x_i | Z_{ij} = 1, \boldsymbol{\theta}, \boldsymbol{\omega}, k),$$

$j = 1, \dots, k$ ,  $i = 1, \dots, n$ . To simplify the notation we consider  $\mathbf{Z} = (Z_1, \dots, Z_n)$  a vector  $nk$  containing all the unobserved indicator vectors  $Z_i$ . Note that the distribution of each  $X_i$  given  $Z_i$  has pdf given by

$$f(x_i | Z_i, \boldsymbol{\theta}, k) = \prod_{j=1}^k [f(x_i | \boldsymbol{\theta}_j)]^{Z_{ij}} \quad (2.2)$$

then the joint distribution of  $(X_i, Z_i)$  can be written as

$$f(x_i, Z_i | \boldsymbol{\theta}, \boldsymbol{\omega}, k) = P(Z_i | \boldsymbol{\omega}, k) f(x_i | Z_i, \boldsymbol{\theta}, k) = \prod_{j=1}^k [\omega_j f(x_i | \boldsymbol{\theta}_j)]^{Z_{ij}}. \quad (2.3)$$

Note that, the vector  $Z_i$  have just one component equal to 1 and the others equal to zero then

$$\prod_{j=1}^k [\omega_j f(x_i | \boldsymbol{\theta}_j)]^{Z_{ij}} = \begin{cases} \omega_1 f(x_i | \boldsymbol{\theta}_1) & \text{if } Z_i = (1, 0, \dots, 0) \\ \omega_2 f(x_i | \boldsymbol{\theta}_2) & \text{if } Z_i = (0, 1, \dots, 0) \\ \vdots & \vdots \\ \omega_k f(x_i | \boldsymbol{\theta}_k) & \text{if } Z_i = (0, 0, \dots, 1) \end{cases}$$

thus,

$$f(x_i | \boldsymbol{\theta}, \boldsymbol{\omega}, k) = \sum_{Z_i} f(x_i, Z_i | \boldsymbol{\theta}, \boldsymbol{\omega}, k) = \sum_{j=1}^k \omega_j f(x_i | \boldsymbol{\theta}_j). \quad (2.4)$$

After the inclusion of the indicator vectors in the model, the augmented data likelihood to  $(\mathbf{x}, \mathbf{Z})$  can be written as

$$L(\boldsymbol{\theta}, \boldsymbol{\omega} | \mathbf{x}, \mathbf{Z}, k) = \prod_{i=1}^n \prod_{j=1}^k [\omega_j f(x_i | \boldsymbol{\theta}_j)]^{Z_{ij}}. \quad (2.5)$$

Finally, the joint distribution of all variables of the model including the augmented version and the prior specifications is

$$P(\mathbf{x}, \boldsymbol{\theta}, \mathbf{Z}, \boldsymbol{\omega} | k) = f(\mathbf{x} | \boldsymbol{\theta}, \mathbf{Z}, \boldsymbol{\omega}, k) P(\boldsymbol{\theta} | \mathbf{Z}, \boldsymbol{\omega}, k) P(\mathbf{Z} | \boldsymbol{\omega}, k) P(\boldsymbol{\omega} | k).$$

A common approach is to impose conditional independence (BOUGUILA; ELGUEBALY, 2012) such that  $P(\boldsymbol{\theta} | \mathbf{Z}, \boldsymbol{\omega}, k) = P(\boldsymbol{\theta} | \mathbf{Z}, k)$ ,  $f(\mathbf{x} | \boldsymbol{\theta}, \mathbf{Z}, \boldsymbol{\omega}, k) = f(\mathbf{x} | \boldsymbol{\theta}, \mathbf{Z}, k)$  leading to the joint distribution

$$f(\mathbf{x}, \boldsymbol{\theta}, \mathbf{Z}, \boldsymbol{\omega} | k) = f(\mathbf{x} | \boldsymbol{\theta}, \mathbf{Z}, k) P(\boldsymbol{\theta} | \mathbf{Z}, k) P(\mathbf{Z} | \boldsymbol{\omega}, k) P(\boldsymbol{\omega} | k), \quad (2.6)$$

where  $P(\mathbf{Z} | \boldsymbol{\omega}, k) = \prod_{i=1}^n \left( \prod_{j=1}^k \omega_j^{Z_{ij}} \right)$  and  $f(\mathbf{x} | \boldsymbol{\theta}, \mathbf{Z}, k) = \prod_{i=1}^n \prod_{j=1}^k [f(x_i | \theta_j)]^{Z_{ij}}$ . This hierarchical representation of the model facilitates the Bayesian analysis because it allows the use of Markov chain Monte Carlo (MCMC) technique. Here, we consider the number of component  $k$  as a known constant however the value of  $k$  can be considered as unknown, and in this case the number of component is also a parameter to be estimated.

## 2.4 The Weibull mixture model

Here, we assume a finite mixture of Weibull distributions for each  $X_i$ ,  $i = 1, \dots, n$ , where the  $j$ th component has scale and shape parameters  $\eta_j$  and  $\delta_j$  respectively. We prefer Weibull distribution since that gives a distribution for which the failure rate is proportional to a power of time and the parameters of the model are easily interpretable. Consequently other distributions were discarded.

Considering Weibull distributions as components in the mixture model, the augmented data likelihood function is given by

$$L(\boldsymbol{\theta}, \boldsymbol{\omega} | \mathbf{x}, k) = \prod_{i=1}^n \prod_{j=1}^k \left[ \omega_j \frac{\delta_j}{\eta_j} \exp \left( - \left( \frac{x_i}{\eta_j} \right)^{\delta_j} \right) \left( \frac{x_i}{\eta_j} \right)^{\delta_j - 1} \right]^{Z_{ij}} \quad (2.7)$$

where  $\boldsymbol{\theta} = (\theta_1, \dots, \theta_k)$  with  $\theta_j = (\delta_j, \eta_j)$  for  $j = 1, \dots, k$ .

Following the Bayesian paradigm, we need to complete the model specification by assigning prior distributions to the parameters. Then, by applying the Bayes theorem the posterior density is proportional to the product the likelihood function (2.7) by the prior density.

We shall assume that all the parameters are a priori independent. Then, within each component, Gamma prior distributions are assigned to the Weibull parameters, i.e.  $\eta_j \sim \text{Gamma}(a_j, b_j)$  and  $\delta_j \sim \text{Gamma}(c_j, d_j)$ ,  $j = 1, \dots, k$ , here the notation  $Y \sim \text{Gamma}(c_j, d_j)$  means that the random variable  $Y$  follows a Gamma distribution with parameters  $c_j$  and  $d_j$ .

Also, since the vector of weights  $\boldsymbol{\omega}$  is defined on the Simplex  $\{\boldsymbol{\omega} \in \mathbb{R}^k : 0 < \omega_j < 1, j = 1, \dots, k, \sum_{j=1}^k \omega_j = 1\}$  we consider a Dirichlet prior distribution for  $\boldsymbol{\omega}$  which pdf is given by

$$p(\boldsymbol{\omega} | \mathbf{v}_1, \dots, \mathbf{v}_k, k) = \frac{\Gamma(\mathbf{v}_1 + \dots + \mathbf{v}_k)}{\Gamma(\mathbf{v}_1) \dots \Gamma(\mathbf{v}_k)} \prod_{j=1}^k \omega_j^{\mathbf{v}_j - 1} \quad (2.8)$$

where  $\mathbf{v}_1 > 0, \dots, \mathbf{v}_k > 0$  are the hyperparameters. In this manuscript, the hyperparameters  $a_j, b_j, c_j, d_j$  and  $\mathbf{v}_j, j = 1, \dots, k$  are held fixed.

Finally, we need to impose identifiability constraints since the labelling of the mixing components is arbitrary and we need some rule to discriminate among the components (see for example [Holmes, Jasra and Stephens \(2005\)](#)). A typical solution, also adopted here, is to impose an ordering constraint  $\mu_1 \leq \mu_2 \leq \dots \leq \mu_k$  where  $\mu_j$  is the mean of the  $j$ th component in the mixture.

Since the posterior density cannot be fully obtained in closed form we use MCMC approach to simulate parameter values and obtain parameter estimates. Details of MCMC methods can be found for example in [Robert and Casella \(2005\)](#). In order to obtain a sample from the joint posterior distribution of the parameters we first obtain the complete conditional distributions. First note that

$$\begin{aligned} P(Z_{ij} = 1 | x_i, \boldsymbol{\theta}, \boldsymbol{\omega}, k) &= \frac{P(Z_{ij} = 1 | \boldsymbol{\theta}, \boldsymbol{\omega}, k) f(x_i | Z_{ij} = 1, \boldsymbol{\theta}, k)}{f(x_i | \boldsymbol{\theta}, \boldsymbol{\omega}, k)} \\ &= \frac{\omega_j \frac{\delta_j}{\eta_j} \exp\left(-\left(\frac{x_i}{\eta_j}\right)^{\delta_j}\right) \left(\frac{x_i}{\eta_j}\right)^{\delta_j - 1}}{\sum_{j=1}^k \left[\omega_j \frac{\delta_j}{\eta_j} \exp\left(-\left(\frac{x_i}{\eta_j}\right)^{\delta_j}\right) \left(\frac{x_i}{\eta_j}\right)^{\delta_j - 1}\right]} \end{aligned} \quad (2.9)$$

for  $i = 1, \dots, n$ . for  $i = 1, \dots, n$ . So, for each observation we just need to sample  $j \in \{1, \dots, k\}$  with probability given by (2.9). Now, combining the likelihood function (2.7) with the prior densities of  $\delta_j$  and  $\eta_j$  it follows that,

$$\begin{aligned} P(\eta_j | \mathbf{x}, Z, \boldsymbol{\theta}_{-\eta_j}) &\propto \eta_j^{a_j - n_j \delta_j - 1} \exp\left\{-\sum_{i: Z_{ij}=1} \left(\frac{x_i}{\eta_j}\right)^{\delta_j} - \eta_j b_j\right\} \\ P(\delta_j | \mathbf{x}, Z, \boldsymbol{\theta}_{-\delta_j}) &\propto \delta_j^{n_j + c_j - 1} \eta^{-n_j \delta_j} \exp\left\{-\sum_{i: Z_{ij}=1} \left(\frac{x_i}{\eta_j}\right)^{\delta_j} - d_j \delta_j\right\} \prod_{i: Z_{ij}=1} x_i^{\delta_j - 1} \end{aligned}$$

where  $n_j = \sum_{i=1}^n Z_{ij}$  denotes the number of observations in the  $j$ th mixture component and  $\boldsymbol{\theta}_{-\delta_j}$  denote the vector of all the parameters of the components of the mixture except  $\delta_j$ .

The complete conditional density of each  $\delta_j$  and  $\eta_j$  is not of any standard form and we use a *Metropolis-Hastings* algorithm. We adopt a random walk Metropolis algorithm by proposing values of  $\log(\delta_j)$  and  $\log(\eta_j)$  from a Normal distribution centred about its current value and fixing the variance to tune the acceptance rates between 0.4 and 0.6.

Finally, the complete conditional density of  $\boldsymbol{\omega}$  is given by

$$P(\boldsymbol{\omega}|\mathbf{x}, Z, \boldsymbol{\theta}) \propto \prod_{j=1}^k \omega_j^{v_j+n_j-1}$$

which represents a Dirichlet distribution with parameters  $v_1 + n_1, \dots, v_k + n_k$ . Sampling from this complete conditional distribution is then accomplished by drawing independent Gamma variables and scaling them to sum to 1.

Here, the *Gibbs sampling* method is used combined with *Metropolis-Hastings* algorithm for obtain sample from the posterior distribution of parameters  $\delta_1, \dots, \delta_k, \eta_1, \dots, \eta_k, \boldsymbol{\omega}$  and  $Z_i$ , for  $i = 1, \dots, n$ , see [Marin, Mengersen and Robert \(2005\)](#). The Gibbs sampling algorithm can be written as follows.

---

**Algorithm 1** – Algorithm for simulating samples from the posterior distribution of the parameters of the mixture of Weibull

---

1. Initialize choosing  $\boldsymbol{\omega}^{(0)}, \delta_j^{(0)}$  and  $\eta_j^{(0)}$ , for  $j = 1, \dots, k$ .
2. For  $t = 1, 2, \dots$  repeat

- a) For  $i = 1, \dots, n$  generate  $Z_i^{(t+1)} \sim \text{Multinomial}(1, \hat{\pi}_{i1}^{(t)}, \dots, \hat{\pi}_{ik}^{(t)})$  wherein

$$\hat{\pi}_{ij}^{(t)} = P(Z_{ij}^{(t)} = 1 | x_i, \delta_j^{(t-1)}, \eta_j^{(t-1)}) = \frac{f(x_i | \delta_j^{(t-1)}, \eta_j^{(t-1)}, k) \omega_j^{(t-1)}}{\sum_{j=1}^k \omega_j^{(t-1)} f(x_i | \delta_j^{(t-1)}, \eta_j^{(t-1)}, k)}. \quad (2.10)$$

- b) Generate  $\boldsymbol{\omega}^{(t)}$  from the  $P(\boldsymbol{\omega} | Z^{(t)})$ .

- c) For  $j = 1, \dots, k$  do

- i. Generate  $(\delta'_j, \eta'_j) \sim \text{Lognormal} \left( \left( \log(\delta_j^{(t-1)}), \log(\eta_j^{(t-1)}) \right), \sigma_j^2 I \right)$   
with  $\sigma_j^2 = 0.05$ .

- ii. Generate  $u \sim \text{Uniform}(0, 1)$

- iii. Compute

$$\alpha \left( \left( \delta_j^{(t-1)}, \eta_j^{(t-1)} \right), \left( \delta'_j, \eta'_j \right) \right) =$$

$$\min \left\{ 1, \frac{P \left( \left( \delta'_j, \eta'_j \right) | \mathbf{x}, Z \right)}{P \left( \left( \delta_j^{(t-1)}, \eta_j^{(t-1)} \right) | \mathbf{x}, Z \right)} \frac{LN \left( \left( \delta_j^{(t-1)}, \eta_j^{(t-1)} \right) | \left( \log(\delta'_j), \log(\eta'_j) \right), \sigma_j^2 I \right)}{LN \left( \left( \delta'_j, \eta'_j \right) | \left( \log(\delta_j^{(t-1)}), \log(\eta_j^{(t-1)}) \right), \sigma_j^2 I \right)} \right\}$$

where  $LN(y|\cdot)$  is the density of Log-Normal distribution evaluate at  $y$ .

- iv. If  $\alpha \left( \left( \delta_j^{(t-1)}, \eta_j^{(t-1)} \right), \left( \delta'_j, \eta'_j \right) \right) < u$  then  $(\delta_j^{(t)}, \eta_j^{(t)}) = (\delta'_j, \eta'_j)$  else  
 $(\delta_j^{(t)}, \eta_j^{(t)}) = (\delta_j^{(t-1)}, \eta_j^{(t-1)})$ .
- 

## 2.5 Choosing the number of components in the mixture model

The specification of a mixture model involves the determination of the number of components  $k$ . Here, instead of endeavouring to apply more complex methods as in [Richardson and Green \(1997\)](#) and [Stephens \(2000\)](#) for example, we compare models by means of Bayes factor

and marginal likelihood (BERKHOF; MECHELEN; GELMAN, 2003). In order to describe de Bayes factor suppose two models  $M_s$  and  $M_r$  with equal prior probabilities  $P(M_s)$  and  $P(M_r)$ . The Bayes factor is obtained as the ratio of marginal likelihood  $m_s(\mathbf{x})$  and  $m_r(\mathbf{x})$ , such that

$$B_{sr} = \frac{f(\mathbf{x}|M_s)}{f(\mathbf{x}|M_r)} = \frac{P(M_s|\mathbf{x}) P(M_r)}{P(M_r|\mathbf{x}) P(M_s)} = \frac{m_s(\mathbf{x})}{m_r(\mathbf{x})} \quad (2.11)$$

where  $P(M_s|\mathbf{x})/P(M_r|\mathbf{x})$  is the posterior odds and  $P(M_s)/P(M_r)$  is the prior odds. Since the Bayes factor is higher than 1 then  $M_s$  has a higher posterior probability.

In particular, the marginal likelihood for the  $k$ -component mixture model is given by,

$$f(\mathbf{x}|k) = \int f(\mathbf{x}|\Theta, k) P(\Theta|k) d\Theta,$$

where  $\Theta = (\boldsymbol{\theta}, \boldsymbol{\omega})$  is a vector containing all the parameters of the model. Computation of the marginal likelihood requires proper prior distributions and the analytic evaluation of this integral is not possible in the context treated here (see for example Chen, Shao and Ibrahim (2000) for an extensive description and comparison of available numerical strategies). In this paper, we compute an approximation to the marginal likelihood based on the MCMC output using the methods described in Chib and Jeliazkov (2001). The estimator is based on the identity

$$m(\mathbf{x}) = \frac{f(\mathbf{x}|\Theta^*, k) P(\Theta^*|k)}{P(\Theta^*|\mathbf{x}, k)} \quad (2.12)$$

where the numerator can be directly computed. Thus the calculation of the marginal likelihood is reduced to finding an estimate of the posterior density at a point  $\Theta^*$ . For estimation efficiency we take the point  $\Theta^* = (\boldsymbol{\theta}^*, \boldsymbol{\omega}^*)$  as the posterior mean of  $\Theta$  in the  $k$ -component model. We now drop the dependence on  $k$  to simplify the notation. Note that the posterior density ordinate can be rewritten as,

$$P(\Theta^*|\mathbf{x}) = P(\boldsymbol{\theta}^*|\mathbf{x}) P(\boldsymbol{\omega}^*|\mathbf{x}, \boldsymbol{\theta}^*) = P(\theta_1^*|\mathbf{x}) \dots P(\theta_k^*|\mathbf{x}) P(\boldsymbol{\omega}^*|\mathbf{x}, \boldsymbol{\theta}^*)$$

where  $\theta_j^* = (\delta_j^*, \eta_j^*)$ , for  $j = 1, \dots, k$ . Our approach is based on an additional  $G$  iterations sampling values of  $Z$  from its complete conditional distributions evaluated at  $(\delta_j^*, \eta_j^*)$  and sampling values of  $(\delta_j, \eta_j)$  from its proposal distribution in the Metropolis-Hastings step also evaluated at  $(\delta_j^*, \eta_j^*)$ .

Chib and Jeliazkov (2001) introduce a way to proximate a pdf of the distribution when it is intractable in the context of MCMC chains produced by *Metropolis-Hastings*. Here, we applied this method over each component of the mixture to approximate each density  $P(\theta_j^*|\mathbf{x})$ . In this context, each density can be written in terms of the expectancy as

$$P(\theta_j^*|\mathbf{x}) = \frac{E \left[ \alpha \left( (\delta_j, \eta_j), (\delta_j^*, \eta_j^*) | \mathbf{x}, Z \right) q \left( (\delta_j, \eta_j), (\delta_j^*, \eta_j^*) | \mathbf{x}, Z \right) \right]}{E \left[ \alpha \left( (\delta_j^*, \eta_j^*), (\delta_j, \eta_j) | \mathbf{x}, Z \right) \right]} \quad (2.13)$$

where  $q((\delta_j, \eta_j), (\delta_j', \eta_j'))$  denote the proposal density of  $(\delta_j, \eta_j)$  in the random walk Metropolis update and  $\alpha(\cdot, \cdot)$  are the acceptance probabilities in the *Metropolis-Hastings* step of the

algorithm. The expectancy in the numerator of equation in (2.13) is related to the distribution defined by  $P(\theta_j, Z|\mathbf{x})$  and the expectancy in the denominator the expectancy is with respect to the distribution defined by  $P(Z|\mathbf{x}, \theta_j^*)q(\theta_j^*, \theta_j|\mathbf{x}, Z)$ . Thus, the marginal density of each  $\theta_j^* = (\delta_j^*, \eta_j^*)$  can be estimated as

$$\hat{P}(\theta_j^*|\mathbf{x}) = \frac{L^{-1} \sum_{l=1}^L \alpha\left((\delta_j^{(l)}, \eta_j^{(l)}), (\delta_j^*, \eta_j^*)|\mathbf{x}, Z^{(l)}\right) q\left((\delta_j^{(l)}, \eta_j^{(l)}), (\delta_j^*, \eta_j^*)|\mathbf{x}, Z^{(l)}\right)}{G^{-1} \sum_{g=1}^G \alpha\left((\delta_j^*, \eta_j^*), (\delta_j^{(g)}, \eta_j^{(g)})|\mathbf{x}, Z^{(g)}\right)} \quad (2.14)$$

where  $\{(\delta_j^{(1)}, \eta_j^{(1)}, Z^{(1)}), \dots, (\delta_j^{(L)}, \eta_j^{(L)}, Z^{(L)})\}$ , in the numerator of equation (2.14), are sampled values from the full run of Algorithm (1). For the denominator we need to sample  $\{(\delta_j^{(1)}, \eta_j^{(1)}, Z^{(1)}), \dots, (\delta_j^{(G)}, \eta_j^{(G)}, Z^{(G)})\}$  from the distribution defined by the pdf  $P(Z|\mathbf{x}, \theta_j^*)q(\theta_j^*, \theta_j|\mathbf{x}, Z)$ . For this purpose we continue the MCMC simulation for additional  $G$  iterations keeping  $\theta_j^*$  fixed and at each iteration of this reduced run we generate

$$\theta_j^g \sim q(\theta_j^*, \theta_j|\mathbf{x}, Z^g)$$

The process of obtainment of the samples to estimate each  $P(\theta_j^*|\mathbf{x})$ , for  $j = 1, \dots, k$ , can be made simultaneously and after this process we can estimate  $P(\boldsymbol{\theta}^*|\mathbf{x})$  as

$$\hat{P}(\boldsymbol{\theta}^*|\mathbf{x}) = \prod_{j=1}^k \hat{P}(\theta_j^*|\mathbf{x})$$

The conditional density ordinates of  $\boldsymbol{\omega}$  is estimated by averaging with respect to the sampled values  $Z^{(g)}$ , for  $g = 1, \dots, G$ , i.e.

$$\hat{P}(\boldsymbol{\omega}^*|\mathbf{x}, \boldsymbol{\theta}^*) = G^{-1} \sum_{g=1}^G P(\boldsymbol{\omega}^*|\mathbf{x}, Z^{(g)}, \boldsymbol{\theta}^*).$$

Finally, the posterior density ordinate is estimated as

$$\hat{P}(\boldsymbol{\Theta}^*|\mathbf{x}) = \prod_{j=1}^k \hat{P}(\theta_j^*|\mathbf{x}) \hat{P}(\boldsymbol{\omega}^*|\mathbf{x}, \boldsymbol{\theta}^*),$$

which is in turn used in (2.12) to obtain an estimate of the marginal likelihood.

### Predictive distribution

A posterior feature of interest is the predictive distribution for a future observation. As discussed by Escobar and West (1995), a density estimation can be obtained by summarizing the unconditional predictive distribution

$$h(x) = p(x_{N+1}|\mathbf{x}) = \int p(x_{N+1}|\boldsymbol{\Theta}) dp(\boldsymbol{\Theta}|\mathbf{x}) = E_{\boldsymbol{\Theta}|\mathbf{x}}[f(x|\boldsymbol{\Theta})]. \quad (2.15)$$

Thus, the Monte Carlo approximation for  $h(x)$  is obtained as

$$\hat{h}(x) = \frac{1}{L} \sum_{l=1}^L f(x|\boldsymbol{\Theta}^{(l)}) \quad (2.16)$$

where  $\left\{ \Theta^{(l)} \right\}_{l=1}^L$  are draws from the joint posterior distribution. In this work, the posterior predictive distribution is used to calculate cumulative probabilities by use of numerical integration such as the Simpson rule, see details of numerical integration methods in [Atkinson \(2008\)](#).

## 2.6 Results

As mentioned in the Section 3.1, we consider the percentage of votes of each municipality in the Sergipe which has 75 municipalities. The distribution of the percentage of votes is presented in the Figure 1. For each data set of vote percentage, obtained considering elections of 1994, 1998, 2002, 2006 and 2010, as seen above, we have implemented the Igorithm 1 to mixtures of Weibull distributions in R language ([R Development Core Team, 2016](#)). In terms of MCMC, we report results corresponding to 10000 iterations following a burn-in period also of 10000 iterations. The convergence of MCMC chain is assessed using separated partial means test proposed by [Geweke \(1992\)](#) and all indicate that the chains have converged. The values of hyperparameters in the prior distributions were specified to produce approximately vague prior. Thus, for the five elections and each number of components in the mixture we specified  $a_j = (4, 5, 5, 7, 7)$ ,  $c_j = (49, 49, 45, 10, 10)$  and  $d_j = (7, 7, 5, 1, 1/2)$ . Also, for all elections we set  $b_j = 1/10$ , and  $v_j = 1$ . The main variation chosen was in the hyper parameter for shape parameter of Weibull distribution as discussed in section 2.2. Thus for elections in 2006 and 2010 smaller values of these hyperparameters were chosen in order to reflect the greater dispersion of the distribution of the data. The acceptance rate in the *Metropolis-Hastings* algorithm for sampling  $\delta_j$  and  $\eta_j$  was controlled to lie within the interval 0.20–0.50 which is usually recommended in the MCMC literature.

Table 1 – Twice the natural logarithm of the Bayes factor of the data of voting percentage under one model resulting of mixture of Weibull distribution relative to another.

Election	$2 \times \log \left( \frac{p(x 2\text{-component})}{p(x 1\text{-component})} \right)$	$2 \times \log \left( \frac{p(x 2\text{-component})}{p(x 3\text{-component})} \right)$	$2 \times \log \left( \frac{p(x 3\text{-component})}{p(x 1\text{-component})} \right)$
1994	560.0	133.3	426.8
1998	753.7	-7.9	761.6
2002	607.9	-3.1	610.9
2006	625.6	-9.4	635.1
2010	562.2	18.3	543.8

We choose a model as the final model among the models with 1, 2 and 3 Weibull components. This model was selected considering twice the natural logarithm of the Bayes factor presented in Table 1, which interpretation can be seen in [Kass and Raftery \(1995\)](#). The results in this Table show that, for all election, models with two or three components are better than a model with one component. However, when comparing models with two or three components the results may vary. For the 1994 and 2010 elections two components in the mixture are sufficient to fit the distribution of the votes. On the other hand, for the 1998, 2002 and 2006 elections three components are needed in the mixture.

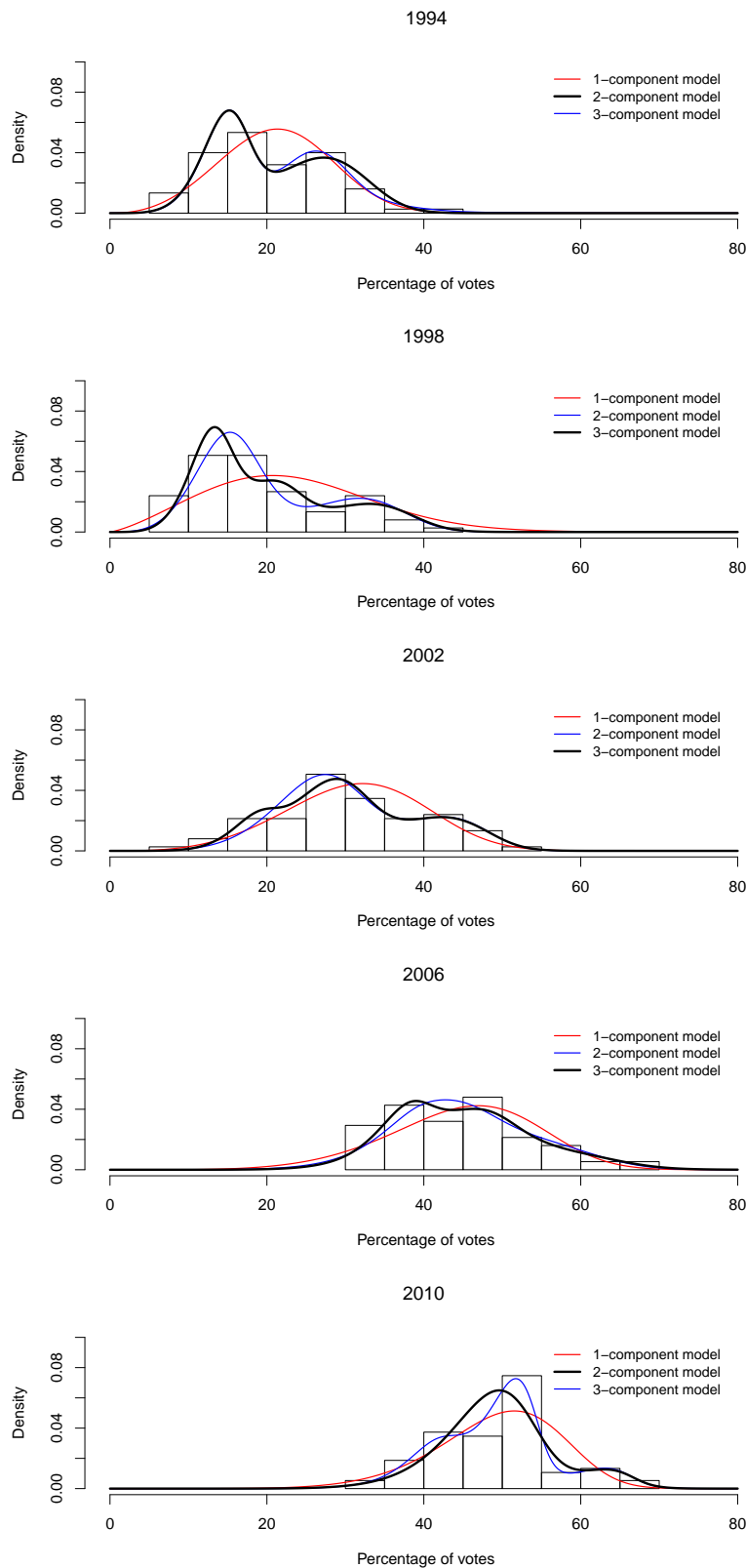


Figure 1 – Histograms of the data of voting percentage obtained by PT in presidential elections, in the cities of Sergipe State, from year 1994 and 1998, when the PT lost the presidential election, to 2002, 2006 and 2010, when the PT candidate was Presidential winner, and its estimated densities based on the posterior predictive distribution for 1, 2 and 3 components.

In addition, considering the posterior mean of the parameters for each model we estimate



the density for each Model and compare with the histogram of votes as showed in Figure 1. We can see that the chosen model, that is the one for which density estimate is closest to the data, coincides with the choice according to Bayes factor.

Table 2 provides posterior means and 95% HPD credible intervals for the parameters in the model chosen according to Bayes factor for each year. The credible intervals were constructed using the package `MCMCpack` of [Martin, Quinn and Park \(2011\)](#).

Using these parameters we can give some interpretations to the results. For example in 1994 we identified two groups of cities in Sergipe. For the first group, formed by 38 cities and with weight 0.45, we found an expected percentage of votes of 14.4 (posterior mean) with variability of 2.9 (posterior standard deviation), whereas in the second group, formed by 37 cities and with weight 0.55, the corresponding values are higher, 26.2 and 5.8 respectively. Likewise, in 2010, two populations were also identified. For the first population, formed by 67 cities with weight 0.84, we found an expected percentage of votes of 47.8 (posterior mean) with variability of 5.7 (posterior standard deviation), whereas in the second group, formed by 8 cities and with weight 0.16, the corresponding values are higher, 60.7 and 4.1 respectively. Note the significant increment of the percent of votes in both populations between 1994 and 2010. In addition, the first population in 2010 has 33 of the cities in the first group in 1994 indicating specifically that this group of cities had a significant increment over time.

Finally, from the best model for each election, the probability that PT obtains more than 50 percent of the votes in the first round was calculated, because if the presidential candidate won the majority of votes in the first ballot the candidate is declared winner of presidential election and the second ballot is not necessary. The probabilities were estimated considering the predictive distribution by numerical integration using the Simpson rule combined with Monte Carlo method as seen in Section 2.5. These corresponding probabilities of winning in the first ballot for PT party considering the Sergipe state for elections in 1994, 1998, 2002, 2006 and 2010 were  $3.15 \times 10^{-6}$ ,  $9.76 \times 10^{-5}$ , 0.0175, 0.273 and 0.459 respectively, it indicate that this probability increased over time.

We should note that as suggested by a referee a Mixture Normal model was also implemented considering an algorithm similar to the one defined in Section 2.3 without *Metropolis-Hastings* step. The results showed, that there is a strong evidence in favour of the Weibull mixture model. Additionally as discussed in Section 2.1 this model can lead to inferences which can be misleading since the Normal is a symmetric distribution and can lead to over-fit when additional component need to be included to capture the asymmetry in the data.

## 2.7 Discussion and further development

This paper proposed a Weibull mixture model to describe the electoral behavior of a Brazilian political party in different elections. The number of votes obtained by PT in the five

Table 2 – Posterior mean and HPD interval of parameters of the best Weibull mixture model chosen by Bayes factor evaluation. Data of voting percentage obtained by PT in presidential elections in the Sergipe State from year 1994 to 2010 was considered for fitting of the models.

Election	k <sup>1</sup>	Posterior mean and HPD Interval (95%)		
		w(weight)	$\delta(shape)$	$\eta(scale)$
1994	2	0.45(0.27, 0.61)	5.81 (4.42, 7.21)	15.57 (14.09, 17.04)
		0.55 (0.39, 0.73)	5.18 (4.00, 6.50)	28.45 (26.97, 31.00)
1998	3	0.47 (0.26, 0.67)	5.38 (3.99, 6.81)	13.91 (12.05, 16.35)
		0.28 (0.09, 0.47)	6.66 (4.72, 8.67)	22.56 (18.13, 29.50)
		0.25 (0.11, 0.42)	6.69 (4.86, 8.52)	34.49 (30.94, 37.92)
2002	3	0.28 (0.05, 0.69)	7.18 (4.70, 9.64)	21.01 (16.52, 29.36)
		0.42 (0.07, 0.66)	8.19 (5.63, 10.88)	31.65 (26.84, 41.88)
		0.30 (0.09, 0.50)	8.57 (6.12, 11.07)	44.01 (40.20, 47.45)
2006	3	0.36 (0.08, 0.72)	11.04 (6.29, 16.50)	39.64 (35.77, 45.15)
		0.42 (0.06, 0.72)	9.65 (5.36, 14.97)	48.69 (43.40, 55.09)
		0.22 (0.02, 0.48)	8.72 (4.93, 13.65)	58.96 (50.89, 67.03)
2010	2	0.84 (0.63, 0.97)	10.10 (6.16, 12.80)	50.17 (48.39, 52.28)
		0.16 (0.03, 0.37)	18.12 (6.27, 29.01)	62.48 (51.91, 66.81)

<sup>1</sup>Number of components in the mixture.

Brazilian presidential elections from 1994 to 2010 were considered for analysis. A fully Bayesian approach was undertaken using MCMC methods.

We note that the results shown in this paper are purely descriptive. They illustrate how the votes of a particular political party in different elections in Brazil in a given geographic area may exhibit multimodality and how the distribution of votes changes over time. Also, we found that the probability of obtaining 50 percent of the votes in the first ballot is increasing over time.

In future developments, the extension of the analysis for all states of Brazil can be considered as well as regression models for explain the electoral conduct. Since the percentage of votes are limited variables, that is, votes is between a minimum and maximum value, models for limited distributions as Beta distributions also can be explored.

The following chapter present a alternative model to data in the unit interval. This model is applied to the percentage of the votes, analyzed here, in the Chapter 6 where we use a mixture of mixed models.

---

# MIXTURE OF SIMPLEX DISTRIBUTIONS WITH UNKNOWN NUMBER OF COMPONENTS

---

---

This chapter addresses the issues that involve Bayesian inference in mixture of Simplex distributions. Unlike what is assumed in the last chapter, here the number of components in the mixture is assumed unknown. In this chapter we develop a mixture model for data on the unit interval.

---

### Abstract

Variables taking values in  $(0, 1)$ , such as rates or proportions, are frequently analyzed by researchers, for instance, political and social data, as well as the Human Development Index. However, sometimes this type of data cannot be modeled adequately using a unique distribution. In this case, we can use mixture of distributions, which is a powerful and flexible probabilistic tool. This manuscript deals with a mixture of Simplex distributions to model proportional data. A fully Bayesian approach is proposed for inference which includes a reversible-jump Markov Chain Monte Carlo procedure. The usefulness of the proposed approach is confirmed by using simulated mixture data from several different scenarios and by using the methodology to analyze municipal Human Development Index data of cities (or towns) in the Northeast region and São Paulo state in Brazil. The analysis shows that among the cities in the Northeast, some appear to have a similar HDI to other cities in São Paulo state.

## 3.1 Introduction

Variable taking values in  $(0, 1)$ , such as index and proportions, are frequently analyzed by researchers, for instance Impartial Anonymous Culture (STENSHOLT, 1999) and the Human Development Index (HDI) (MCDONALD; RANSOM, 2008; CIFUENTES *et al.*, 2008). Sometimes, the data cannot be modeled adequately using a unique distribution as is the case for the proportion of votes obtained by a political party in the Presidential Elections in each city of a country analyzed in the study conducted by Paz, Bazán and Elher (2015). In addition, different components can be identified in the HDI data of several regions in Brazil, see index in PNUD, IPEA and FJP. (2013).

The mixture models can be a powerful and flexible probabilistic tool for modeling many kinds of data, see for example McLachlan and Peel (2004). In financial data, Faria and Gonçalves (2013), can be cited. In addition, a mixture of distributions has been widely analyzed for Normal data, see for example Tanner and Wong (1987), Gelfand and Smith (1990), Diebolt and Robert (1994), Richardson and Green (1997). For data in  $(0,1)$ , there are some studies which consider a finite mixture of Beta distributions (BOUGUILA; ZIOU; MONGA, 2006; BOUGUILA; ELGUEBALY, 2012). However, other probability distributions with support in the interval  $(0,1)$  can be found in the statistics literature, which have not yet been completely analyzed in the context of mixture models, for example the Simplex distribution. The Simplex distribution is a dispersion model proposed by Barndorff-Nielsen and Jorgensen (1991) and has recently been considered as a complementary and alternative regression model when compared to the beta regression model (LÓPEZ, 2013; SONG; TAN, 2000).

This manuscript deals with a new framework for modeling the bounded variables with multimodality as a complementary model to the corresponding beta model. The model proposed

considers a mixture of Simplex distributions with the number of components unknown (Simplex mixture model). This work is motivated by the municipal HDI data in Brazil. Thus, the aim is to identify the number of components and the characteristics of each population identified by the model considering the HDI of the cities of São Paulo state and the Northeast region of Brazil. In order to deal with the problem of estimating the number of components of the mixture model, a reversible-jump Markov chain Monte Carlo (RJMCMC) approach was adopted (see [Green \(1995\)](#) and [Richardson and Green \(1997\)](#)). A RJMCMC procedure is adopted for the mixture of Simplex distributions with a convenient transition function. The results obtained, considering the proposal, are promising since the performance of the method is tested by applying it to simulated data sets from mixtures of Simplex distributions by considering several different scenarios.

In future developments, it can be considered that the phenomenon can be explained by sociological and economic factors which should be included. In addition, the response variable might be associated to geospatial information as potential covariates.

The remainder of the chapter is organized as follows: In Section 2, the mixture of Simplex distributions is presented. Section 3 addresses the Bayesian inference approach by considering a new estimation RJMCMC method for the mixture of Simplex distributions. Section 4 is dedicated to investigating if our algorithm is able to estimate the mixture parameters and select the number of components considering several scenarios of generated data. In Section 5, an analysis of the municipal HDI data is presented. Finally, some conclusions are drawn in Section 6 and in the [Appendice A](#) we present a summary of the algorithm used for simulating samples of the jointly posterior distributions of parameters model.

## 3.2 Simplex Mixture Distribution

Consider initially a sequence of  $k$  continuous random variables all taking values in  $(0,1)$ , each following a distribution with probability density function (pdf)  $P(\cdot|\theta_j)$ ,  $j = 1, \dots, k$ . The parameter values in  $\theta_1, \dots, \theta_k$  can be different leading to a mixture in the sequence of random variables (r.v.). Then, the pdf of a new r.v  $Y$  is defined as

$$P(y|\boldsymbol{\theta}, \boldsymbol{\omega}, k) = \sum_{j=1}^k \omega_j P(y|\theta_j), \quad 0 < y < 1, \quad (3.1)$$

where  $(\boldsymbol{\theta}, \boldsymbol{\omega}, k)$  denotes a vector containing all unknown parameters in the model with  $\boldsymbol{\theta} = (\theta_1, \dots, \theta_k)$ ,  $\boldsymbol{\omega} = (\omega_1, \dots, \omega_k)$ ,  $\omega_j$  is called the mixing proportion satisfying  $\omega_j > 0$  and  $\sum_{j=1}^k \omega_j = 1$ , and  $k$  is the number of components in the mixture, which is assumed unknown. The densities  $P(\cdot|\theta_j)$  shall be referred to as the  $j$ th component density in the mixture and  $k$  as the number of components of the mixture.

The density (3.1) is called a mixture density; its corresponding distribution function is called a mixture of distributions. Details about formulation, interpretation and properties of finite mixture models can be seen in [McLachlan and Peel \(2004\)](#).

In this work, the component densities  $P(\cdot|\theta_j)$  are taken to belong to the Simplex distribution (JØRGENSEN, 1997) whose pdf is given by

$$S(y|\mu, \sigma^2) = \left(2\pi\sigma^2(y(1-y))^3\right)^{-1/2} \exp\left\{-\left(\frac{1}{2\sigma^2}\right)\left(\frac{(y-\mu)^2}{y(1-y)\mu^2(1-\mu)^2}\right)\right\} I_{(0,1)}(y), \quad (3.2)$$

where  $0 < \mu < 1$  is the location parameter and  $\sigma^2 > 0$  is the dispersion parameter. The mean of the Simplex distribution is given by  $E(Y) = \mu$ . Since the component densities  $P(\cdot|\theta_j)$  are assumed to belong to the Simplex distribution family, we shall refer to the component densities in the mixture as Simplex components and the model given by (3.1) as a Simplex Mixture (SM). We shall also rewrite the pdf of this model as

$$P(y|\boldsymbol{\theta}, \boldsymbol{\omega}, k) = \sum_{j=1}^k \omega_j S(y|\theta_j), \quad 0 < y < 1, \quad (3.3)$$

where  $\theta_j = (\mu_j, \sigma_j^2)$ ,  $j = 1, \dots, k$ .

### 3.3 Inference

Consider  $n$  independent r.v.  $\mathbf{Y} = (Y_1, \dots, Y_n)$  of SM model and  $\mathbf{y} = (y_1, \dots, y_n)$  a realization of  $\mathbf{Y}$  where  $y_i$  is the observed value of the  $Y_i$ , for  $i = 1, \dots, n$ , then the likelihood corresponding to a SM model with  $k$ -component is:

$$L(\boldsymbol{\theta}, \boldsymbol{\omega}, k|\mathbf{y}) = \prod_{i=1}^n \sum_{j=1}^k \omega_j S(y_i|\theta_j).$$

A way to simplify the inference process of a mixture model is to consider an unobserved random vector  $Z_i = (Z_{i1}, \dots, Z_{ik})$  such that  $Z_{ij} = 1$  if the  $i$ th observation belongs to the  $j$ th mixture component and  $Z_{ij} = 0$  otherwise,  $i = 1, \dots, n$ . Note that  $\sum_{j=1}^k Z_{ij} = 1$ , and we suppose each random vector  $Z_1, \dots, Z_n$  to be independently distributed according to a multinomial distribution with parameters 1 and  $\boldsymbol{\omega} = (\omega_1, \dots, \omega_k) = (P(Z_{i1} = 1|\boldsymbol{\omega}, k), \dots, P(Z_{ik} = 1|\boldsymbol{\omega}, k))$ , for  $i = 1, \dots, n$ . Then

$$P(Z_{ij} = 1|y_i, \theta_j, \boldsymbol{\omega}, k) \propto P(Z_{ij} = 1|\boldsymbol{\omega}, k)P(y_i|Z_{ij} = 1, \theta_j, \boldsymbol{\omega}, k) = \omega_j S(y_i|\theta_j),$$

$j = 1, \dots, k$ ,  $i = 1, \dots, n$ . To simplify the notation, we consider  $Z = (Z_1, \dots, Z_n)$  the vector of dimension  $nk$  containing all unobserved indicator vectors  $Z_i$ .

The conditional pdf of  $Y_i$  given  $Z_i$  and all parameters can be written as

$$P(y_i|Z_i, \boldsymbol{\theta}, \boldsymbol{\omega}, k) = P(y_i|Z_{ij} = 1, \theta_j, \omega_j, k) = S(y_i|\theta_j), \text{ for } j \text{ such that } Z_{ij} = 1.$$

Then the pdf of each  $Y_i$  given  $Z_i$  is given by

$$P(y_i|\boldsymbol{\theta}, Z_i, \boldsymbol{\omega}, k) = S(y_i|\theta_j) = \prod_{j=1}^k [S(y_i|\theta_j)]^{Z_{ij}}. \quad (3.4)$$

Finally, the joint distribution of  $(Y_i, Z_i)$  has the pdf given by

$$P(y_i, Z_i | \boldsymbol{\theta}, \boldsymbol{\omega}, k) = P(Z_i | \boldsymbol{\theta}, \boldsymbol{\omega}, k) P(y_i | Z_i, \boldsymbol{\theta}, \boldsymbol{\omega}, k) = \prod_{j=1}^k [\omega_j S(y_i | \theta_j)]^{Z_{ij}}, i = 1, \dots, n.$$

Therefore, after the inclusion of the indicator vectors in the model, the augmented data likelihood to  $(\mathbf{y}, Z)$  can be written as

$$L(\boldsymbol{\theta}, \boldsymbol{\omega}, k | \mathbf{y}, Z) = \prod_{i=1}^n \prod_{j=1}^k [\omega_j S(y_i | \theta_j)]^{Z_{ij}}. \quad (3.5)$$

The joint distribution of all variables of the model including the augmented version of the data and the prior specifications is

$$P(\mathbf{y}, \boldsymbol{\theta}, Z, \boldsymbol{\omega}, k) = P(\mathbf{y} | \boldsymbol{\theta}, Z, \boldsymbol{\omega}, k) P(Z | \boldsymbol{\omega}, k) P(\boldsymbol{\theta} | Z, \boldsymbol{\omega}, k) P(\boldsymbol{\omega} | k) P(k).$$

A common approach is to impose conditional independence (BOUGUILA; ELGUEBALY, 2012) such that  $P(\boldsymbol{\theta} | Z, \boldsymbol{\omega}, k) = P(\boldsymbol{\theta} | k)$ ,  $P(\mathbf{y} | \boldsymbol{\theta}, Z, \boldsymbol{\omega}, k) = P(\mathbf{y} | \boldsymbol{\theta}, Z)$  leading to the joint distribution

$$P(\mathbf{y}, \boldsymbol{\theta}, Z, \boldsymbol{\omega}, k) = P(\mathbf{y} | \boldsymbol{\theta}, Z) P(Z | \boldsymbol{\omega}, k) P(\boldsymbol{\theta} | k) P(\boldsymbol{\omega} | k) P(k), \quad (3.6)$$

where  $P(Z | \boldsymbol{\omega}, k) = \prod_{i=1}^n P(Z_i | \boldsymbol{\omega}, k) = \prod_{i=1}^n \left( \prod_{j=1}^k \omega_j^{Z_{ij}} \right)$  and  $P(\mathbf{y} | \boldsymbol{\theta}, Z) = \prod_{i=1}^n P(y_i | \boldsymbol{\theta}, Z_i, \boldsymbol{\omega}, k)$  with  $P(y_i | \boldsymbol{\theta}, Z_i, \boldsymbol{\omega}, k)$  given by (3.4).

The mixture model presented here precludes the use of an improper prior. This is because an improper prior leads to an improper posterior, when some of the component become empty. Thus, for the component parameters  $\theta_j = (\mu_j, \sigma_j^2)$  with  $\phi_j = \sigma_j^{-2}$ ,  $j = 1 \dots k$ , we choose independent priors, that is,  $P(\boldsymbol{\theta} | Z, \boldsymbol{\omega}, k) = P(\boldsymbol{\mu} | k) P(\boldsymbol{\phi} | k)$  such that

$$\mu_j | k \sim \text{Uniform}(0, 1) \text{ and } \phi_j | k \sim \text{Gamma}(a, b), \quad j = 1, \dots, k, \quad (3.7)$$

where the hyperparameters  $a$  and  $b$  are fixed. The scale parameters,  $\mu'_j$ 's, are unknown and assume values in the interval  $(0, 1)$ , therefore the unit Uniform seems a good choice for a vague prior. The Gamma distribution with parameters  $a = b = \varepsilon$ ,  $\varepsilon$  being a small value, is often chosen as a prior distribution for the precision parameter. In the simulations and the application, we consider  $a = 2$  and  $b = 1/2$  then it is expected  $E(\phi) = 4$  and  $V(\phi) = 8$ . Alternative values for hyper-parameters  $a$  and  $b$  are also used. Considering empirical results, we recommend that the mean and variance of the Gamma prior is in the interval  $(0, 10)$ . For  $P(\boldsymbol{\omega} | k)$ , since the vector of weights  $\boldsymbol{\omega}$  is defined on the

Simplex

$\{\boldsymbol{\omega} \in \mathbb{R}^k : 0 < \omega_j < 1, j = 1, \dots, k, \sum_{j=1}^k \omega_j = 1\}$ , it is natural to consider a Dirichlet prior distribution for  $\boldsymbol{\omega}$  given  $k$ , then  $\boldsymbol{\omega} | k \sim \text{Dirichlet}(v_1, \dots, v_k)$ . Some values for the hyper-parameters  $v_j$  were tested for the simulated data sets. We found good results for  $v_1 = v_2 = \dots = v_k = 1$  that is noninformative in the sense of the equiprobability. Finally, for  $K$  we adopted a Uniform discrete distribution between 1 and  $k_{max}$ . The application of the model for several simulated data sets showed a good performance of the model for the values of hyper-parameter used in this work.

Hence, the full conditional posterior distributions can be obtained, and consequently a Markov chain Monte Carlo method (MCMC) (ROSS, 2006, pages, 245 - 271) can be used to sample from the joint probability distribution of the parameters  $(\boldsymbol{\theta}, \boldsymbol{\omega}, k)$ , given the observed data  $\mathbf{y}, Z$ . Then the sample of the joint posterior distribution produced by MCMC is used for Bayesian inference.

The full conditional distributions of the parameters for  $j$ th components are given by

$$P(\phi_j | \mathbf{y}, Z, \mu_j) \propto \phi_j^{n_j/2+a-1} \exp \left\{ -\phi_j \left( \sum_{i \in \{i: Z_{ij}=1\}} \frac{(y_i - \mu_j)^2}{2y_i(1-y_i)\mu_j^2(1-\mu_j)^2} + b \right) \right\} \quad (3.8)$$

$$P(\mu_j | \mathbf{y}, Z, \phi_j) \propto \exp \left\{ -\frac{\phi_j}{2\mu_j^2(1-\mu_j)^2} \sum_{i \in \{i: Z_{ij}=1\}} \left( \frac{(y_i - \mu_j)^2}{y_i(1-y_i)} \right) \right\}, \quad (3.9)$$

where  $n_j = \sum_{i=1}^n Z_{ij}$  denotes the number of observations drawn from a  $j$ th component of the mixture. Note that  $(\phi_j | \mathbf{y}, Z, \mu_j) \sim \text{Gamma} \left( n_j/2 + a, \sum_{i \in \{i: Z_{ij}=1\}} \frac{(y_i - \mu_j)^2}{2y_i(1-y_i)\mu_j^2(1-\mu_j)^2} + b \right)$ . In addition, the full conditional density of  $\boldsymbol{\omega}$  is

$$P(\boldsymbol{\omega} | \mathbf{y}, Z) \propto \prod_{j=1}^k \omega_j^{v_j+n_j-1}, \quad (3.10)$$

that is the pdf of a Dirichlet distribution, that is,  $(\boldsymbol{\omega} | \mathbf{y}, Z) \sim \text{Dirichlet}(v_1 + n_1, \dots, v_k + n_k)$  where  $v_1, \dots, v_k$  are the parameters of the Dirichlet prior. A description of the whole algorithm to simulate from the joint posterior distribution is given in Appendix A in the end of the thesis. A reversible-jump to estimate the number of components in the mixture is described in the following subsection.

## Reversible-jump

Reversible-jump (RJ) MCMC was introduced by Green (1995) as an extension to MCMC in which the dimension of the model is uncertain. Richardson and Green (1997) extends this method for mixtures of Normal distributions. For the limited data, Bouguila and Elguebaly (2012) develop a procedure to deal with mixtures of beta distributions. However, for mixtures of Simplex distributions, the RJMCMC is not available. In this subsection, we describe a RJMCMC for mixtures of Simplex distributions.

The move in the RJ step, called split-combine moves, allows the increase or reduction of the number of components by one in each step. In each move, the reversible-jump compares two models with different numbers of Simplex components. The split-combine moves form a reversible pair. For this pair, we choose the proposal distribution  $T_{k \rightarrow k^*}$  according to informal considerations in order to obtain a reasonable probability of acceptance. The notation  $T_{k \rightarrow k^*}$  means the proposal transition function for the move from a model with  $k$  Simplex components to a model with  $k^*$  Simplex components. This move is chosen with probability  $p_{k^*|k}$ . Since the parametric space of parameters  $(\boldsymbol{\theta}, \boldsymbol{\omega}, k)$  is different from  $(\boldsymbol{\theta}^*, \boldsymbol{\omega}^*, k^*)$ , the smaller parameter space should be increased. We generate a three-dimensional random vector  $u$  from a  $g(u)$  to



complete the parameter space. Green (1995) shows that the balance condition is determined by the acceptance probability to this move given by  $\alpha((\boldsymbol{\theta}^*, \boldsymbol{\omega}^*, k^*) | (\boldsymbol{\theta}, \boldsymbol{\omega}, k)) = \min\{1, A\}$  where

$$A = \frac{L((\boldsymbol{\theta}^*, \boldsymbol{\omega}^*, k^*) | \mathbf{y}, Z) P((\boldsymbol{\theta}^*, \boldsymbol{\omega}^*) | k^*) P(k^*) p_{k|k^*}}{L((\boldsymbol{\theta}, \boldsymbol{\omega}, k) | \mathbf{y}, Z) P((\boldsymbol{\theta}, \boldsymbol{\omega}) | k) P(k) p_{k^*|k} g(u)} |J|, \quad (3.11)$$

where  $J$  is the Jacobian of the transformation. The probability of the inverse move is given by  $\alpha((\boldsymbol{\theta}, \boldsymbol{\omega}, k) | (\boldsymbol{\theta}^*, \boldsymbol{\omega}^*, k^*)) = \min\{1, A^{-1}\}$ .

The choice between whether to split or combine is made randomly with probability  $b_k$  and  $d_k = 1 - b_k$  respectively, depending on  $k$ . Note that  $d_1 = 0$  and  $b_{k_{max}} = 0$  with  $k_{max}$  representing the maximum value assumed for  $k$ , as seen in the previous subsection. If  $2 < k < k_{max}$  we adopt  $b_k = d_k = 0.5$ .

If the split move is chosen, we select randomly one component  $j_*$  to break into two new components  $(j_1, j_2)$  and create a new state with  $k^* = k + 1$  components. In order to specify the new values of parameters for the two components, Richardson and Green (1997) propose generating a vector  $u = (u_1, u_2, u_3)$  from beta distributions, and set these parameters using a deterministic transformation called proposal transition function. This function must be bijective and provide adequate values of parameters. Then, we generate  $u_1 \sim Beta(2, 2)$ ,  $u_2 \sim Beta(1, 1)$  and  $u_3 \sim Beta(2, 2)$ , and in order to set the parameters we propose the proposal transition function such that

$$\begin{aligned} \omega_{j_1} &= \omega_{j_*} u_1, & \omega_{j_2} &= \omega_{j_*} (1 - u_1), \\ \mu_{j_1} &= \mu_{j_*} - u_2 u_1 (\mu_{j_*} - \mu_{j_*}^2), & \mu_{j_2} &= \mu_{j_*} + u_2 (1 - u_1) (\mu_{j_*} - \mu_{j_*}^2), \\ \phi_{j_1}^{-1} = \sigma_{j_1}^2 &= \sigma_{j_*}^2 u_3 (1 - u_2^2) / u_1, & \phi_{j_2}^{-1} = \sigma_{j_2}^2 &= \sigma_{j_*}^2 (1 - u_3) (1 - u_2^2) / (1 - u_1). \end{aligned} \quad (3.12)$$

All observations previously allocated to  $j_*$  are reallocated doing  $z_i = j_1$  or  $z_i = j_2$  following the same criteria used in the step (2a) of the Algorithm 4.

The combine proposal begins by choosing a pair of components  $(j_1, j_2)$ , where the first is chosen through a discrete Uniform distribution and the second is chosen by making  $j_2 = j_1 + 1$ , the  $k$ th component cannot be chosen in the first place. These two components are merged, reducing  $k$  by 1. The new component is labelled  $j_*$  and contains all observations previously allocated to  $j_1$  and  $j_2$  doing  $z_i = j_*$ . The parameters for the component  $j_*$  are set as

$$\omega_{j_*} = \omega_{j_1} + \omega_{j_2}, \mu_{j_*} = \frac{\mu_{j_1} \omega_{j_2} + \mu_{j_2} \omega_{j_1}}{\omega_{j_*}} \text{ and } \sigma_{j_*}^2 = \frac{\sigma_{j_2}^2 \left(\frac{\omega_{j_2}}{\omega_{j_*}}\right)}{\left(1 - \left(\frac{\mu_{j_2} - \mu_{j_1}}{\mu_{j_*} - \mu_{j_*}^2}\right)^2\right) \left(\frac{\sigma_{j_2}^2 \omega_{j_2}}{\sigma_{j_1}^2 \omega_{j_1} + \sigma_{j_2}^2 \omega_{j_2}}\right)}.$$

This process is reversible, i.e., if we first split one component into two and then combine the components  $j_1$  and  $j_2$ , we can recover the previous state. We can also compute the corresponding values of  $u_i$ 's in

$$\text{the merge move as } u_1 = \frac{\omega_{j_1}}{\omega_{j_*}}, u_2 = \frac{\mu_{j_2} - \mu_{j_1}}{\mu_{j_*} - \mu_{j_*}^2} \text{ and } u_3 = \frac{\omega_{j_1} \sigma_{j_1}^2}{\omega_{j_1} \sigma_{j_1}^2 + \omega_{j_2} \sigma_{j_2}^2}.$$

The acceptance probabilities for split and combine are  $\min\{1, A\}$  and  $\min\{1, A^{-1}\}$  respectively, according to (3.11), with

$$A = \frac{(k+1) \left( \prod_{i \in \{i: Z_{ij_1}=1\}} S(y_i | \mu_{j_1} \sigma_{j_1}^2) \right) \left( \prod_{i \in \{i: Z_{ij_2}=1\}} S(y_i | \mu_{j_2} \sigma_{j_2}^2) \right)}{\left( \prod_{i \in \{i: Z_{ij_*}=1\}} S(y_i | \mu_{j_*} \sigma_{j_*}^2) \right)}$$

$$\begin{aligned} & \times \frac{P(k+1)}{P(k)} \frac{\omega_{j_1}^{v-1+n_1} \omega_{j_2}^{v-1+n_2}}{\omega_{j_*}^{v-1+n_1+n_2}} \frac{P(\sigma_{j_2}^2)P(\sigma_{j_2}^2)}{P(\sigma_{j_*}^2)} \frac{P(\mu_{j_2})P(\mu_{j_2})}{P(\mu_{j_*})} \\ & \times \frac{d_{k+1}}{b_k P_{alloc}(u)} \frac{1}{2} (\sigma_{j_1}^2 + \sigma_{j_2}^2) (\omega_{j_1} + \omega_{j_2}) \left[ 2(\mu_{j_1} + \mu_{j_2}) - (\mu_{j_1} + \mu_{j_2})^2 \right], \end{aligned}$$

where  $d_{k+1}$  is the probability of choosing the merge movement between components  $j_1$  and  $j_2$ ,  $b_k$  is the probability of choosing the split movement of the component  $j_*$ ,  $P_{alloc}$  is the probability of a specific allocation defined as the product of conditional posterior probabilities used to allocate the observations,  $g(u)$  is the joint distribution of  $u = (u_1, u_2, u_3)$  given by the product of densities of beta distributions with parameters  $(2, 2)$ ,  $(k+1)$  is the ratio  $\frac{(k+1)!}{k!}$  from the order statistics densities for the parameters  $(\mu, \sigma^2)$  and the last term of the equation is the Jacobian of the transformations. The second term in (3.13) is the rate of the densities of a prior distributions.

### 3.4 Analysis of simulated data sets

The aim of this section is to investigate the performance of the proposed algorithm to estimate the mixture parameters and the number of components of the mixture model. For this purpose, we implemented the Algorithm 4, shown in Appendix A, in R Development Core Team (2016). The analysis was conducted considering six scenarios for Simplex distribution mixture models, i.e. six different mixtures were considered to generate data. The data was generated independently from the Simplex distribution mixtures, i.e.  $Y \sim SM(\mu, \sigma^2, \omega, k)$  with  $k \in \{2, 3\}$ . The parameters of the six models are shown in the first column of Table 3 and are denoted as  $\mathcal{M}_1, \dots, \mathcal{M}_6$ . For each model, we simulated three data sets. The first had a size of  $n = 1000$  and the other two had a size of  $n < 1000$ , as seen in the second column of Table 3. The value of  $k_{max}$  was fixed to 5 and the hyper-parameters for the Gamma prior were fixed at  $a = 2$  and  $b = 1/2$ . After discarding the first 100000 iterations, we used 100000 iterations with thinning equal to 10 for the inference process.

The posterior relative frequency of  $k$ , shown in Table 3, gives evidence that the reversible-jump MCMC correctly estimated the number of components. In this table, we can see that, even when the component averages are closer (model  $\mathcal{M}_1$ ), the method chooses the correct model in most instances. In addition, Table 4 shows the estimated values of the parameters for the six models. Posterior mean and empirical standard deviation (SD) are shown in this table. It can be observed that the SD decreases as  $n$  increases and the estimated values of the parameters are always close to the true values. Finally, the real histogram and estimated densities are shown in Figure 2 which confirm the adequate performance of the estimation method. The results related to the posterior relative frequency for the number of components obtained from simulated data, shown in Table 3, are in concordance with the results presented in Bouguila and Elguebaly (2012), for mixtures of Beta distributions with unknown number of component.

It can be observed in the simulation process that the convergence speed is improved if the

Table 3 – Parameters used to simulate the data sets and the posterior relative frequency for the number of components obtained from each simulated data set of size  $n$ .

Model	n	Posterior relative frequency				
		k=1	k=2	k=3	k=4	k=5
$\mathcal{M}_1 \begin{cases} \mu = (0.34, 0.72) \\ \sigma^2 = (0.8, 1.5) \\ w = (0.5, 0.5) \end{cases}$	1000	0	<b>0.9874</b>	0.0125	0.0001	0
	500	0	<b>0.9804</b>	0.0194	0.0002	0
	100	0.2790	<b>0.6866</b>	0.0333	0.0010	0.
$\mathcal{M}_2 \begin{cases} \mu = (0.08, 0.40) \\ \sigma^2 = (2, 1) \\ w = (0.65, 0.35) \end{cases}$	1000	0	<b>0.9941</b>	0.0059	0	0
	500	0	<b>0.9852</b>	0.0145	0.0003	0
	100	0.1060	<b>0.8532</b>	0.0389	0.002	0.0001
$\mathcal{M}_3 \begin{cases} \mu = (0.23, 0.58) \\ \sigma^2 = (1.8, 0.8) \\ w = (0.30, 0.70) \end{cases}$	1000	0	<b>0.9876</b>	0.0123	0.0001	0
	500	0	<b>0.9843</b>	0.0154	0.0002	0.0001
	100	0.1648	<b>0.7711</b>	0.0591	0.005	0
$\mathcal{M}_4 \begin{cases} \mu = (0.30, 0.55, 0.80) \\ \sigma^2 = (0.20, 0.10, 0.20) \\ w = (0.20, 0.30, 0.50) \end{cases}$	1000	0.01	0.051	<b>0.9370</b>	0.0024	0
	500	0.0062	0.0523	<b>0.9387</b>	0.0028	0
	300	0.0199	0.3393	<b>0.6316</b>	0.009	0.0003
$\mathcal{M}_5 \begin{cases} \mu = (0.15, 0.47, 0.75) \\ \sigma^2 = (5.0, 0.2, 0.8) \\ w = (0.25, 0.45, 0.30) \end{cases}$	1000	0.0001	0.1140	<b>0.8667</b>	0.0189	0.0003
	500	0.0005	0.1660	<b>0.8157</b>	0.0177	0.0001
	400	0.0023	0.3831	<b>0.5906</b>	0.0240	0
$\mathcal{M}_6 \begin{cases} \mu = (0.10, 0.50, 0.90) \\ \sigma^2 = (6.0, 0.5, 8.0) \\ w = (0.3, 0.50, 0.20) \end{cases}$	1000	0.0047	0.0208	<b>0.9518</b>	0.0224	0.0003
	500	0.0126	0.0583	<b>0.8984</b>	0.0388	0.001
	400	0.0106	0.1409	<b>0.7929</b>	0.0536	0.0020

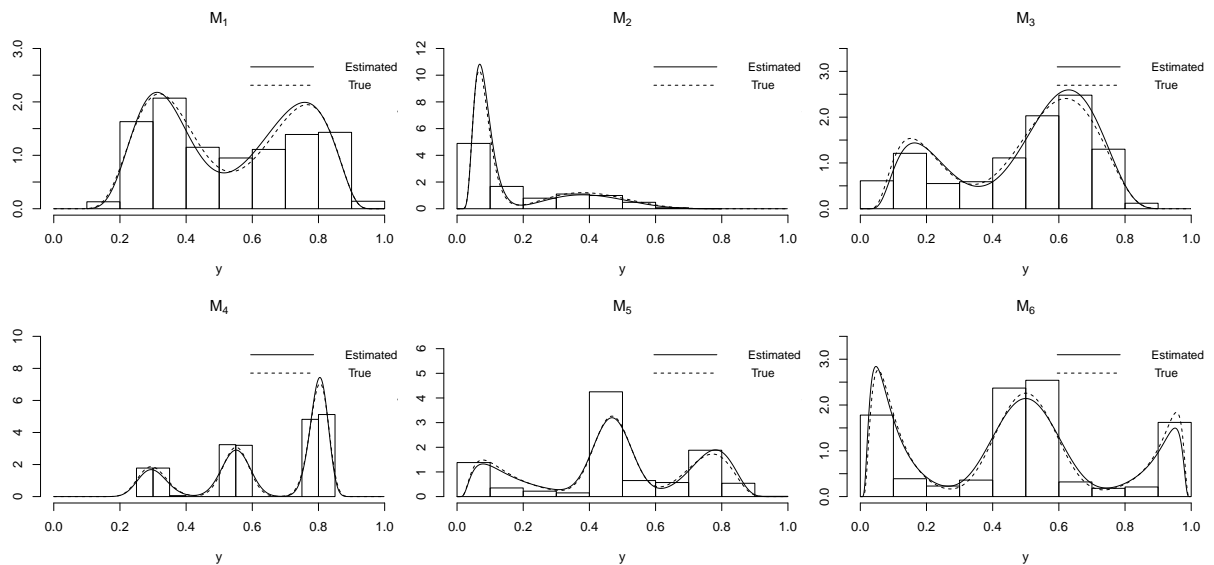


Figure 2 – Histograms and posterior density function.

Table 4 – Posterior mean of the parameters and empirical standard deviation (SD) for simulated data sets considering six models with  $k = 2$  and  $k = 3$  described in Table 3.

Model	n		Estimative to $\mu$	Estimative to $\sigma^2$	Estimative to $\omega$
$\mathcal{M}_1$	1000	Mean	(0.33, 0.71)	(0.69, 1.57)	(0.46, 0.54)
		SD's	(0.005, 0.006)	(0.061, 0.134)	(0.019, 0.016)
	500	Mean	(0.33, 0.72)	(0.77, 1.43)	(0.47, 0.53)
		SD's	(0.010, 0.011)	(0.1225, 0.1990)	(0.031, 0.031)
	100	Mean	(0.35, 0.69)	(0.61, 1.83)	(0.41, 0.59)
		SD's	(0.041,0.047)	(0.556, 0.741)	(0.123,0.123)
$\mathcal{M}_2$	1000	Mean	(0.082,0.40)	(1.99, 0.98)	(0.70, 0.3)
		SD's	(0.001, 0.006)	(0.1178, 0.0915)	(0.015,0.015)
	500	Mean	(0.079, 0.40)	(2.00, 0.89)	(0.68, 0.32)
		SD's	(0.002, 0.008)	(0.3417, 0.1115)	(0.021, 0.021)
	100	Mean	(0.091,0.40)	(2.96, 0.77)	(0.72,0.28)
		SD's	(0.008, 0.026)	(0.811, 0.661)	(0.058,0.058)
$\mathcal{M}_3$	1000	Mean	(0.21,0.61)	(1.86,0.91)	(0.29, 0.71)
		SD's	(0.009, 0.005)	(0.2386, 0.0659)	(0.018, 0.018)
	500	Mean	(0.19,0.60)	(2.03, 1.08)	(0.27, 0.73)
		SD's	(0.012, 0.003)	(0.970, 0.1327)	(0.028, 0.028)
	100	Mean	(0.24, 0.58)	(2.00, 1.11)	(0.23, 0.77)
		SD's	(0.078, 0.028)	(2.085, 0.706)	(0.111, 0.111 )
$\mathcal{M}_4$	1000	Mean	(0.30, 0.55, 0.80)	(0.20, 0.12, 0.18)	(0.18, 0.31, 0.51)
		SD's	(0.007, 0.004, 0.03)	(0.07, 0.08, 0.05)	(0.01, 0.02, 0.02)
	500	Mean	(0.30 0.55 0.80)	(0.23,0.10,0.20)	(0.17 0.31 0.52)
		SD's	(0.009 0.007 0.007)	(0.10, 0.10 0.06)	(0.02 0.02 0.02)
	300	Mean	(0.30, 0.55, 0.80)	(0.27, 0.13, 0.22)	(0.17, 0.32, 0.52)
		SD's	(0.025, 0.016, 0.014)	(0.32, 0.25, 0.13)	(0.034, 0.042, 0.041)
$\mathcal{M}_5$	1000	Mean	(0.16, 0.47, 0.76)	(5.25, 0.21, 0.77)	(0.24, 0.45, 0.31)
		SD's	(0.01, 0.01, 0.01)	(0.7, 0.1, 0.1)	(0.02, 0.02, 0.02)
	500	Mean	(0.14, 0.46, 0.75)	(3.19, 0.26, 0.76)	(0.20 0.47 0.33)
		SD's	(0.02, 0.01, 0.01)	(0.65, 0.26, 0.19)	(0.02, 0.03, 0.03)
	400	Mean	(0.15, 0.48, 0.76)	(4.41, 0.37, 0.83)	(0.20, 0.45, 0.35)
		SD's	(0.031, 0.038, 0.028)	(2.9, 1.8, 0.35)	(0.04, 0.06, 0.07)
$\mathcal{M}_6$	1000	Mean	(0.10, 0.50, 0.89)	(7.34, 0.58, 7.50)	(0.31, 0.51, 0.18)
		SD's	(0.012, 0.009, 0.026)	(0.9, 0.9, 1.9)	(0.02, 0.02, 0.02)
	500	Mean	(0.098, 0.49, 0.88)	(6.05, 0.73, 9.25)	(0.31, 0.51, 0.18)
		SD's	(0.02, 0.01, 0.06)	(6,3,4)	(0.03, 0.04, 0.06)
	400	Mean	(0.097, 0.49, 0.87)	(4.843, 1.00, 10.45)	(0.31, 0.49, 0.20)
		SD's	(0.02, 0.03, 0.08)	(1, 5, 6)	(0.05, 0.05, 0.06)
		(0.078, 0.028)	(2.085, 0.706)	(0.111, 0.111 )	

Table 5 – Relative frequency of  $k$  to the MHDI data set considering alternative SM models.

$k$	1	2	3	4	5
	0.073	0.923	0.005	0.0004	0

Table 6 – Posterior estimates of the parameters and the empirical standard deviation for the MHDI data set.

$\hat{\mu}$		$\hat{\sigma}^2$		$\hat{w}$	
means	SD's	means	SD's	means	SD's
(0.59, 0.73)	(0.004, 0.010)	(0.09, 0.21)	(0.03, 0.076)	(0.69, 0.31)	(0.03, 0.031)

initial value of the number of component is set as  $k^{(0)} = k_{max}$ . The convergence is also affected by the acceptance of the MR step in the *Gibbs sampling* algorithm. A strategy to avoid low rate of acceptance in the MR step is presented in Appendix A. These results also provide evidence that the choice of hyper-parameters discussed in Section 3.3 and the proposal transition function were appropriate.

### 3.5 Analysis of a municipal HDI data set in Brazil

The Human Development Index (HDI) is a summary measure of long-term progress in three basic dimensions of human development that takes into account education, income and longevity indexes. The HDI is the geometric mean of normalized indexes for each of the three dimensions of human development. In this work, we analyze a municipal HDI (MHDI) data set, i.e., the MHDI of the cities (or towns) of São Paulo state and the Northeast region of Brazil. The Northeast was chosen because it is the third largest region of Brazil and the largest in number of states and considered a region with poor distribution of resources, see index in [PNUD, IPEA and FJP. \(2013\)](#). This region includes the states of Alagoas, Bahia, Ceará, Maranhão, Paraíba, Pernambuco, Piauí, Rio Grande do Norte and Sergipe. There are 1,794 cities in the Northeastern region and 645 in São Paulo state leading to a sample of size  $n = 2439$ . The histogram of the data is shown in Figure 3, where the multimodality phenomenon can be observed. This phenomenon is already expected because the MHDI depends on characteristics that can be similar to some cities or towns.

The MHDI data set was analyzed with the SD model where we set  $a = 2$  and  $b = 1/2$  in the Gamma prior distribution for  $\sigma_j^{-2}$ , for  $j = 1, \dots, k$ . In order to reduce prior information, we set  $v_1 = v_2 = \dots = v_k = 1$  for the Dirichlet prior distribution and  $k_{max} = 5$  for prior distribution of  $k$ . Table 5 shows the posterior distribution of parameter  $k$  with high posterior probability to  $k = 2$ . There is evidence for two components in the data. The mean and empirical SD of the parameter estimate are shown in Table 6.

The results of the analysis show that there is strong evidence for two components with similar characteristics. The first component has, in mean, smaller MHDI than the second

2010

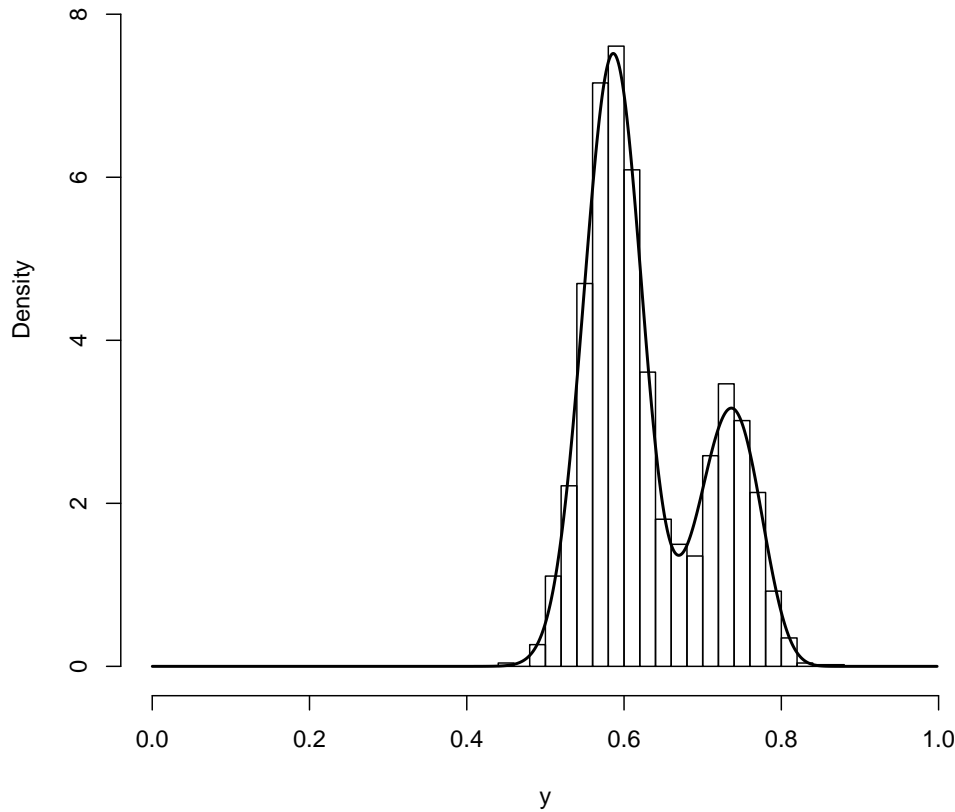


Figure 3 – Real histogram and Estimated density function for the MHDl data set.

component. The component with less MHDl has larger mixing proportions than the second component, as expected. Figure 4(A) shows in black the cities classified as belonging to the first component. It can be observed that there are some cities in the Northeastern region classified in the first component. These cities have better MHDl than those which are classified in the second component, shown in Figure 4(B), cities in black.

### 3.6 Final comments

A flexible model for proportions considering a mixture of Simplex distributions with an unknown number of components was proposed. The main advantage of this model is its flexibility for working with bounded data with multimodalities identified in the components or populations of the data. A Fully Bayesian approach considering MCMC with the reversible-jump algorithm for a mixture of Simplex distributions was developed, as seen in Section 3.3. In this approach, the prior distribution and the hyper-parameters were chosen conveniently, motivated by applying the methodology to real data sets and by the simulation study. In addition, a proposal transition function was adopted in the RJ step that seems adequate in view of the results found

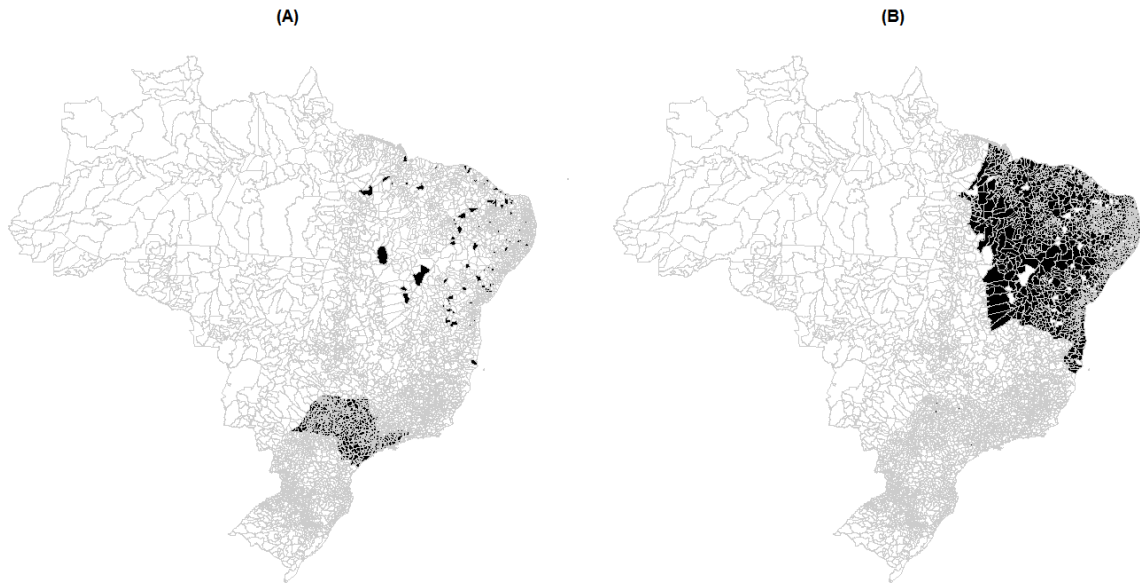


Figure 4 – Classification of HDI of cities in the states São Paulo and Northeastern region of Brazil where the cities classified in the second component are in black in (A) and cities classified in the first component are in black in (B).

in the simulation study. The algorithms have been coded and implemented in R software (R Development Core Team, 2009) and are available from the corresponding author upon request.

An application to generated data sets from a mixture of Simplex distributions with 2 and 3 components were conducted. For these applications, we found that the method provides a good estimation of the number of components, as well as of the other parameters of the model since the estimated values lie close to the real values of the parameters. Results from the simulated data sets also show that the empirical standard deviation decreases as the size of the sample increases. Another application was conducted with a real data set and we found small empirical standard deviations for the sample of the estimates of the parameters of the components and mixing proportions, as can be shown in Table 6, providing evidence that the parameters are well estimated.

The proposed model can be extended to other problems, for instance, for modeling the response variable  $Y$  as a function of covariates. In this case, the mixing proportions may or may not be modelled as functions of a vector of predictor variables, not necessarily having elements in common with the covariates. An example of a link function that can be used for mixing proportions is presented in [McLachlan and Peel \(2004, p. 145\)](#). In addition, as it was observed that the acceptance of RJ decreases when the size of the sample increases, a strategy to avoid persistent rejection of proposed moves in a RJ algorithm can be added to improve the *Gibbs sampling* algorithm ([GREEN; MIRA, 2001](#); [AL-AWADHI; HURN; JENNISON, 2004](#)).

Here, we proposed an flexible model to proportions considering a mixture of Simplex distribution with an unknown number of components. In the next Chapter we investigate the use

of covariates to explain a response variable considering a mixture of Simplex regression models.



---

# MODELING MHDI WITH A FINITE MIXTURE OF SIMPLEX REGRESSION MODELS

---

---

In the Chapters [2](#) and [3](#) we describe mixture models without covariates and apply it in percentage and proportion data, respectively. This chapter presents a mixture of Simplex regression model which is applied to the real data analyzed in the Chapter [3](#).

---

### Abstract

This manuscript deals with an analysis of the municipal human development index as a function of the municipal human poverty index. We propose a regression model where the response follow a mixture of Simplex distribution. Estimation is performed by a Bayesian approach making use of *Gibbs sampling* algorithm.

## 4.1 Introduction

The finite mixtures of distributions are the simplest mixture models which are used for model based on clustering. In this case the model is given by a convex combination of a known number of different distributions where each of them is referred to as component. [Goldfeld and Quandt \(1973\)](#) introduced the finite mixture of linear regression models which have been applied in various fields in the literature as a useful class of finite mixture models. Many efforts have been made to these models and their extensions such as finite mixture of generalized linear models, see [Hurn, Justel and Robert \(2003\)](#). This is because finite mixtures of regression models allow to relax the assumption that the regression coefficients and dispersion parameters are the same for all observations. This work presents the mixture of Simplex regression model as an extension of mixture of linear regression and apply it to MHDI data presented in the previous chapter.

The remainder of the manuscript is organized as follows: In Section 4.2 we present the Mixture of Simplex regression model. The Section 4.3 is dedicate to Models specification and some criteria of comparison. Finally, the results are drawn in Section 4.3.

## 4.2 Model specification

Let  $(x_i, y_i)$  be observations where  $y_i$  represents the observed value of random variable  $Y_i$  taken value in  $(0, 1)$  and  $\mathbf{x}_i = \left( \mathbf{x}_i^{(M)T}, \mathbf{x}_i^{(D)T} \right)^T$  a vector of explanatory variables with dimensions  $q$  and  $d$ , respectively, both with 1 in the first component. In addition, let's assume that  $Y_i$  are independent with density

$$f_i(y_i|\mathbf{x}_i, \boldsymbol{\beta}, \boldsymbol{\delta}, \boldsymbol{\omega}) = \sum_{j=1}^k \omega_j S(y_i|\mathbf{x}_i, \boldsymbol{\beta}_j, \boldsymbol{\delta}_j), \quad (4.1)$$

where

$$S(y_i|\mathbf{x}_i, \boldsymbol{\beta}_j, \boldsymbol{\delta}_j) = \left( 2\pi\sigma_{ij}^2(y_i(1-y_i))^3 \right)^{-1/2} \exp \left\{ - \left( \frac{1}{2\sigma_{ij}^2} \right) \left( \frac{(y_i - \mu_{ij})^2}{y_i(1-y_i)\mu_{ij}^2(1-\mu_{ij})^2} \right) \right\} I_{(0,1)}(y_i)$$

is the  $j$ th component density of the mixture model with  $k$  component given by (4.1). Additionally,  $\mu_{ij}$  and  $\sigma_{ij}$  are the mean and dispersion parameters for the  $i$ th observation in the  $j$ th component,

for  $i = 1, \dots, n$  and  $j = 1, \dots, k$ . In order to formulate the regression model, consider

$$h_1(\mu_{ij}) = \mathbf{x}_i^{(M)T} \boldsymbol{\beta}_j \quad \text{and} \quad h_2(\sigma_{ij}) = \mathbf{x}_i^{(D)T} \boldsymbol{\delta}_j, \quad (4.2)$$

$$(4.3)$$

where  $h_1$  and  $h_2$  are link functions to mean and dispersion and the components of the vectors  $\boldsymbol{\beta} = (\boldsymbol{\beta}_1, \dots, \boldsymbol{\beta}_k)$  and  $\boldsymbol{\delta} = (\boldsymbol{\delta}_1, \dots, \boldsymbol{\delta}_k)$  are  $q$  and  $d$ -dimensional vectors of unknown regression parameters, and  $\mathbf{x}_i^{(D)T}$  is a vector of covariates that can be the same or part of the covariates in  $\mathbf{x}_i^{(M)T}$ .

Since the component density of the mixture model in (4.1) is a pdf of Simplex distribution the model specified relates to a mixture of Simplex regression model with  $k$  components were  $\boldsymbol{\omega} = (\omega_1, \dots, \omega_k)$  are the weights of the mixture model.

### 4.3 Bayesian inference

Let's consider a unobserved random vector  $Z_i = (Z_{i1}, \dots, Z_{ik})$  such that  $Z_{ij} = 1$  if the  $i$ th observation belongs to the  $j$ th mixture component and  $Z_{ij} = 0$  otherwise,  $i = 1, \dots, n$ . The augmented data likelihood to  $(\mathbf{y}, \mathbf{x}, Z)$  can be written as

$$L(\boldsymbol{\beta}, \boldsymbol{\delta}, \boldsymbol{\omega} | \mathbf{y}, \mathbf{x}, Z) = \prod_{i=1}^n \prod_{j=1}^k [\omega_j S(y_i | \mathbf{x}_i, \boldsymbol{\beta}_j, \boldsymbol{\delta}_j)]^{Z_{ij}}, \quad (4.4)$$

where  $Z = (Z_1, \dots, Z_n)$ , and  $\mathbf{x} = (\mathbf{x}_1, \dots, \mathbf{x}_n)$  represent all covariates in the model.

Assuming that  $y_i$  is assigned to component  $j$ ,  $Z_{ij} = 1$ , and taking into account the conditional independence such that  $P(\boldsymbol{\beta}, \boldsymbol{\delta} | Z, \boldsymbol{\omega}) = P(\boldsymbol{\beta}, \boldsymbol{\delta} | Z)$  and  $P(\mathbf{y} | \boldsymbol{\beta}, \boldsymbol{\delta}, \mathbf{x}, Z, \boldsymbol{\omega}) = P(\mathbf{y} | \boldsymbol{\beta}, \boldsymbol{\delta}, \mathbf{x}, Z)$  (BOUGUILA; ELGUEBALY, 2012), the likelihood to  $\boldsymbol{\beta}$  and  $\boldsymbol{\delta}$ , given  $Z$ , can be write as

$$L(\boldsymbol{\beta}, \boldsymbol{\delta} | \mathbf{x}, \mathbf{y}, Z, k) = \prod_{i=1}^n \prod_{j=1}^k [S(y_i | \mathbf{x}_i, \boldsymbol{\beta}_j, \boldsymbol{\delta}_j)]^{Z_{ij}} \quad (4.5)$$

$$= \prod_{j=1}^k \left[ \frac{\exp \left\{ - \sum_{i \in \{i: Z_{ij}=1\}} \left( \frac{(y_i - \mu_{ij})^2}{2\sigma_{ij}^2 y_i (1 - y_i) \mu_{ij}^2 (1 - \mu_{ij})^2} \right) \right\}}{\prod_{i \in \{i: Z_{ij}=1\}} \left( 2\pi \sigma_{ij}^2 (y_i (1 - y_i))^3 \right)^{1/2}} \right]. \quad (4.6)$$

$$(4.7)$$

In order to formulate a Bayesian estimate of the parameters of proposed model, each of the  $k$  factors in (4.7) can be combined with a prior distribution leading to the its full conditional posterior distribution. We specify proper prior distributions as the following

$$\boldsymbol{\beta}_{lj} \sim \text{Normal}(0, 100), \text{ for } l = 0, \dots, q - 1 \text{ and } j = 1, \dots, k,$$

$$\begin{aligned}\boldsymbol{\delta}_{lj} &\sim \text{Normal}(0, 100), \text{ for } l = 0, \dots, d-1 \text{ and } j = 1, \dots, k, \\ \boldsymbol{\omega} &\sim \text{Dirichlet}(\mathbf{v}_1, \dots, \mathbf{v}_k).\end{aligned}\tag{4.8}$$

$$\tag{4.9}$$

These proper priors are a common choice in the literature, see, for example, [Viele and Tong \(2002\)](#). Here, we assume that the priors distribution of parameters are independent to each other. With this information, it is possible to use several Bayesian mechanisms in order to estimate the parameters. Here, we have chosen a Gibbs sampling approach due to because the relative ease to be implemented. The algorithm used in this chapter is similar to that used in the Chapter 3, and summary of this algorithm is shown in Algorithm 2.

---

**Algorithm 2** – Algorithm for simulate samples from the jointly posterior distribution of the parameters of the mixture of L-logistic regression models

---

1. Initialize choosing  $\boldsymbol{\omega}_j^{(t)} = \boldsymbol{\omega}^{(0)}$ ,  $\boldsymbol{\beta}_j^{(t)} = \boldsymbol{\beta}_j^{(0)}$ ,  $\boldsymbol{\delta}_j^{(t)} = \boldsymbol{\delta}_j^{(0)}$  and  $Z_i^{(t)} = Z_i^{(0)}$ , for  $j = 1, \dots, k$  and  $i = 1, \dots, n$ .
2. For  $t = 0, 1, 2, \dots$  repeat

- a) For  $i = 1, \dots, n$  draw  $Z_i^{(t+1)} \sim \text{Multinomial}(1, \pi_{i1}^{(t)}, \dots, \pi_{ik}^{(t)})$ , wherein

$$\pi_{ij}^{(t)} \propto \omega_j f(y_i | \mathbf{x}_i, \boldsymbol{\beta}_j^{(t)}, \boldsymbol{\delta}_j^{(t)})$$

3. Generate  $\boldsymbol{\omega}^{(t+1)}$  from the its full conditional posterior distribution.
  4. For  $j = 1, \dots, k$ 
    - a) Generate  $\boldsymbol{\beta}_j^{(t+1)}$  from the its full conditional posterior distribution using a Metropolis-Hasting step.
    - b) Generate  $\boldsymbol{\delta}_j^{(t+1)}$  from the its full conditional posterior distribution using a Metropolis-Hasting step.
- 

## 4.4 Data analysis

In the Chapter 3 we model the municipal human development index (MHDI), considering some cities of Brazil. As already described, the HDI is a summary measure of long-term progress in three basic dimensions of human development that takes into account education, income and longevity indexes. The HDI is the geometric mean of normalized indexes for each of the three dimensions of human development, as seen in previously Chapter.

The analysis of the MHDI data set as a function of the proportion of poor people per municipality (PPPM) is presented here where the data is modeled by a mixture of two Simplex distribution. The PPPM is defined as the proportion of individuals in each city with household income equal or less than half minimum wage (R\$ 255,00), August 2010 ([ESTATÍSTICA, 2014](#)). We consider MIDH and PPPM of the cities of Northeast region and São Paulo state in Brazil. In this region are the states of Alagoas, Bahia, Ceará, Maranhão, Paraíba, Pernambuco, Piauí, Rio Grande do Norte and Sergipe. There are 1794 cities in the Northeastern region and 645 in São Paulo state leading to a sample of size  $n = 2439$ .

### Sub-models related comparisons

We consider four mixture of Simplex distributions to model the MHDI data.

$$\left\{ \begin{array}{ll} \text{null model } (M_0) : & \text{logit}(\mu_{ij}) = \beta_{0j} \text{ and } \log(\sigma_{ij}) = \delta_{0j}, \\ \text{mean-model } (M_1) : & \text{logit}(\mu_{ij}) = \mathbf{x}_i^T \boldsymbol{\beta}_j \text{ and } \log(\sigma_{ij}) = \delta_{0j}, \\ \text{dispersion model } (M_2) : & \text{logit}(\mu_{ij}) = \beta_{0j} \text{ and } \log(\sigma_{ij}) = \mathbf{x}_i^T \boldsymbol{\delta}_j, \\ \text{full model } (M_3) : & \text{logit}(\mu_{ij}) = \mathbf{x}_{1i}^T \boldsymbol{\beta}_j \text{ and } \log(\sigma_{ij}) = \mathbf{x}_i^T \boldsymbol{\delta}_j, j = 1, 2, \end{array} \right.$$

where  $\boldsymbol{\beta}_j = (\beta_{0j}, \beta_{1j})$ ,  $\boldsymbol{\delta}_j = (\delta_{0j}, \delta_{1j})$  and  $\text{logit}(x) = \log\left(\frac{x}{1-x}\right)$ . Here, we assume that  $\mathbf{x}_i^T$  is a vector of explanatory variables with 1 in the first component and in the second component the  $i$ th value of covariate PPPM, for  $i = 1, \dots, 2439$ .

The sub-models ( $M_0, M_1, M_2$  and  $M_3$ ) were compared by estimating the marginal likelihood and deviance information as expected Akaike information criteria (EAIC), expected Bayesian information criteria (EBIC) and deviance information criteria (DIC) introduced by Spiegelhalter *et al.* (2002). The MCMC output was used to approximate these criteria. The estimate of marginal likelihood was obtained based on the identity

$$m(\mathbf{y}) = \frac{\prod_{i=1}^n f(y_i | \mathbf{x}_i, \boldsymbol{\beta}, \boldsymbol{\delta}, \boldsymbol{\omega}, M) p(\boldsymbol{\beta}, \boldsymbol{\delta}, \boldsymbol{\omega} | M)}{p(\boldsymbol{\beta}, \boldsymbol{\delta}, \boldsymbol{\omega} | \mathbf{y}, \mathbf{x}, M)} \quad (4.10)$$

where  $f(y_i | \mathbf{x}_i, \boldsymbol{\beta}, \boldsymbol{\delta}, \boldsymbol{\omega}, M)$  is the density of  $i$ th observation to current sub-model,  $M$ ,  $p(\boldsymbol{\beta}, \boldsymbol{\delta}, \boldsymbol{\omega} | M)$  is the prior to the parameters and  $p(\boldsymbol{\beta}, \boldsymbol{\delta}, \boldsymbol{\omega} | \mathbf{y}, \mathbf{x}, M)$  is the density of posterior distribution. The approximate  $p(\boldsymbol{\beta}, \boldsymbol{\delta}, \boldsymbol{\omega} | \mathbf{y}, \mathbf{x}, M)$  is obtained as described in Chapter 2 (PAZ; BAZÁN; ELHER, 2015) where is used an approach introduced by Chib and Jeliazkov (2001) to approximate the marginal densities in the mixture models when its do not have know form. The estimate of DIC is obtained as

$$DIC = \bar{D} + P_D \quad (4.11)$$

where  $P_D = \bar{D} - \hat{D}$  with

$$\bar{D} = G^{-1} \sum_{g=1}^G \left( -2 \sum_{i=1}^n \log \left( f_i(y_i | \mathbf{x}_i, \boldsymbol{\beta}^{(g)}, \boldsymbol{\delta}^{(g)}, \boldsymbol{\omega}^{(g)}) \right) \right) \quad (4.12)$$

and

$$\hat{D} = -2 \sum_{i=1}^n \log \left( f_i(y_i | \mathbf{x}_i, \bar{\boldsymbol{\beta}}, \bar{\boldsymbol{\delta}}, \bar{\boldsymbol{\omega}}) \right), \quad (4.13)$$

where the notation  $\bar{\theta}$  main the posterior mean of the parameter  $\theta$  and  $\theta^{(g)}$  represent the  $g$ th estimate of the parameter  $\theta$ , all estimates are obtained from MCMC output. The EAIC and EBIC are estimated by

$$\begin{aligned} EAIC &= \bar{D} + 2 \times P, \\ EBIC &= \bar{D} + 10 \times \log n, \end{aligned} \quad (4.14)$$

where  $P$  is the number of model parameters.

## Results

For the MHDI data set analyzed previously in the Chapter 3, we observe in Figure 5 that still there are some evidence of heterogeneity in the data, even though inclusion of the covariate. Then, we assume that the data can be modeling by a mixture of Simplex distribution with two components, that is, we assume  $k = 2$ .

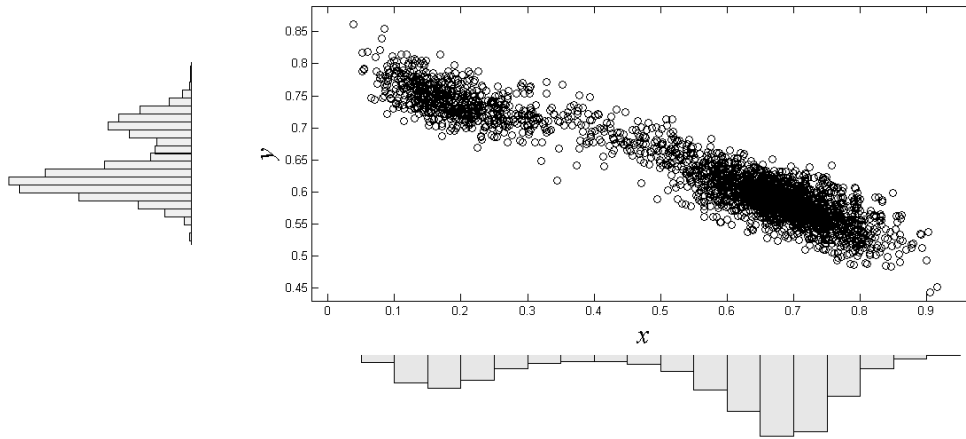


Figure 5 – Scatter plot with marginal histograms of the data.

From specified initial values, we first iterate the sampling procedure to "burn-in" phase of 5000 interactions, already considered thinning of 10 iterations, and the 5000 remaining were used in the analysis. The acceptance rate for update moves, in the Metropolis-Hastings step, is kept around 0.3, this rate was chosen based on the convergence of the algorithm, i.e., we choose the acceptance rate for which the algorithm had convergence.

Table 7 shows the criteria used to compare the four sub-models, including the sub-model without covariate ( $M_0$ ). In this table we can observe that the selected model by the criteria used here, is the sub-model  $M_3$ , which have the covariate in the mean and dispersion. However, taking into account the parsimonious criteria, we can observe that the model  $M_1$  is more plausible than  $M_3$ . Thus, we choose the model  $M_1$  as a final model among those compared. That is, the final model is such that

$$\begin{cases} \text{component 1 : } \text{logit}(\mu_{i1}) = 1.301 - 1.396MP_i \text{ and } \log(\sigma_{i1}) = -3.045, \\ \text{component 2 : } \text{logit}(\mu_{i2}) = 1.598 - 1.827MP_i \text{ and } \log(\sigma_{i2}) = -3.189. \end{cases}$$

Table 7 – Model comparison criteria to the models proposed to MHDH data.

Criteria	Model			
	$M_0$	$M_1$	$M_2$	$M_3$
DIC	-6521.251	-11122.33	-6537.52	-11146.01
EAIC	-6505.701	-11112.94	-6523.045	-11133.90
EBIC	-6447.707	-11066.55	-6465.052	-11075.91
Log marginal likelihood	3627.944	5544.9343	3769.03	5553.879

Table 8 – Number of observations classified across the models and components.

Model	Component	$M_0$		$M_1$	
		1	2	1	2
$M_0$	1	1703	0	1703	0
	2	0	736	622	114
$M_1$	1	1703	622	2325	0
	2	0	114	0	114

For the chosen model and the model without covariate, we build the contingency Table 8 to show the classifications of observations in the model  $M_0$  and  $M_1$  and across these models. In addition, the classification of observations can be seen in the scatter-plot presented in Figure 6. In the contingency table and in the scatter plot, we can observe that the second component, the component with less weight, decrease in number of observations if the covariates are included in the model. This fact is because more information about the poverty is included in the model changing the distribution of the weights. The distribution of the weights, or probability of the mixture, to model  $M_0$  and  $M_1$  can be seen in the Table 9.

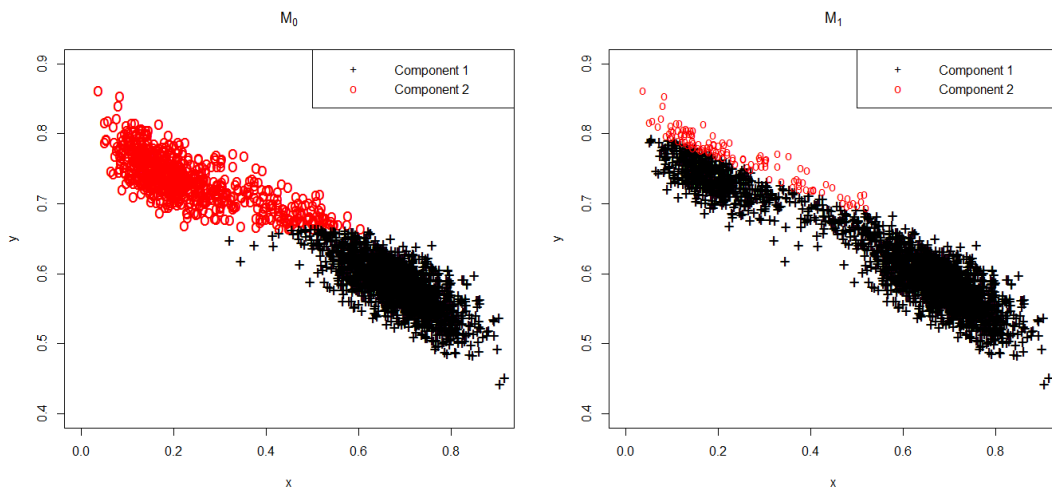


Figure 6 – Scatter plot of the classified data.

Finally, table 9 shows the posterior mean of the parameters of the modes  $M_0$  and  $M_1$  and 95% HPD credible intervals (MARTIN; QUINN; PARK, 2011). We can observe that the zero is

out of range of the HPD interval to  $\beta_{11}$  and  $\beta_{12}$  given evidence that the covariate is positively significant in the model  $M_1$ . The empirical standard deviation is also presented in Table 9 where we can observe that its values are all close to zero given evidence that the all parameters are well estimated.

Table 9 – Posterior mean, credibility intervals and standard empirical deviation of the estimated parameters for sub-model  $M_0$  and  $M_1$ .

Model	Parameter	$\beta_{01}$	$\beta_{02}$	$\beta_{11}$	$\beta_{12}$	$\delta_{01}$	$\delta_{02}$	$\delta_{11}$	$\delta_{12}$	$\omega_1$	$\omega_2$
$M_0$	Mean	0.585	0.732	-	-	-2.336	-1.603	-	-	0.693	0.307
	Lower limit	0.583	0.728	-	-	-2.418	-1.740	-	-	0.672	0.287
	Uper limit	0.587	0.735	-	-	-2.255	-1.471	-	-	0.713	0.328
	SD	0.001	0.002	-	-	0.004	0.014	-	-	0.011	0.011
$M_1$	Mean	1.301	1.598	-1.396	-1.827	-3.045	-3.189	-	-	0.810	0.190
	Lower limit	1.279	1.537	-1.426	-1.924	-3.092	-3.473	-	-	0.714	0.123
	Uper limit	1.318	1.653	-1.360	-1.734	-2.993	-2.948	-	-	0.877	0.286
	SD	0.011	0.031	0.019	0.050	0.027	0.132	-	-	0.043	0.043

## 4.5 Conclusion

In this chapter, we present a mixture of Simplex regression model with known number of components. Our interest lies in the relationship between the MHDI and the municipal human poverty index, where the MHDI is take as a response and the PPPM is take as a covariate. For this purpose, we consider four mixtures of Simplex regression with fixed and known number of components, which are sub-models of mixture of Simplex regression model seen in the Section 4.2. The four sub-models are compared using some criteria of comparisson and the sub-model with the covariate in the mean and dispersion and the sub-model with convariate just in the mean had the best fit. Since the diference between these sub-models is not significant, the sub-model with covariate just in the mean was chouse as the final model for the MHDI data. Based on the results, we found that the municipal human poverty index is significant to explain the MHDI, as expected.

In this analysis, based on the results from the previous chapter and the Figure 5, we assume  $k = 2$  components in the mixture. However, additional analysis have to made to confirm if the real number of components in the mixture is in fact  $k = 2$ . Because, with the inclusion of the covariate, the number of components can increase or decrease. One way to check if the number of components was changed is to proceed the same analysis for mixture with some others number of components. For example, we can consider additionally mixture of Simplex regression models with  $k = 1, 3, 4$ . Finally, we can compare the choosen sub-models for each mixture using the criteria presented in Section 4.4.

In this chapter was presented a mixture of regression model for data in the unit interval. The distribution considered here was already proposed in the literature however it was not analyzed in the mixture context. Here, we apply a mixture of regression model to MHDI data without considering the effect of the stats. In the next chapter we propose a new regression model



which will be considered again in the Chapter 6 in the context of the mixed models. This model is apply to MHDI data without mixture, but the effect of the states is considered.



---

# L-LOGISTIC REGRESSION MODELS: PRIOR SENSITIVITY ANALYSIS, ROBUSTNESS TO OUTLIERS AND APPLICATIONS

---

---

This chapter is addressed to the L-Logistic distribution introduced by [Tadikamalla and Johnson \(1982\)](#). In this work, a convenient parameterization of this distribution is proposed in order to develop regression models.

---

### Abstract

(TADIKAMALLA; JOHNSON, 1982) developed the  $L_B$  distribution to variables with bounded support by considering a transformation of the standard logistic distribution. In this manuscript, a convenient parameterization of this distribution is proposed in order to develop regression models. This distribution, called here L-logistic distribution, provides great flexibility and includes the uniform distribution as a particular case. Several properties of this distribution are studied, and a Bayesian approach is adopted for the parameters estimation. Simulation studies, considering prior sensitivity analysis, and comparison with Beta distribution show the robustness to outliers of the proposed estimation method and the efficiency of the algorithm MCMC adopted. Applications to estimate the vulnerability to poverty and to explain the anxiety are performed. The results to applications show that the L-logistic regression models provides a better fit than the correspondent Beta regression models.

## 5.1 Introduction

Modeling data that are restricted to the interval  $(0,1)$ , as for example the proportion of children vulnerable to poverty or anxiety as function of the stress, are frequently considered by researchers. It is noteworthy that different regression models for modeling this type of data have been proposed in the past few years. For example, Buckley (2003), Ferrari and Cribari-Neto (2004), Paz, Bazán and Milan (2015), Lemonte and Bazán (2016), Gómez-Déniz, Sordo and Calderín-Ojeda (2014) and Bayes, Bazán and Castro (2017), regression models have been proposed among others. However, there are still continuous distributions with bounded support that need further study. This is the case for the L-logistic distribution, which was originally proposed by Tadikamalla and Johnson (1982) through a transformation of the standard logistic distribution. This construction is similar to the  $S_B$  system proposed by Johnson (1949). This distribution was studied by, among others, Tadikamalla and Johnson (1990) and Johnson and Tadikamalla (1991), which proposed the method of moments and the percentile points method to fit this distribution. Additionally, Wang and Rennolls (2005) use a Maximum Likelihood estimation. However, regression models were not studied by considering this distribution.

In this manuscript, we discuss properties of this distribution by considering some parameterizations. Specifically, we present a new parameterization of the L-logistic that consider the median as a parameter. Therefore, we propose a new regression model considering this distribution in the context of quantile regression (QR) models, which were introduced by Koenker and Bassett (1978). Thus, we propose a median regression model so as to represent the relationship between the median (central location) of the response and a set of covariates considering a convenient link function. If the data are highly skewed, since the median is a natural robust measure of the center, the conditional median modeling can be more useful than conditional mean modeling adopted in Beta regression models (BUCKLEY, 2003; FERRARI; CRIBARI-NETO, 2004).

Different from previous studies of L-logistic distribution, we proposed a Bayesian approach using a Markov chain Monte Carlo (MCMC) method for the modeling framework. The issues of model fitting are addressed by means of a hybrid algorithm that combines Metropolis-Hasting with Gibbs sampling algorithms. Initially, we report studies with simulated data sets to investigate the performance of the proposed estimation method concerning influence of the prior distributions and a comparison with Beta distribution for data with outliers. The results obtained from the simulation studies give evidence that the proposed estimation method works well, and that the model proposed is more robust than the Beta model when data with outliers are considered. Also, real applications of social and psychological data are considered to show that there are advantages of our approach in some situations. Firstly, we found that there are evidence that the proportion of children vulnerable to poverty of the municipalities of the state of Alagoas in Brazil, for the 2010 season, is best analyzed considering the L-logistic distribution in comparison with Beta distribution. Additionally, the application of the Beta and L-Logistic regression models to data of anxiety (SMITHSON; VERKUILEN, 2006) give evidence that the L-logistic regression model has a better fit than the correspondent Beta regression model for the modeling this data set.

The rest of the manuscript is organized as follows. In Section 5.2, we present the pdf, the cdf, the quantile function, and also describe the computer generation of the L-Logistic distribution. In Section 5.3, we study some characteristics of the distribution, other parametrizations, some related distributions, moments, and the skewness and kurtosis measures of the L-Logistic distribution. Section 5.4 is dedicated to the Bayesian estimation of the distribution parameters. Some methods for model comparison and diagnosis are also discussed in this section. Section 5.5 presents the results of three simulations studies that examines a prior sensitivity analysis, the estimation of the parameters, and the comparison with Beta distribution to contaminated data. Section 5.6 discusses applications of the proposed distribution, including the proposal of several types of regression models to real data sets. Finals remarks are made in Section 5.7.

## 5.2 The L-Logistic Distribution

We say that the r.v.  $Y$  follows a L-Logistic distribution if its probability density function (pdf) is given by

$$f(y|m,b) = \frac{b(1-m)^b m^b y^{b-1} (1-y)^{b-1}}{[(1-m)^b y^b + m^b (1-y)^b]^2}, \quad 0 < y < 1, 0 < m < 1, b > 0. \quad (5.1)$$

The parameters  $m$  and  $b$  allow L-Logistic distribution, denoted by  $Y \sim LL(m,b)$ , to take on a variety of density shapes (see Figures 7 and 8). Note that when we set  $m = 0.5$  and  $b = 1$  in (5.1), then the pdf of the L-Logistic distribution simply becomes the pdf of the standard Uniform distribution. Here,  $m$  is the median of the distribution, which scales the graph to the left or right on the horizontal axis. On the other hand,  $b$  is a parameter that governs the shape of the

distribution. The L-Logistic density is uni-modal (or “uni-antimodal”), increasing, decreasing, or constant, depending on the values of its parameters. More details on this issue are presented with another parametrization of the L-Logistic model is presented in Section 5.3.

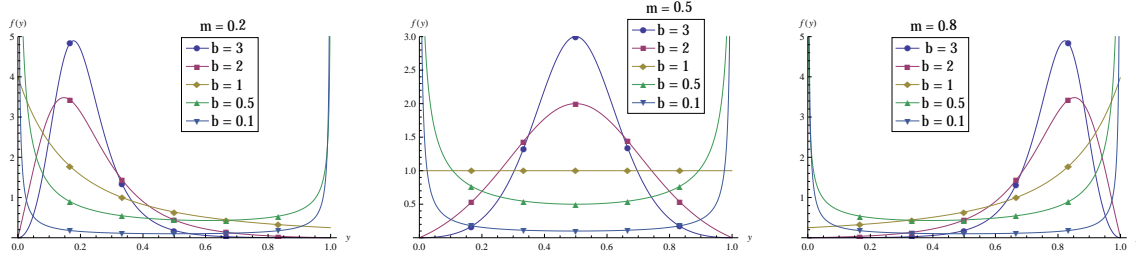


Figure 7 – L-Logistic probability density function for scale parameter  $m = 0.1, 0.5$  and  $0.7$  and some values of parameter  $b$ .

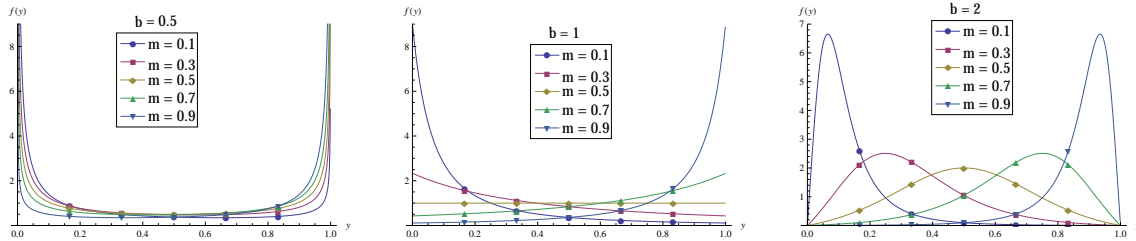


Figure 8 – L-Logistic probability density function for shape parameter  $b = 0.1, 1$  and  $4$  and some values of scale parameter  $m$ .

The cumulative distribution function (cdf) of the L-Logistic distribution is given by

$$F_Y(y|m, b) = \left( 1 + \left( \frac{m(1-y)}{y(1-m)} \right)^b \right)^{-1}, \quad 0 < y < 1, \quad (5.2)$$

which can be readily inverted to yield the quantile function

$$Q_Y(p) = F_Y^{-1}(p) = \frac{mp^{1/b}}{(1-p)^{1/b}(1-m) + p^{1/b}m}, \quad 0 \leq p \leq 1. \quad (5.3)$$

This would readily enable a quantile-based analysis of this model. Note that if  $p = 1 - p = 0.5$ , then  $Q(p) = m$ , which means that the parameter  $m$  is indeed the 50<sup>th</sup> percentile or the median of the L-Logistic distribution.

Equation (5.3) facilitates simple random variate generation. If  $U \sim Uniform(0, 1)$ , then

$$X = Q(U) = \frac{mU^{1/b}}{(1-U)^{1/b}(1-m) + U^{1/b}m} \sim LL(m, b). \quad (5.4)$$

We can also express the inter-quantile range (IQR) as

$$IQR = Q(0.75) - Q(0.25) = \frac{m3^{1/b}}{(1-m) + 3^{1/b}m} - \frac{m}{3^{1/b}(1-m) + m}. \quad (5.5)$$

The IQR has a breakdown point of 50%, and is often preferred over range. When the distribution is symmetric, half IQR equals the median absolute deviation (MAD), and is often used to detect outliers in data.

The functions that provide the probability density function, cumulative density function, quantile function and random generation for the L-Logistic distribution with parameters  $m$  and  $b$  are available in the CRAN of R program, see [Paz and Bazan \(2017\)](#) for details.

## 5.3 Properties of the L-Logistic distribution

This section discusses some properties of the L-Logistic distribution such as related distributions, measures of skewness and kurtosis, mode and the moments. Some pertinent derivations are presented in Appendix B, given in the end of this thesis.

### Related Distributions

The following property shows the relation between L-Logistic distribution and logistic distribution.

**Propriety 5.3.1.** If  $Y \sim LL(m, b)$ , then  $Z = b \log \left( \frac{Y(1-m)}{m(1-Y)} \right)$  has the standard logistic distribution.

Next, we present two alternative parametrizations to the L-Logistic distribution.

**Propriety 5.3.2.** If  $Y \sim LL(m, b)$ , then, with  $m = \frac{e^{-\frac{\delta}{b}}}{1+e^{-\frac{\delta}{b}}}$  ( $\delta > 0$ ), the pdf and cdf of the alternative parametrization of L-Logistic distribution, denoted by  $Y \sim LL(\delta, b)$ , is given by

$$f(y|\delta, b) = \frac{be^{\delta}y^{b-1}(1-y)^{b-1}}{[y^b e^{\delta} + (1-y)^b]^2}, \quad 0 < y < 1, \quad (5.6)$$

and

$$F_Y(y|\delta, b) = \left( 1 + e^{-\delta} \left( \frac{1-y}{y} \right)^b \right)^{-1}, \quad 0 < y < 1, \quad (5.7)$$

respectively, where  $b$  and  $\delta$ ,  $b, \delta > 0$ , are both shape parameters.

The pdf of  $Y \sim LL(\delta, b)$  was introduced by [Tadikamalla and Johnson \(1982\)](#) and [Wang and Rennolls \(2005\)](#), which extended this pdf for  $L_B$  system on any bounded interval by introducing two extra parameters. These authors referred to this distribution as logit-logistic distribution.

**Propriety 5.3.3.** If  $Y \sim LL(m, b)$ , then, with  $\alpha = \frac{1}{1+(\frac{m}{1-m})^b}$  ( $\alpha \in (0, 1)$ ), the pdf and cdf of the alternative parametrization of L-Logistic distribution, denoted by  $Y \sim LL(\alpha, b)$ , is given by

$$f(y|\alpha, b) = \frac{b\alpha(1-\alpha)y^{b-1}(1-y)^{b-1}}{[y^b\alpha + (1-y)^b(1-\alpha)]^2}, \quad 0 < y < 1, \quad (5.8)$$

and

$$F_Y(y|\alpha, b) = \left( 1 + \left( \frac{1-\alpha}{\alpha} \right) \left( \frac{1-y}{y} \right)^b \right)^{-1}, \quad 0 < y < 1, \quad (5.9)$$

where  $b > 0$  and  $\alpha \in (0, 1)$  are both shape parameters.

Note that, although the expressions of the alternative parametrizations of the L-Logistic distribution above have simple expressions for the pdf and cdf, they do not have the median as a natural parameter. In addition, an advantage of the parametrization in (5.1) and (5.2) in comparison with the alternatives parameterizations is that the parameter  $m$  is the median of the distribution and, consequently, the interpretation of this parametrization becomes easier.

**Propriety 5.3.4.** If  $Y \sim LL(\delta, b)$ , then  $Z' = \delta + b \log\left(\frac{Y}{1-Y}\right)$  has the standard logistic distribution.

The reciprocal property was first introduced by [Tadikamalla and Johnson \(1982\)](#), based on an equivalent transformation described by ([JOHNSON; KOTZ; BALAKRISHNAN, 1994](#), pg. 34-49) and first investigated by [Johnson \(1949\)](#), for the case of the standard Normal distribution.

**Propriety 5.3.5.** A extension of the L-Logistic distribution to a variable with support on bounded interval  $(c, d)$  is given by

$$f(y|m, b, c, d) = \frac{(d-c)b(1-m)^b m^b (y-c)^{b-1} (d-y)^{b-1}}{[(1-m)^b (y-c)^b + m^b (d-y)^b]^2}, \quad (5.10)$$

$c < y < d$ , with  $c, d \in \mathbb{R}$ .

## Mode

**Propriety 5.3.6.** For  $b > 1$ , the mode  $y_0$  of the L-Logistic distribution is the solution of the equation

$$\left( \frac{1-m}{m} \right)^b = \left( \frac{1-y_0}{y_0} \right)^b \frac{b+2y_0-1}{b-2y_0+1}. \quad (5.11)$$

Note that, upon taking  $\delta = -b \log\left(\frac{m}{1-m}\right)$ , the mode  $y_0$  can be obtained by solving the equation

$$\delta = \log \left( \left( \frac{1-y_0}{y_0} \right)^b \frac{b+2y_0-1}{b-2y_0+1} \right). \quad (5.12)$$

In addition, from (5.11) and (5.12), if  $y_0 = m = 0.5$ , then  $\delta = 0$  for all values of  $b$ . Thus, we can study the behavior of the mode by studying the function given in (5.12). For this purpose, we take the derivative of the right-hand side of (5.12) with respect to  $y_0$  to obtain the equation

$$\frac{\partial \delta}{\partial y_0} = \frac{b(b^2-1)}{(y_0-1)y_0 \{(b^2-1) + 4y_0 - 4y_0^2\}}. \quad (5.13)$$



(5.11) is negative for  $b > 1$ , the situation where  $\delta$  decreases as  $y_0$  (mode) increases (first derivative test), then the mode lies in  $(0, 1/2)$  if  $\delta > 0$  (or  $m < 1/2$ ) and for  $\delta < 0$  (or  $m > 1/2$ ) the mode is in  $(1/2, 1)$ . If  $b < 1$ , (5.13) is positive whenever  $(b^2 - 1) + 4y_0 - 4y_0^2 > 0$ , that is, whenever  $\frac{1-b}{2} < y < \frac{1+b}{2}$ , the situation where  $\delta$  increases as  $y$  increases. Thus, from (5.12) and (5.13), the minimum of the pdf lies in  $(\frac{1-b}{2}, 1/2)$  for  $\delta < 0$  or  $m > 1/2$ , and in  $(1/2, \frac{1+b}{2})$  for  $\delta > 0$  or  $m < 1/2$ .

### Skewness and Kurtosis Measures

First, we have the following symmetry property.

**Propriety 5.3.7.** The L-Logistic density is symmetric when  $m = 0.5$  for all values of the  $b$ .

For the case when the L-Logistic density is asymmetric, the degree of skewness can be quantified by some measures of skewness. Since the L-Logistic distribution is related to the logistic distribution, the skewness measure introduced by Arnold and Groeneveld (1995), denoted by  $\gamma_M$ , seems to be an appropriate skewness measure. The measure  $\gamma_M$  is based on the mode of distribution and is given by

$$\gamma_M = 1 - 2F(M), \quad (5.14)$$

where  $M$  is the mode of the distribution and  $F(\cdot)$  is the distribution function. The value of  $\gamma_M$  lies in  $(-1, 1)$ , and if  $\gamma_M$  is near 1, it indicates extreme right skewness. On the other hand, if  $\gamma_M$  is near -1, it indicates extreme left skewness.

We also consider another measure of skewness called octile skewness (denoted here by  $\gamma_p$ ), first proposed. This skewness measure is given by

$$\gamma_p = \frac{Q(1-p) + Q(p) - 2m}{Q(1-p) - Q(p)}, \quad (5.15)$$

which is a function of high and low percentiles defined by  $p \in (0, 0.5)$  with  $Q(\cdot)$  being as in (5.3). The maximum value of  $\gamma_p$  is 1, representing extreme right skewness, while the minimum is -1 representing extreme left skewness. This measure is also zero for any symmetric distribution. However, the function in (5.15) depends on the value of  $p$ . We can remove this dependence by integrating over  $p$  or to decide which value of  $p$  is appropriate for use. In Brys, Hubert and Struyf (2003), there is a comparison between several robust skewness measures in which accuracy, robustness, and computational complexity are considered. The most interesting skewness measure of all the measures investigated is octile skewness. Octile skewness takes  $p = 0.125$  in (5.15), that is, it is given by

$$\gamma_{125} = \frac{Q(0.875) + Q(0.125) - 2m}{Q(0.875) - Q(0.125)}. \quad (5.16)$$

For the L-Logistic distribution, we made use of this particular skewness measure instead of removing the dependence over  $p$  through integration.

Moreover, the kurtosis of the L-Logistic distribution can also be derived easily by using the quantiles. The kurtosis measure introduced by Moors (1988) is given by

$$k_Q = \frac{Q(0.875) - Q(0.625) - Q(0.375) + Q(0.125)}{Q(0.75) - Q(0.25)}, \quad (5.17)$$

with  $k_Q \in (0, \infty)$ .

Figure 9 presents the results of the measures of skewness and kurtosis described here for some values of the parameter  $m$  as a function of the shape parameter  $b$ ,  $b > 1$ . In this figure, we can see that the two measures of skewness become close as  $b$  increases, as intuition suggests. Moors (1988) justified the use of the kurtosis measure in (5.17) by the interpretation that the two terms in the numerator of (5.17) are large (small) if relatively little (much) probability mass is concentrated in the neighborhood of  $Q(0.75)$  and  $Q(0.25)$ . This corresponds to large (small) dispersion around (roughly)  $E_Y[Y] \pm Var_Y[Y]$  where  $E_Y[Y]$  and  $Var_Y[Y]$  are the mean and variance of  $Y$ , respectively.

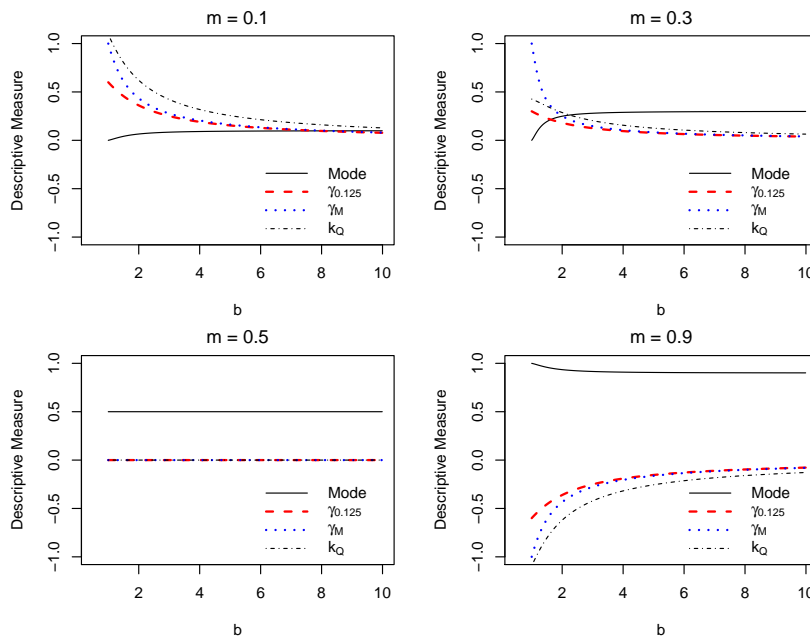


Figure 9 – The mode, skewness ( $\gamma_M$  and  $\gamma_{0.125}$ ) and kurtosis ( $k_Q$ ) of the L-Logistic distribution for some values of the parameters.

## Moments

The following proposition gives an expression for the moments of the L-Logistic distribution.

**Propriety 5.3.8.** If  $Y \sim LL(m, b)$ , then the higher moments of  $Y$  about zero is given by

$$E[Y^t] = \int_0^1 \left[ 1 + \left( \frac{1-v}{v} \right)^{1/b} \left( \frac{1-m}{m} \right) \right]^{-t} dv. \quad (5.18)$$

The integral in (5.18) cannot be expressed in an analytical form. However, we can use numerical integration to evaluate some moments as  $E_Y(Y)$ ,  $E_Y(Y^2)$  and  $Var_Y(Y) = E_Y(Y^2) - E_Y(Y)^2$ . Table 10 shows some values of the first and second moments and the variance of the L-Logistic distribution. In addition, Figure 10 shows the graphs of the mean and variance as functions of the shape parameter  $b$ , for some values of the parameter  $m$ . For this purpose, the integral in (5.18) was evaluated by the Gauss quadrature.

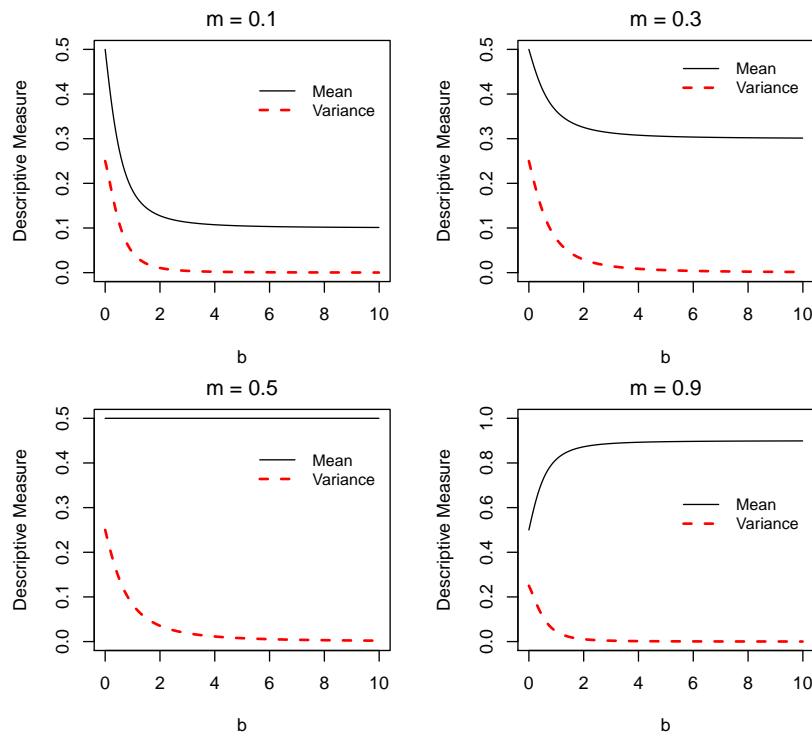


Figure 10 – Descriptive measures of the L-Logistic distributions for some values of the parameters

Table 10 –  $E_Y[Y]$ ,  $E_Y[Y^2]$ , and  $Var_Y(X)$  of the L-Logistic distribution for some values of  $b$  and  $m$ .

$m$	0.2	0.5	0.8	0.2	0.5	0.8
$b$	1	1	1	3	3	3
$E_Y[Y]$	0.283	0.5	0.717	0.216	0.5	0.784
$E_Y[Y^2]$	0.145	0.333	0.579	0.056	0.269	0.625
$Var_Y[Y]$	0.065	0.083	0.065	0.010	0.019	0.01

## 5.4 Bayesian inference

In this section, we describe the Bayesian approach for the estimation of parameters of the L-Logistic model. If we consider a random sample  $\mathbf{Y} = (Y_1, \dots, Y_n)$  from the distribution in (5.2), then the likelihood function is given by

$$L(\theta|\mathbf{y}) = \prod_{i=1}^n \frac{b(1-m)^b m^b y_i^{b-1} (1-y_i)^{b-1}}{[(1-m)^b y_i^b + m^b (1-y_i)^b]^2}. \quad (5.19)$$

## Prior specification

To complete the Bayesian specification of the model, since parameters  $m$  and  $b$  have different behavior, we assume independence between them, and the following structure is then considered:

$$\pi(m, b) = \pi(m)\pi(b), \quad (5.20)$$

where  $\pi(m)$  and  $\pi(b)$  are the prior densities for  $m$  and  $b$ , respectively.

Assuming that  $m \sim \text{Uniform}(0, 1)$ , where  $\text{Uniform}(0, 1)$  represents the Uniform distribution on the unit interval, and prior  $\pi(b)$  for the parameter  $b$ , the joint posterior distribution for  $(m, b)$  is given by

$$\pi(m, b | \mathbf{y}) \propto \prod_{i=1}^n \frac{b(1-m)^b m^b y_i^{b-1} (1-y_i)^{b-1}}{[(1-m)^b y_i^b + m^b (1-y_i)^b]^2} \pi(b). \quad (5.21)$$

The prior  $\pi(b)$  can be, for example, the pdf of the Gamma distribution with parameter vector  $(\varepsilon, \varepsilon)$ ,  $\varepsilon$  being a small value. Some priors for parameter  $b$  are compared and discussed in Section 5.7.

Since the posterior distribution is not available in a closed-form, the Markov Chain Monte Carlo (MCMC) approach (GELMAN *et al.*, 2013, pp. 259 – 349) is used to estimate the model parameters. Initially, we consider the full conditional posterior distributions for the parameters  $(m, b)$  given by

$$\pi(m | b, \mathbf{y}) = K_1^{-1} \frac{(1-m)^{nb} m^{nb}}{\prod_{i=1}^n [(1-m)^b y_i^b + m^b (1-y_i)^b]^2}, \quad (5.22)$$

$$\pi(b | m, \mathbf{y}) = K_2^{-1} \prod_{i=1}^n \left( \frac{b(1-m)^b m^b y_i^{b-1} (1-y_i)^{b-1}}{[(1-m)^b y_i^b + m^b (1-y_i)^b]^2} \right) \pi(b), \quad (5.23)$$

$$(5.24)$$

where  $K_1$  and  $K_2$  are normalizing constants.

Thus, a hybrid algorithm that combines Metropolis-Hastings and Gibbs sampling was implemented in R language (R Development Core Team, 2016) to obtain a sample from the posterior distribution of model parameters  $(m, b)$ . These codes are available upon request.

## Model comparison criteria

In order to compare different models, we made use of some model comparison criteria. Specifically, we consider the Expected Akaike information criterion (EAIC), the expected Bayesian information criterion (EBIC), the deviance information criterion (DIC), and the Watanabe-Akaike information criterion. For a review of these criteria, one may refer to Gelman *et al.* (2013). The EAIC, EBIC and DIC can be estimated as

$$EAIC = \bar{D} + 2 \times p, \quad (5.25)$$

$$EBIC = \bar{D} + 10 \times \log n, \quad (5.26)$$

$$DIC = \bar{D} + p_D, \quad (5.27)$$

where

$$\bar{D} = G^{-1} \sum_{g=1}^G \left( -2 \sum_{i=1}^n \log(f(y_i | b^g, m^g)) \right),$$

$$\hat{D} = -2 \sum_{i=1}^n \log(f(y_i | \bar{m}, \bar{b})),$$

$p_D = \bar{D} - \hat{D}$ , and  $p$  represents the number of model parameters. Here, the notation  $\bar{\theta}$  means the posterior mean of  $\theta$  and  $\theta^{(g)}$  represents the  $g$ th parameters of vector  $\theta$  of a sequence of size  $G$  generated from the posterior distribution by *MCMC* method.

WAIC is a Bayesian approach for estimating the expected pointwise log predictive density (elpdd) for a new dataset, and is given by

$$elpdd = \sum_{i=1}^n E_y[\log(P(\tilde{y}))], \quad (5.28)$$

where the expectancy  $E_y[\cdot]$  and the pdf  $P(\cdot)$  are related to the predictive distribution (pd) induced by the posterior distribution  $P(m, b | \mathbf{y})$  and  $\tilde{y}$  is a future data. To estimate the elppd, we start by computing the pointwise log predictive density (lppd) as

$$\begin{aligned} lppd &= \sum_{i=1}^n \log P(y_i) \\ &= \sum_{i=1}^n \log \int f(y_i | m, b) dP(m, b | \mathbf{y}) \\ &\approx \sum_{i=1}^n \log \left( G^{-1} \sum_{g=1}^G f(y_i | m^g, b^g) \right) = l\hat{p}pd. \end{aligned} \quad (5.29)$$

Thus, we can use a factor of correction for the effective number of parameters to adjust for overfitting. In the literature, there are two factors of correction that can be viewed as approximations to cross-validation (GELMAN *et al.*, 2013, pp. 169-174). The factor of correction used here makes use of the variance in each term of the log predictive density and is used here to obtain the WAIC. This factor is given by

$$\begin{aligned} p_{WAIC} &= \sum_{i=1}^n \text{Var}_{(m,b)|\mathbf{y}}[\log(f(y_i | (m, b)))] \\ &\approx \sum_{i=1}^n \left( \frac{1}{G-1} \sum_{g=1}^G \left[ \log(f(y_i | m^g, b^g)) - G^{-1} \sum_{g=1}^G \log(f(y_i | m^g, b^g)) \right]^2 \right) = \hat{p}_{WAIC_2}. \end{aligned} \quad (5.30)$$

Finally, we can compute the WAIC as an approximation to the elppd as

$$WAIC = l\hat{p}pd - \hat{p}_{WAIC}. \quad (5.31)$$

## Posterior predictive checking

One point that will be of interest is the predictive distribution for a future observation. The density estimation of the posterior predictive distribution is obtained by integrating the unconditional predictive distribution  $P(\tilde{y}|\mathbf{y}) = \int P(\tilde{y}|\boldsymbol{\theta})dP(\boldsymbol{\theta}|\mathbf{y})$ , where  $P(\boldsymbol{\theta}|\mathbf{y})$  is the density of the posterior distribution of the parameters of the assumed model. In practice, we will be interested in simulating draws from the posterior predictive distribution of unknown observables  $\tilde{y}$ . Thus, the predictive distribution can be used to compare the predicted value under the assumed model  $y^{rep}$  with the actual data  $y$ , where  $y^{rep}$  can be thought as an estimate of  $\tilde{y}$ , or as an attempt to replicate the observed data based on the parameters. If the model fits well, then this predicted value should be similar to the observed data.

Using MCMC techniques, we could simulate values of the posterior predictive distribution by generating  $y^{rep}$  from the distribution assumed by its model structure with the parameters generated from the posterior distribution. Let us consider  $y_1, \dots, y_n$  as observations generated independently from the L-Logistic distribution. Then, we can generate  $y^{rep}$  from the L-Logistic distribution with appropriate parameters, that is,

$$y_i^{rep,(s)} \sim LL(m^s, b^s),$$

$s = 1, \dots, S$ , where  $(m^1, b^1), \dots, (m^S, b^S)$  is a sample generated from the posterior distribution. After generating  $(y_1^{rep,(s)}, \dots, y_n^{rep,(s)})$ , we order the sample to achieve  $(y_{(1)}^{rep,(s)}, \dots, y_{(n)}^{rep,(s)})$ , the ordered generated value. We can then compare the distribution of the ordered generated values  $y_{(i)}^{rep,(s)}$  with the ordered observed values  $y_{(i)}$ . Finally, error bar can be constructed or posterior predictive values can be obtained by making use of the discrepancy measure, allowing for an evaluation of the model fit. Details about the predictive model checking are discussed by [Gelman \*et al.\* \(2013\)](#).

## 5.5 Simulation studies

This section presents three simulation studies, one that examines a prior sensitivity analysis, another investigates the recovery of the parameters of the model by the proposed estimation method and, finally, we present a comparison between our approach and a existing approaches to model data with outliers. For this purpose, the Bayesian method is applied on simulated data sets from the L-Logistic distribution, considering different scenarios. For the estimation of parameters, we generated 20,000 samples from the posterior distribution given in (5.21), then the first 10,000 samples were discarded and sequences of 10 observations were eliminated, and finally the resulting samples of size 1,000 were used for inference.

### **Prior sensitivity analysis**

Prior sensitivity analysis plays an important role in applied Bayesian analyses. This is especially true for Bayesian models used for new distribution, where the interpretability of the corresponding parameters becomes important. In this section, we consider a simulation study to evaluate the sensitivity of different choices of prior distributions for parameter  $b$  since this is different from parameter  $m$ , which is clearer in its interpretation. Specifically, we assume prior independence between parameters  $b$  and  $m$ , considering a unit Uniform distribution for parameter  $m$ .

The models estimated with different prior distributions were compared by using WAIC, EAIC, EBIC and DIC. We considered five different prior distributions for  $b$ , considering simulated data sets from the L-Logistic distribution for some pairs of parameters  $m$  and  $b$ . The values for  $m$  and  $b$  used are as follows:  $b \in \{0.5, 1, 5\}$  and  $m \in \{0.2, 0.5, 0.9\}$ , leading to nine scenarios or pairs of parameters, corresponding to nine models simulated. We simulated samples of size  $n = 100$ ,  $y_1, \dots, y_n$ , from the L-Logistic distribution upon considering these pairs of parameters, then nine distinct simulated datasets were considered for the analysis.

Based on the works of [Figueroa-Zúñiga, Arellano-Valle and Ferrari \(2013\)](#), we consider for the parameter  $b$  three relatively non-informative and two informative prior distributions. The non-informative prior distributions are the Gamma distribution with parameter vector  $(0.001, 0.001)$  ( $b \sim \text{Gamma}(0.001, 0.001)$ ), denoted by prior A, the Uniform distribution with parameter vector  $(1, 100)$  for  $U$  ( $U \sim \text{Uniform}(0, 100)$ ) with  $b = U^2$ , denoted by prior B, and the central Student t distribution with parameters vector  $(10, 0, 2)$  ( $L \sim \text{St}(0, 100)$ ) for  $L$  with  $\log(b) = L$ , denoted by prior C. The prior B is chosen because it is less informative than the usual Gamma with parameter vector  $(\varepsilon, \varepsilon)$ . For the informative prior distributions, we consider  $b \sim \text{Gamma}(2.5, 1)$ , denoted by prior D, and  $b \sim \text{Gamma}(50, 1)$ , denoted by prior E. Note that prior E provides incorrect information about parameter  $b$ , while prior D provides almost correct information. For all cases, the prior distribution for parameter  $m$  is the Uniform distribution with parameters 0 and 1, that is,  $m \sim \text{Uniform}(0, 1)$ . For all simulated datasets in the nine scenarios, we found that with prior E the estimated model achieves the worst fit among all models fitted with the other priors. However, for the models using all other prior distributions, the values of WAIC, EAIC, EBIC, and DIC are quite close, showing no significant difference, giving evidence that the estimated models provide almost the same quality of fit for the analyzed samples. Thus, for these cases, the posterior distribution does not seem to be sensitive with respect to the specification of these prior distributions. The values of WAIC, EAIC, EBIC, and DIC for the fitted models, considering these different prior distributions, was shown in Table T of the Appendix B.

For a more detailed analysis, additionally, we chose two non-informative priors, A and C, and the worst informative prior E to present HPD intervals and point estimates. Prior A was chosen for this second analysis because it is simplest among the non-informative priors considered before, while prior C is chosen because it presents lower values for EAIC, EBIC and

DIC than prior B in most of the studied cases. Table 11 reports the posterior mean and the 95% HPD interval (obtained by the package of [Martin, Quinn and Park \(2011\)](#)), considering these priors. We observe that when prior E is used, the HPD interval does not contain the true value of  $b$ . On the other hand, the non-informative A and C priors provide intervals containing the true value of the parameters for all cases analyzed. However, prior A provides the estimated values for the parameter  $b$  (posterior mean) closer to the true value than prior C in most cases.

Considering the results discussed before, we conclude that prior A is the best choice among the priors analyzed for the parameter  $b$  of L-Logistic distribution, for these data sets. Moreover, Gamma distribution is commonly used in the literature for shape parameters (specifically, for precision parameters) and, based on the analysis presented here, the posterior distribution seems to be not sensitive with respect to the specification of this prior distribution.

Table 11 – Posterior mean with 95% HPD interval, prior distributions for parameter  $b$  and true values of the parameters of L-Logistic distribution used to simulate the data sets.

Real Value	Prior A		Prior C		Prior E	
	Mean	HPD (95%)	Mean	HPD (95%)	Mean	HPD (95%)
$m = 0.20$	0.24	(0.14, 0.35)	0.24	(0.14, 0.35)	0.24	(0.16, 0.33)
$b = 0.50$	0.58	(0.49, 0.67)	0.59	(0.50, 0.68)	0.78	(0.68, 0.90)
$m = 0.20$	0.21	(0.17, 0.26)	0.21	(0.17, 0.26)	0.22	(0.18, 0.26)
$b = 1.00$	1.15	(0.98, 1.34)	1.17	(0.98, 1.35)	1.55	(1.34, 1.77)
$m = 0.20$	0.20	(0.19, 0.21)	0.20	(0.19, 0.21)	0.20	(0.19, 0.21)
$b = 5.00$	5.77	(4.89, 6.76)	5.77	(4.83, 6.77)	7.51	(6.49, 8.55)
$m = 0.50$	0.54	(0.40, 0.68)	0.54	(0.39, 0.68)	0.54	(0.43, 0.66)
$b = 0.50$	0.58	(0.48, 0.67)	0.59	(0.49, 0.68)	0.78	(0.67, 0.88)
$m = 0.50$	0.52	(0.44, 0.58)	0.52	(0.45, 0.59)	0.52	(0.47, 0.58)
$b = 1.00$	1.16	(0.97, 1.33)	1.17	(0.98, 1.36)	1.55	(1.33, 1.76)
$m = 0.50$	0.50	(0.49, 0.52)	0.50	(0.49, 0.52)	0.50	(0.49, 0.52)
$b = 5.00$	5.78	(4.79, 6.77)	5.79	(4.83, 6.68)	7.52	(6.51, 8.68)
$m = 0.90$	0.90	(0.85, 0.95)	0.90	(0.86, 0.95)	0.91	(0.87, 0.94)
$b = 0.50$	0.58	(0.49, 0.68)	0.59	(0.49, 0.69)	0.78	(0.68, 0.89)
$m = 0.90$	0.90	(0.88, 0.93)	0.90	(0.88, 0.93)	0.91	(0.89, 0.93)
$b = 1.00$	1.15	(0.98, 1.34)	1.17	(0.98, 1.36)	1.55	(1.34, 1.77)
$m = 0.90$	0.90	(0.90, 0.91)	0.90	(0.90, 0.91)	0.90	(0.90, 0.91)
$b = 5.00$	5.78	(4.96, 6.78)	5.79	(4.80, 6.72)	7.51	(6.60, 8.76)

### Parameter recovery

Here, we present a study of parameter recovery for the parameters of the L-Logistic distribution using prior A for the shape parameter  $b$ , and the unit Uniform prior for the parameter  $m$ . This evaluation was made based on the  $\sqrt{MSE}$  and bias, for simulated data sets from this distribution. The mean and variance of an estimator  $\hat{\theta}$  can be computed by Monte Carlo simulations using the approximations

$$E_{\hat{\theta}}[\hat{\theta}] \approx G^{-1} \sum_{s=1}^G \hat{\theta}^s, \tag{5.32}$$

$$Var_{\hat{\theta}}[\hat{\theta}] \approx G^{-1} \sum_{s=1}^G (\hat{\theta}^s - E_{\hat{\theta}}[\hat{\theta}])^2, \tag{5.33}$$



where  $\hat{\theta}^1, \dots, \hat{\theta}^G$  are obtained from  $G$  different simulated samples. Thus, the MSE of  $\hat{\theta}$  is the function of  $\theta$  defined by

$$E_{\hat{\theta}} [(\hat{\theta} - \theta)^2] = \text{Var}_{\hat{\theta}}[\hat{\theta}] + (E_{\hat{\theta}}[\hat{\theta}] - \theta)^2 \approx G^{-1} \sum_{s=1}^G (\hat{\theta}^s - \theta)^2, \quad (5.34)$$

where  $E[\hat{\theta}] - \theta$  is the bias of  $\hat{\theta}$ . Of course, a good estimator should produce mean, standard deviation (square root of variance), and bias close to zero.

For the analysis presented here, we generated samples of size  $n = 50$ ,  $n = 100$ , and  $n = 500$ . The values of the parameters were set as  $m \in \{0.2, 0.5, 0.9\}$  and  $b \in \{0.3, 0.5, 1, 2, 4\}$ . For these data sets, we estimated the parameters of the L-Logistic model by using the Bayesian method under the assumption of independent priors for  $m$  and  $b$ . Bayes estimator used here is the mean of the posterior distribution (estimator with respect to squared error loss function).

Table 12 shows the values of the  $\sqrt{MSE}$  and bias from the simulated data sets. The estimates for these quantities were obtained from  $G = 1,000$  Monte Carlo replications. We can see that the  $\sqrt{MSE}$  and bias are close to zero even when the sample size is  $n = 50$ . For these samples, the estimator performs well as both  $\sqrt{MSE}$  and bias are very small, for all the analyzed cases. Therefore, we can conclude that the proposed estimation method for the parameters of the L-Logistic distribution works well.

### **Robustness to outliers in L-Logistic**

Now, a simulation study is presented to investigate the robustness to outliers in the L-Logistic distribution, i.e we study the relative performance of the procedure for estimating the Beta and L-Logistic models, considering data generated from Beta and contaminated with outliers.

The contaminated Beta data were generated following Bayes, Bazán and García (2012) in two steps. First, the datasets are generated from a Beta distribution with location parameter  $\mu = 0.2$ , considering two values for the dispersion parameter,  $\phi = 10, 30$ , and three sample sizes,  $n = 50, 100, 200$ . Second, these data were contaminated with outliers generated from a Uniform distribution with parameters 0.999 and 1. The proportions of outliers considered were 0.02, 0.05 and 0.08 for each dataset, i.e.,  $r = 2\%, 5\%, 8\%$  of the data in each dataset were randomly replaced by outliers. This gave  $r \times n/100$  total outliers in each dataset containing  $n$  values. The combination of values of  $\phi$ ,  $n$  and  $r$  provides  $2 \times 3 \times 3 = 18$  scenarios to be analyzed.

In order to compare the fit of Beta and L-Logistic models to the each contaminated data, WAIC, EAIC, EBIC and DIC were obtained for Beta and L-Logistic models for 100 replications in each scenario. Thus, the percentage of cases was computed in which the L-Logistic model achieved a lower value for WAIC, EAIC, EBIC and DIC than the Beta model. The results are presented in Table 13 where we can see no significant difference between the two analyzed models when the DIC is used to select the model. However, the L-Logistic model performed

Table 12 – Bias and root mean square error ( $\sqrt{MSE}$ ) of the Bayesian estimator of the parameters  $m$  and  $b$ .

Real value			m=0.2		m=0.5		m=0.9	
n	parameter		Bias	$\sqrt{MSE}$	Bias	$\sqrt{MSE}$	Bias	$\sqrt{MSE}$
b=0.3	50	m	-1.5e-02	1.5e-02	-1.7e-01	1.7e-01	-2.2e-01	2.2e-01
		b	1.9e-02	1.9e-02	1.9e-02	1.9e-02	1.8e-02	1.8e-02
	100	m	-5.3e-02	5.3e-02	-2.4e-01	2.4e-01	-2.8e-01	2.8e-01
		b	-7.0e-03	7.0e-03	-7.1e-03	7.1e-03	-5.6e-03	5.6e-03
	500	m	-1.8e-02	1.8e-02	-9.8e-02	9.8e-02	-1.0e-01	1.0e-01
		b	-1.2e-02	1.2e-02	-1.1e-02	1.1e-02	-1.1e-02	1.1e-02
b=0.5	50	m	-2.4e-02	2.4e-02	-1.3e-01	1.3e-01	-1.3e-01	1.3e-01
		b	2.8e-02	2.8e-02	3.2e-02	3.2e-02	3.2e-02	3.2e-02
	100	m	-4.3e-02	4.3e-02	-1.7e-01	1.7e-01	-1.6e-01	1.6e-01
		b	-8.8e-03	8.8e-03	-9.5e-03	9.5e-03	-8.1e-03	8.1e-03
	500	m	-1.6e-02	1.6e-02	-6.1e-02	6.1e-02	-5.5e-02	5.5e-02
		b	-1.9e-02	1.9e-02	-1.8e-02	1.8e-02	-1.9e-02	1.9e-02
b=1	50	m	-1.9e-02	1.9e-02	-7.1e-02	7.1e-02	-6.0e-02	6.0e-02
		b	5.7e-02	5.7e-02	5.6e-02	5.6e-02	5.2e-02	5.2e-02
	100	m	-2.7e-02	2.7e-02	-9.3e-02	9.3e-02	-7.3e-02	7.3e-02
		b	-1.7e-02	1.7e-02	-1.6e-02	1.6e-02	-2.2e-02	2.2e-02
	500	m	-9.9e-03	9.9e-03	-3.3e-02	3.3e-02	-2.4e-02	2.4e-02
		b	-3.4e-02	3.4e-02	-3.5e-02	3.5e-02	-3.5e-02	3.5e-02
b=2	50	m	-1.2e-02	1.2e-02	-3.7e-02	3.7e-02	-2.7e-02	2.7e-02
		b	1.2e-01	1.2e-01	1.2e-01	1.2e-01	1.3e-01	1.3e-01
	100	m	-1.5e-02	1.5e-02	-4.9e-02	4.9e-02	-3.3e-02	3.3e-02
		b	-3.4e-02	3.4e-02	-4.3e-02	4.3e-02	-3.4e-02	3.4e-02
	500	m	-6.5e-03	6.5e-03	-1.7e-02	1.7e-02	-1.1e-02	1.1e-02
		b	-7.2e-02	7.2e-02	-6.9e-02	6.9e-02	-7.2e-02	7.2e-02
b=4	50	m	-7.2e-03	7.2e-03	-1.8e-02	1.8e-02	-1.2e-02	1.2e-02
		b	2.6e-01	2.6e-01	2.4e-01	2.4e-01	2.6e-01	2.6e-01
	100	m	-8.7e-03	8.7e-03	-2.4e-02	2.4e-02	-1.6e-02	1.6e-02
		b	-7.0e-02	7.0e-02	-6.9e-02	6.9e-02	-6.2e-02	6.2e-02
	500	m	-2.7e-03	2.7e-03	-7.6e-03	7.6e-03	-4.9e-03	4.9e-03
		b	-1.6e-01	1.6e-01	-1.4e-01	1.4e-01	-1.4e-01	1.4e-01

better than Beta models in all analyzed cases by considering WAIC, EAIC and EBIC. In this table, we also present the bias and MSE for the estimators of  $m$  and  $\mu$  obtained by replication in each scenario, considering 0.2 as the real value for the parameters  $m$  and  $\mu$ . The bias and MSE are always smaller for the  $m$  than the  $\mu$  estimator showing that for any scenario with outliers, there was an improvement in the accuracy (bias and MSE decrease) for the estimation of the model's parameters when using an L-Logistic rather than the Beta model for a contaminated dataset. In order to illustrate the results, the estimated densities for the scenario in which  $n = 100$ ,  $r = 5\%$  and  $\phi = 10$  is shown in Figure 11 where the L-Logistic model seems to fit the data better than the Beta model. From these results, we conclude that the L-logistic distribution is more robust than Beta distribution for the cases analyzed here given evidence that the L-Logistic is more robust than Beta for the modeling these kind of the data.

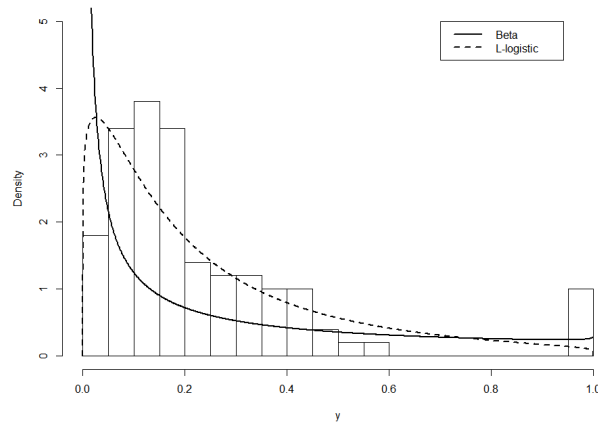


Figure 11 – Estimated densities for Beta and L-Logistic models for de scenarios with  $n = 100$ ,  $\phi = 10$  and  $r = 5\%$ .

Table 13 – Comparison of Bias, MSE and percentage of selection of the model L-Logistic versus Beta considering WAIC, EAIC, EBIC and DIC for different scenarios of contaminated Beta data (two values of  $\phi$  3% of outliers and three sample sizes) by considering 100 dataset replications in each scenario.

$\phi$	% of outlier	n	Beta		L-Logistic		Comparison			
			Bias	$\sqrt{MSE}$	Bias	$\sqrt{MSE}$	WAIC	EAIC	EBIC	DIC
10	2	50	8.0e-02	8.2e-02	-1.8e-02	2.5e-02	1.0e+00	1.0e+00	1.0e+00	4.6e+00
		100	7.9e-02	8.0e-02	-1.8e-02	2.3e-02	1.0e+00	1.0e+00	1.0e+00	4.7e+00
		200	7.8e-02	7.8e-02	-1.9e-02	2.1e-02	1.0e+00	1.0e+00	1.0e+00	4.4e+00
	5	50	1.3e-01	1.3e-01	-1.5e-02	2.3e-02	1.0e+00	1.0e+00	1.0e+00	5.2e+00
		100	1.5e-01	1.5e-01	-9.1e-03	1.6e-02	1.0e+00	1.0e+00	1.0e+00	4.4e+00
		200	1.5e-01	1.5e-01	-1.2e-02	1.4e-02	1.0e+00	1.0e+00	1.0e+00	5.4e+00
8	50	2.1e-01	2.1e-01	1.4e-02	2.3e-02	1.0e+00	1.0e+00	1.0e+00	3.9e+00	
	100	2.1e-01	2.1e-01	1.2e-02	2.0e-02	1.0e+00	1.0e+00	1.0e+00	5.0e+00	
	200	2.0e-01	2.0e-01	5.5e-03	1.1e-02	1.0e+00	1.0e+00	1.0e+00	5.4e+00	
30	2	50	8.3e-02	8.4e-02	-5.0e-03	1.2e-02	1.0e+00	1.0e+00	1.0e+00	5.3e+00
		100	8.2e-02	8.2e-02	-5.2e-03	9.0e-03	1.0e+00	1.0e+00	1.0e+00	4.2e+00
		200	8.0e-02	8.1e-02	-5.9e-03	8.0e-03	1.0e+00	1.0e+00	1.0e+00	5.3e+00
	5	50	1.4e-01	1.4e-01	2.6e-03	1.2e-02	1.0e+00	1.0e+00	1.0e+00	4.8e+00
		100	1.6e-01	1.6e-01	5.4e-03	8.9e-03	1.0e+00	1.0e+00	1.0e+00	5.0e+00
		200	1.6e-01	1.6e-01	3.1e-03	6.1e-03	1.0e+00	1.0e+00	1.0e+00	5.5e+00
	8	50	2.1e-01	2.2e-01	2.4e-02	2.6e-02	1.0e+00	1.0e+00	1.0e+00	4.6e+00
		100	2.1e-01	2.1e-01	1.8e-02	2.1e-02	1.0e+00	1.0e+00	1.0e+00	4.8e+00
		200	2.1e-01	2.1e-01	1.8e-02	1.9e-02	1.0e+00	1.0e+00	1.0e+00	4.3e+00

## 5.6 Applications to a real data set

In order to illustrate the advantages of the use of the L-Logistic distribution in comparison to Beta distribution, in sub-section 6.1 we estimate the distribution of the vulnerability to poverty. Later in sub-section 6.2, we consider initially different regression models to explain the anxiety as function of stress.

### **Estimating the distribution of the vulnerability to poverty**

In this section, we consider a real dataset, which contains the proportion of children (0-14 years olds) vulnerable to poverty. The data came from the municipalities of the state of Alagoas in Brazil, and was collected in 2010. The state of Alagoas is located in the eastern part of the Northeastern Region of Brazil and is made up of 102 municipalities. This state is one of the poorest states of Brazil and its HDI (Human Development Index) is the country's worst, based on information available in [PNUD, IPEA and FJP. \(2013\)](#). Thus, we are interested in modeling the proportion of children vulnerable to poverty (PPOBC). Here, a child is considered vulnerable to poverty if the per capita household income is at most BRL 255, in 2010. The PPOBC data set comprises 102 observations and is modeled here using the L-Logistic distribution and the Beta distribution that is often used to model data when a distribution over some finite interval is needed; see [Gupta and Nadarajah \(2004\)](#). Here, we use the re-parametrized Beta distribution discussed by [Ferrari and Cribari-Neto \(2004\)](#) in the context of regression analysis.

The Bayesian methodology was used to estimate the parameters of both models. For the L-Logistic distribution with parameters  $m$  and  $b$ , we considered prior A discussed early in Section 5.5. Since the Beta distribution have parameters  $0 < \mu < 1$  and  $\phi > 0$ , we considered the same prior A for this model as well.

Table 14 – Estimates and 95% HPD intervals for the parameters of the L-Logistic and Beta models, and statistics for model comparison.

Model	Parameter	Criteria				
		WAIC	EAIC	EBIC	DIC	
L-Logistic	$m$	0.86(0.85, 0.87)	155.1322	-304.2996	-299.0496	-306.3422
	$b$	4.04(3.42, 4.72)				
Beta	$\mu$	0.85(0.84, 0.86)	150.8993	-295.3312	-290.0813	-297.3437
	$\phi$	37.81(27.55, 47.83)				

The final result on the estimation is presented in Table 14. This table also shows the values of statistics for model comparison in order to evaluate the ability of L-Logistic and Beta models to fit the data. According to this table, it is clear that the L-Logistic model is better for modeling the PPOBC data than the Beta model. In addition, Figure 12 shows two graphs with the mean values and errors bars with 95% credibility intervals plotted against the corresponding observed value of the data. The errors bars were constructed from 1000 samples (ordered, and of size 102) generated from the L-Logistic and Beta distributions, respectively, for each graph, with the estimated parameters. In the case of the L-Logistic model, the bars crossed by the diagonal line  $y = x$  indicate that the model is quite suitable for the data. On the other hand, in the case of the Beta model, we observe high deviations between the predicted and observed data, mainly in the tail of the distribution. In this case, an observation is flagged as an outlier, since the corresponding posterior interval does not contain the observed value. Thus, Figure 12 provides evidence that the Beta model is unsuitable for these data. Finally, the estimated and the observed

histogram of the PPOBC data are presented in Figure 5.6, which confirms that the L-Logistic model provides a better fit for these data than the Beta model.

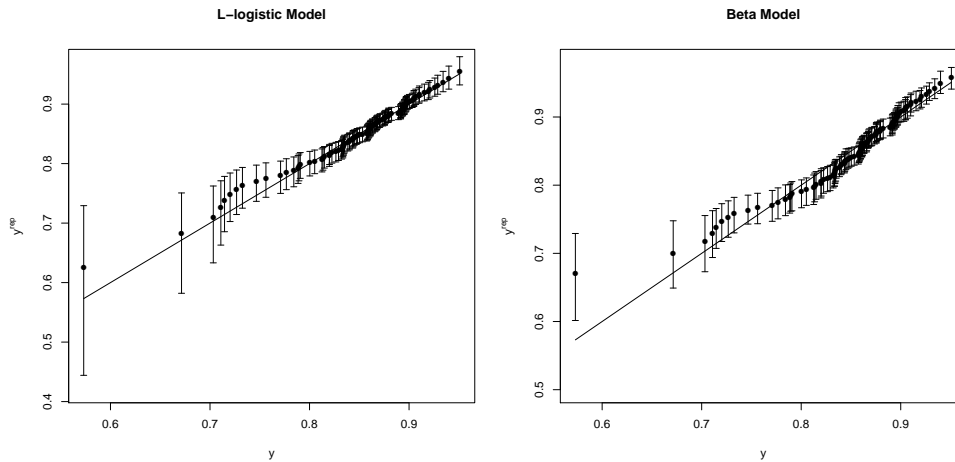


Figure 12 – Posterior predictive error bars with 95% confidence intervals of the generated values  $y_{(i)}^{rep}$  versus ordered observed data  $y_{(i)}$  for the PPOBC data, using L-Logistic and Beta models.

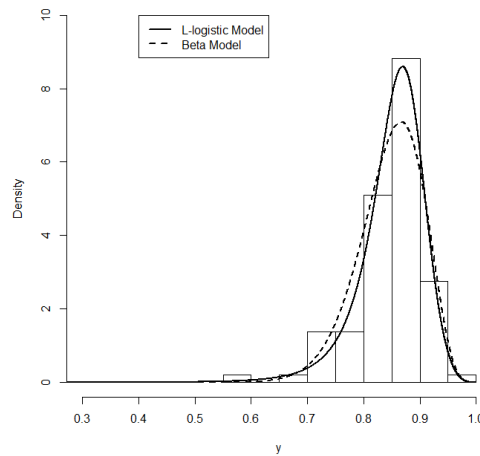


Figure 13 – Estimated density of PPOBC data.

### **Regression analysis with L-Logistic distribution**

Regression analysis estimates the potential differential effect of a covariate on mean or quantiles on the conditional distribution (HAO; NAIMAN, 2007). Here, we are interested in studying the conditional (or regression) median as a function of the covariates, when the response variable takes values in a bounded interval. In the analysis with the L-Logistic distribution, we assume that conditional on the explanatory variables (covariates), the random variable  $Y_i$ ,  $i = 1, \dots, n$ , are mutually independent with L-Logistic distribution,  $Y_i \sim LL(m_i, b_i)$ . Thus, given  $\mathbf{x}_{1i}^T$  and  $\mathbf{x}_{2i}^T$  ( $q$  and  $d$ -dimensional vectors, respectively, containing the explanatory variables both

with 1 as the first component), the likelihood of the observed sample  $\mathbf{y} = (y_1, \dots, y_n)$  can be written as

$$L(\boldsymbol{\beta}, \boldsymbol{\delta} | \mathbf{y}, \mathbf{X}) = \prod_{i=1}^n \frac{b_i (1 - m_i)^{b_i} m_i^{b_i} y_i^{b_i-1} (1 - y_i)^{b_i-1}}{[(1 - m_i)^{b_i} y_i^{b_i} + m_i^{b_i} (1 - y_i)^{b_i}]^2}, \quad (5.35)$$

where  $\mathbf{X}$  is the matrix containing all explanatory variables, and

$$h_1(m_i) = \mathbf{x}_{1i}^T \boldsymbol{\beta} \quad \text{and} \quad h_2(b_i) = \mathbf{x}_{2i}^T \boldsymbol{\delta}, \quad (5.36)$$

with  $\boldsymbol{\beta} = (\beta_0, \dots, \beta_{q-1})$  and  $\boldsymbol{\delta} = (\delta_0, \dots, \delta_{d-1})$  representing, respectively, the  $q$  and  $d$ -dimensional vectors of unknown regression parameters. In (5.36),  $h_1$  and  $h_2$  are strictly monotone and twice differentiable real link functions. This method allows to fit the model adequately with a variety of link functions, which ensures parameter  $m$  is in the interval  $(0, 1)$  and the shape parameter  $b$  is in the interval  $(0, \infty)$ . Some link functions for a scale parameter is discussed in Ferrari and Cribari-Neto (2004). A common link function for the parameter  $m$  is the logit function,

$$\text{logit}(m_i) = \mathbf{x}_{1i}^T \boldsymbol{\beta}. \quad (5.37)$$

For the shape parameter, a common link function is the log-linear link function. For easy interpretation, here we follow Smithson and Verkuilen (2006) and consider  $h_2 = -\log(b_i)$ , that is,

$$\log(b_i) = -\mathbf{x}_{2i}^T \boldsymbol{\delta}. \quad (5.38)$$

In addition, we adopt the following proper prior distributions with large variance as is frequently considered in the literature:

$$\begin{aligned} \beta_j &\sim \text{Normal}(0, 100), \text{ for } j = 0, \dots, q-1, \\ \delta_l &\sim \text{Normal}(0, 100), \text{ for } l = 0, \dots, d-1. \end{aligned} \quad (5.39)$$

Thus, samples of the joint posterior distribution of  $\boldsymbol{\beta}$  and  $\boldsymbol{\delta}$  can be obtained by considering the MCMC method to simulate from the posterior distribution, with pdf given by

$$\pi(\boldsymbol{\beta}, \boldsymbol{\delta} | \mathbf{y}) \propto L(\boldsymbol{\beta}, \boldsymbol{\delta} | \mathbf{y}, \mathbf{x}) \pi(\boldsymbol{\beta}) \pi(\boldsymbol{\delta}). \quad (5.40)$$

In order to illustrate the regression analysis with L-Logistic distribution, we analyzed a know data set in the literature previously analyzed using Beta distribution. This real data come from a sample of nonclinical women in Townsville, Queensland, Australia. The data contain 166 observations on two variables, namely the stress score and the anxiety score. Both variables were assessed on the depression anxiety stress Scales, ranging from 0 to 42, but linearly transformed to the open unit interval by Smithson and Verkuilen (2006). The scatterplot of the anxiety versus stress variable, and the histograms of the data, are presented in Figure 15. The histogram given in this figure suggest that the anxiety is strongly skewed.

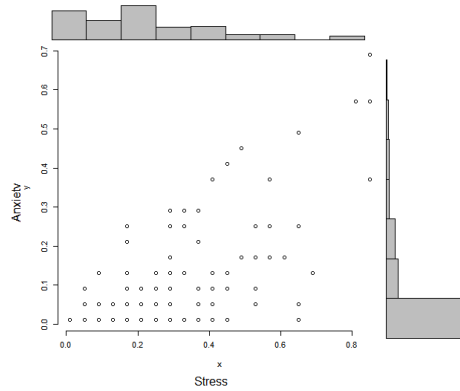


Figure 14 – Scatterplot and histograms of the real data.

Considering the data, we propose four possible regression sub-models using L-Logistic distribution. We consider a null regression model without any covariate, a scale regression model considering only covariate effects in parameter  $m$ , a shape regression model considering only covariate effects in the shape parameter  $b$ , and full regression model considering both effects, as follows:

$$Y_i \sim LL(m_i, b_i) \text{ and } \begin{cases} \text{null model } (L_0) : & \text{logit}(m_i) = \beta_0 \text{ and } \log(b_i) = -\delta_0, \\ \text{median-model } (L_1) : & \text{logit}(m_i) = \mathbf{x}_{1i}^T \boldsymbol{\beta} \text{ and } \log(b_i) = -\delta_0, \\ \text{shape model } (L_2) : & \text{logit}(m_i) = \beta_0 \text{ and } \log(b_i) = -\mathbf{x}_{1i}^T \boldsymbol{\delta}, \\ \text{full model } (L_3) : & \text{logit}(m_i) = \mathbf{x}_{1i}^T \boldsymbol{\beta} \text{ and } \log(b_i) = -\mathbf{x}_{2i}^T \boldsymbol{\delta}, \end{cases}$$

for  $i = 1, \dots, 166$ . In addition, we also consider equivalent regression models using the Beta distribution, as follows:

$$Y_i \sim \text{Beta}(\mu_i, \phi_i) \text{ and } \begin{cases} \text{null model } (B_0) : & \text{logit}(\mu_i) = \beta_0 \text{ and } \log(\phi_i) = -\delta_0, \\ \text{mean-model } (B_1) : & \text{logit}(\mu_i) = \mathbf{x}_{1i}^T \boldsymbol{\beta} \text{ and } \log(\phi_i) = -\delta_0, \\ \text{dispersion model } (B_2) : & \text{logit}(\mu_i) = \beta_0 \text{ and } \log(\phi_i) = -\mathbf{x}_{1i}^T \boldsymbol{\delta}, \\ \text{full model } (B_3) : & \text{logit}(\mu_i) = \mathbf{x}_{1i}^T \boldsymbol{\beta} \text{ and } \log(\phi_i) = -\mathbf{x}_{2i}^T \boldsymbol{\delta}, \end{cases}$$

for  $i = 1, \dots, 166$ .

Here, the Bayesian approach is considered for the inference process with prior distribution for the unknown regression parameters as given in (5.39). All algorithms were prepared in R language and we report the results corresponding to 10,000 iterations following a burn-in period also of 10,000 iterations. The convergence of MCMC chain was assessed by using the separated partial means test of Geweke (1992), which provided evidence for the chains to have converged.

Table 15 – Model comparison criteria for model comparison.

Sub model	L-Logistic model				Beta model			
	WAIC	EAIC	EBIC	DIC	WAIC	EAIC	EBIC	DIC
0	259.34	-512.94	-506.72	-514.92	239.45	-472.90	-466.67	-474.90
1	277.67	-545.51	-536.17	-548.56	243.28	-478.06	-468.72	-481.07
2	316.57	-624.78	-615.44	-627.83	283.41	-556.95	-547.62	-559.92
3	319.65	-627.47	-615.02	-631.48	301.91	-591.85	-579.41	-595.82

The regression models investigated were compared by the use of EAIC, EBIC, DIC and WAIC criteria, which result are shown in Table 15. The parameter estimates for these models are shown in Table 16. Considering Table 15, we can observe that the regression models considering L-Logistic distribution provide better fit than the corresponding Beta regression models, for all criteria considered. These results also give evidence that the  $L_3$  and  $L_2$  are the best models among the ones using the L-Logistic distribution. Though there is no significant difference between the  $L_2$  and  $L_3$  regressions, we consider the  $L_3$  regression model to be a reasonable choice for this data set, due to its expected influence on covariates in the shape parameter (SMITHSON; VERKUILEN, 2006). Additional diagnostic analysis could show further evidence that the M3 model is an appropriate model for the data.

Moreover, a posterior distribution of residuals was obtained and a posterior mean of this distribution was computed (GELMAN *et al.*, 2013). That is, for  $i = 1, \dots, 166$ , we have  $\hat{r}_i = G^{-1} \sum_{g=1}^G \frac{y_i - \hat{y}^g}{SD(Y_i | \boldsymbol{\beta}^g)}$ , where  $\boldsymbol{\beta}^1, \dots, \boldsymbol{\beta}^G$  are obtained from the posterior distribution,  $\hat{y}$  is the estimated value for a data point  $y_i$ , and  $SD(Y | \boldsymbol{\beta}^g)$  is the standard deviation of posterior values of  $Y$ , both obtained given a single random draw  $\boldsymbol{\beta}^g$  of the posterior distribution. Figure 15 shows the standard residual versus the estimated values in which we can see that the  $L_3$  regression model provides better fit than the  $B_3$  model, which confirms that  $L_3$  model are better than the corresponding  $B_3$  model.

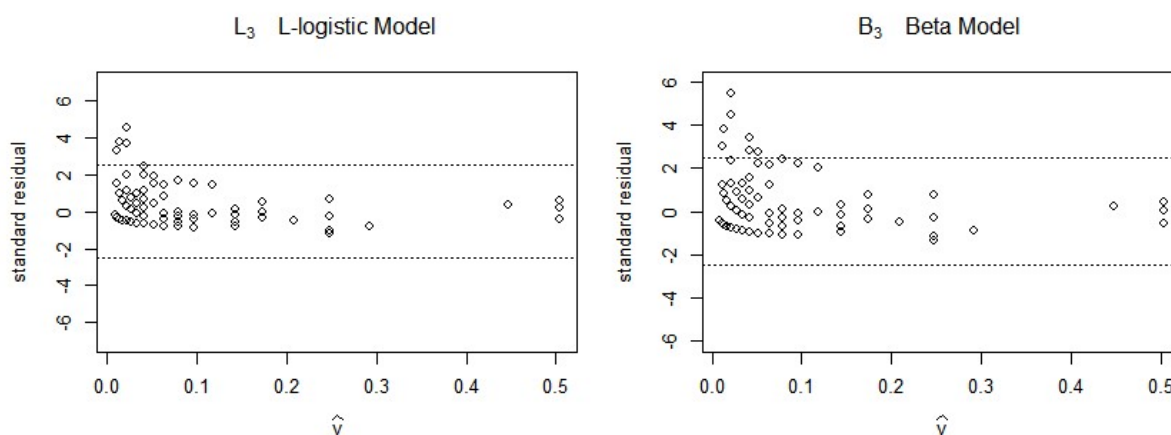


Figure 15 – Standard residual versus adjusted values for the L-Logistic and Beta models.



Table 16 – Parameter estimates and 95% HPD intervals for the L-Logistic and Beta models.

Model		Coefficient			
		$\hat{\beta}_0$ (HPD)	$\hat{\beta}_1$ (HPD)	$\hat{\delta}_0$ (HPD)	$\hat{\delta}_1$ (HPD)
L-Logistic	$L_0$	-3.354 (-3.615, -3.106)	-	-0.08(-0.20, 0.04)	-
	$L_1$	-4.78(-5.04, -4.52)	5.78(4.99, 6.58)	-0.44(- 0.57,-0.31)	-
	$L_2$	-4.03(-4.27, -3.78)	-	-0.87(-1.14,-0.58)	2.56( 1.62,3.42)
	$L_3$	-4.77(-5.00, -4.53)	5.64(4.68, 6.61)	-0.76(-1.03, -0.48)	1.14( 0.19, 2.05)
Beta	$B_0$	-2.239(-2.430, -2.04)	-	-1.78(- 2.02,-1.54)	-
	$B_1$	-3.47(-3.75, -3.18)	3.74(3.11, 4.37)	-2.44(-2.7,-2.20)	-
	$B_2$	-2.54(-2.80, -2.27)	-	-2.49(-2.98,-1.95)	1.53(0.42,2.55)
	$B_3$	-4.02 (-4.30, -3.72)	4.95( 4.09, 5.83)	-3.94(-4.45,-3.44)	4.28 ( 2.78,5.79)

Finally, in Table 16, by considering the 95% HPD intervals for all coefficients of the models under analysis, we can see that the estimatives are very precise. For the model chosen, that is,  $L_3$  model, we observe that the HPD intervals for the estimates of the parameters  $\beta_1$  and  $\delta_1$  do not contain zero given evidence that is parameters are significant in the model. That is, stress is important in both parameters of the distribution of anxiety. Then, the final model is given by

$$Y_i \sim LL(\hat{m}_i, \hat{b}_i) \text{ with } \begin{cases} \hat{m}_i = \frac{\exp\{-4.765+5.642 \times stress_i\}}{1+\exp\{-4.765+5.642 \times stress_i\}}, \\ \hat{b}_i = \exp\{0.755 - 1.137 \times stress_i\}, \end{cases}$$

for  $i = 1, \dots, 166$ . Considering this estimated model, we can conclude that the stress, positively, influence the anxiety. Considering  $b_i$  a dispersion parameter, see [Lemonte and Bazán \(2016\)](#), the stress influence also the dispersion, i.e., higher values of stress are associated with increased variability in anxiety.

## 5.7 Final remarks

The L-Logistic distribution, introduced by [Tadikamalla and Johnson \(1982\)](#), is a bounded continuous distribution that possesses some nice properties, as discussed in Section 5.3. Considering the parameterization introduced in this manuscript, we propose a Bayesian estimator by considering MCMC method as an alternative to the moments and maximum likelihood methods developed previously in the literature. In the Bayesian context, a non-informative prior distribution can be adopted for the parameter  $m$  since it lies in the unit interval, enabling the use of unit Uniform distribution as a non-informative prior distribution. Two simulation studies are presented in Section 5.6 for evaluating the posterior distribution with respect to the specification of the prior distribution for the shape parameter  $b$  and to evaluate the performance of the Bayesian estimator chosen. In the first study, for the studied cases, we observe that the non-informative prior distributions provide correct information about parameter  $b$ , based on the results of WAIC, EAIC, EBIC, and DIC. Thus, the posterior distribution is not sensitive with respect to the specification of these prior distributions. Some characteristics of the posterior distribution are also calculated using two non-informative and one informative prior distributions for the parameter  $b$ . In this study, we observe that the prior information is dominated by the

sample information. In the second study, we evaluate the estimates of the parameters of the L-Logistic distribution obtained by using Bayesian method upon considering the prior Gamma distribution with parameters vector  $(0.001, 0.001)$ . We observe that the  $\sqrt{MSE}$  and bias lie both close to zero even when the sample size is small. Hence, for the samples analyzed, the estimator seems to provide reasonable estimates.

The main motivation of the parameterization introduced was the development of regression models considering the L-Logistic distribution. Thus, we also introduce conditional median regression models, which is a special case of quantile regression wherein the conditional 0.5th quantile is modeled as a function of covariates. By considering as application to a known data set (anxiety explained by stress), we show that the L-Logistic regression models can be a good alternative to Beta model. An advantage of this approach is the possibility of modeling other quantiles in order to describe a non-central position of a distribution. So, one may choose a specific position for his/her needs. For example, it is possible to consider a regression model to explain other quantiles the anxiety considering the influence of the stress in our application. Thus, conditional quantile models offer the flexibility to focus on these population segments, whereas conditional mean models do not. However, since quantile regression curves are estimated individually, the quantile curves can cross, leading to an invalid distribution for the response. Thus, this problem, referred to as crossing in the literature, needs to be studied carefully. Some authors have proposed methods to deal with this problem; see, for example, [Cai and Jiang \(2015\)](#).

Two application was considered in this work. Firstly, we consider an application to social data wherein the proportion of children vulnerable to poverty of the municipalities of the state of Alagoas in Brazil, for the 2010 season, is modeled. Second, we analyze a known data set, previously analyzed using Beta distribution by [Smithson and Verkuilen \(2006\)](#), which contain the stress score and the anxiety score. Here, the anxiety variable is modeled as a function of the stress. For the case of the L-Logistic distribution, we use a regression model proposed in this work. In these applications, we observe that the L-Logistic distribution seems to fit better than the Beta model for both of the case.

For future, we aim to develop techniques for mixed quantile regression for the L-Logistic distribution. Moreover, we intend to explore mixtures of L-Logistic distributions in a Bayesian framework as well as a multivariate version of this distribution.

In the next chapter we extend the model treated here for a mixture of distributions in the mixed model context.

---

# FINITE MIXTURE OF MIXED L-LOGISTIC REGRESSION: A BAYESIAN APPROACH

---

---

In this chapter, we extend the proposed model of the previous chapter to a finite mixture of mixed L-Logistic regression models. We apply this model to a simulated data set and to the real data shown in the Chapter 2.

---

## Abstract

A median regression model can be used to investigate the relationship between the central location of the response and a set of covariates, as similar in the conditional mean regression modeling. Here, we consider an alternative quantile regression approach to model data of proportions, using the L-Logistic distribution. Specifically, we consider a mixture of mixed-effects regression for model the median of proportion data. For this model, a Bayesian estimation considering a *Gibbs sampling* algorithm with *Metropolis-Hasting* inside is developed. Thus, the model is applied to a simulated data providing good estimates for the parameters. Applications to real data are also performed for mixed L-Logistic models with and without mixture.

## 6.1 Introduction

Studies where the response variables is a proportion or rates are very common in social area. If the response variable is observed over time a correlation among observations might occur and it should be taken into account in the analysis. A way to deal of this correlation is the inclusion of a random effect in the linear predictor to explain variability not considered in the model that can influence the result. [Figuroa-Zúñiga, Arellano-Valle and Ferrari \(2013\)](#)

and Bonat, Ribeiro and Zeviani (2015) adopt the mixed Beta regression model for independent response variable taking values in the interval  $(0, 10)$ , where the mean is conveniently linked with a mixed-effects regression structure by a convenient link function. However, if the distribution of the data is highly skewed, then the median can be more informative than the mean, as shown in the Chapter 5. In this case, the conditional median modeling has the potential to be more useful than conditional mean modeling, and can be used to investigate the relationship between the central location of the response and a set of covariate. In this work, we focus on the conditional quantiles of the response variable, which take value in the unit interval. Specifically, the median is linked with a mixed-effects regression structure by a link function, instead of the mean. The mixed quantile regression model considered here consider the L-Logistic distribution, which was originally proposed by Tadikamalla and Johnson (1982) and investigated more fully in the Chapter 5. We focus on extension of the median regression model considering L-Logistic distribution from a Bayesian point of view, and a mixed quantile regression model. The specific model considered here take into account the correlation that can arise when, for example, the response variable is observed over the time, i.e., when we have longitudinal data. These models are applied to simulated and real data set. In addition, we propose a finite mixture of mixed-effects regression.

The rest of the work is organized as follows. In Section 6.2, we present the pdf of L-Logistic distribution with the new parametrization. Section 6.3 is dedicated to describe the median regression model for this distribution. The mixed quantile regression model considering the L-Logistic distribution is presented in Section 6.4 with a Bayesian estimate and application to simulated and real data sets. In the section 6.5, the finite mixture of mixed-effects regression is developed, which is applied to longitudinal data. Finally, remarks are shown in Section 6.6.

## 6.2 L-Logistic distribution

By considering the notation used in Chapter 5, the random variable (r.v.)  $Y$  follows a L-Logistic distribution parameterized in terms of its median  $m$  ( $0 < m < 1$ ) and shape parameter  $\varphi$  ( $\varphi > 0$ ) if its probability density function (pdf) is given by

$$f(y|m, \varphi) = \frac{\varphi(1-m)^\varphi m^\varphi y^{\varphi-1} (1-y)^{\varphi-1}}{[(1-m)^\varphi y^\varphi + m^\varphi (1-y)^\varphi]^2}, \quad 0 < y < 1 \quad (6.1)$$

We denote by  $Y \sim LL(m, \varphi)$  a r.v.  $Y$  following the L-Logistic distribution with parameters  $m$  and  $\varphi$ .

The L-Logistic distribution is very flexible in term of the variety of density shapes that its two parameters allow to accommodate (see Figures 16 and 17). Note that when we set  $m = 0.5$  and  $b = 1$  in (6.1), then the pdf of the L-Logistic distribution simply becomes the pdf of the standard uniform distribution. Here,  $m$  is the median of the distribution, which scales the graph to the left or right on the horizontal axis. On the other hand,  $b$  is a parameter that governs the

shape of the distribution. The L-Logistic density is uni-modal (or “uni-antimodal”), increasing, decreasing, or constant, depending on the values of its parameters. More details on this issue are presented with another parametrization of the L-Logistic model is presented in Section 5.3.

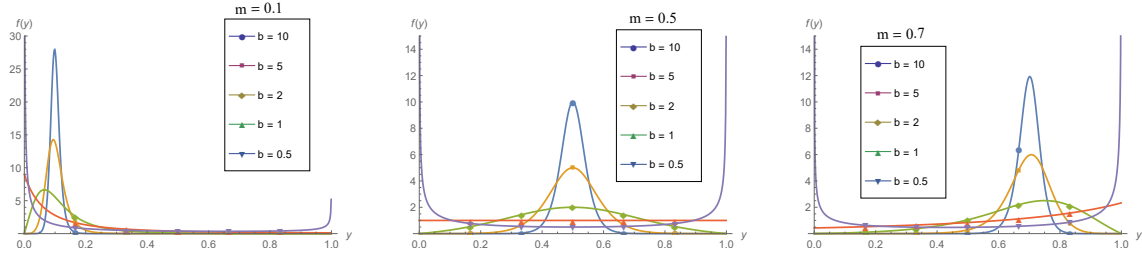


Figure 16 – L-Logistic probability density function for scale parameter  $m = 0.2, 0.5$  and  $0.8$  and some values of parameter  $b$ .

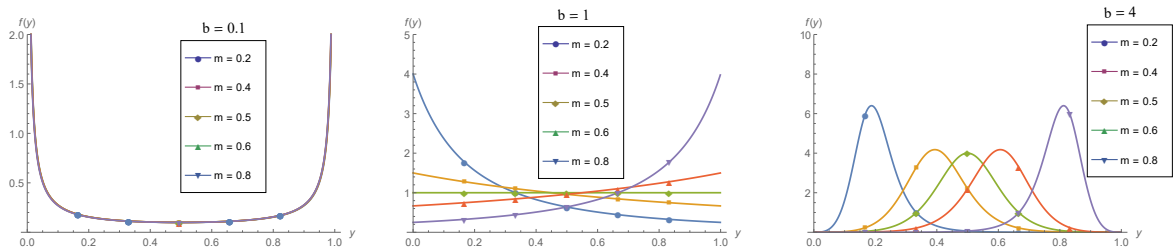


Figure 17 – L-Logistic probability density function for shape parameter  $b = 0.5, 1$  and  $2$  and some values of scale parameter  $m$ .

As mentioned in the Chapter 5, the functions that provide the probability density function, cumulative density function, quantile function and random generation for the L-Logistic distribution with parameters  $m$  and  $b$  are available in the CRAN of R program, details can be see in Paz and Bazan (2017).

### 6.3 L-Logistic median regression model

In the regression analysis with the L-Logistic distribution, we assume that conditional on the explanatory variables (covariates), the random variable  $Y_i$ ,  $i = 1, \dots, n$ , are mutually independent with L-Logistic distribution, i.e.,

$$Y_i \sim LL(m_i, \varphi_i), \quad (6.2)$$

where

$$m_i = g^{-1}(\mathbf{x}_i^T \boldsymbol{\beta}) = g^{-1}(\eta_i). \quad (6.3)$$

In (6.3),  $\boldsymbol{\beta} = (\beta_0, \dots, \beta_{q-1})$  represent the  $q$ -dimensional vector of unknown regression parameters, and  $g(\cdot)$  is the link function that relates the median of the response to the linear predictors ( $\eta_i$ ) in the model. Let’s call this model L-Logistic regression (LLR).

Given  $\mathbf{x}_i(p \times 1)$  ( $q$ -dimensional vector contain the explanatory variables with 1 in the first component), the likelihood of the observed sample  $\mathbf{y} = (y_1, \dots, y_n)$  can be written as

$$L(\boldsymbol{\beta}, \varphi | \mathbf{y}, \mathbf{X}) = \prod_{i=1}^n \frac{\varphi(1-m_i)^\varphi m_i^\varphi y_i^{\varphi-1} (1-y_i)^{\varphi-1}}{[(1-m_i)^\varphi y_i^\varphi + m_i^\varphi (1-y_i)^\varphi]^2}, \quad (6.4)$$

where  $\mathbf{X} = (\mathbf{x}_1, \mathbf{x}_2, \dots, \mathbf{x}_n)$  is the matrix containing all explanatory variables.

### Bayesian estimation for LLR model

For the development of the Bayesian estimation, it is necessary to consider prior distributions for all unknown model parameters. The priors distribution of parameters is choose here under the assumption that they are independent to each other. Then, following the Chapter 5, for the parameter  $\varphi$  we consider the transformation  $\log(\varphi) = -\delta$ ,  $\delta \in \mathbb{R}$ , and the normal prior for  $\delta$  such that  $\delta \sim N(0, 100)$ . For the regression coefficients  $\boldsymbol{\beta}$  we adopt the multivariate normal distribution as a prior, i.e.,  $\boldsymbol{\beta} \sim N_q(\mathbf{0}, \Sigma)$ . All hyperparameters is assumed known. Assuming prior distributions for the parameters, the posterior density has the form

$$\pi(\mathbf{b}, \boldsymbol{\beta}, \delta | \mathbf{y}) \propto L(\boldsymbol{\beta}, \varphi | \mathbf{y}, \mathbf{X}) \pi(\boldsymbol{\beta}) \pi(\delta) = \prod_{i=1}^n f(y_i | \mathbf{x}_i, \boldsymbol{\beta}, \delta) \pi(\boldsymbol{\beta}) \pi(\delta), \quad (6.5)$$

where  $\delta = -\log\{\varphi\}$ . Therefore, the full conditional posterior distributions for  $\boldsymbol{\beta}$  and  $\delta$  can be written as

$$\begin{aligned} \pi(\boldsymbol{\beta} | \mathbf{x}_i, \delta) &\propto \prod_{i=1}^n f(y_i | \mathbf{x}_i, \boldsymbol{\beta}, \delta) \pi(\boldsymbol{\beta}), \\ \pi(\delta | \mathbf{x}_{ij}, \boldsymbol{\beta}) &\propto \prod_{i=1}^n f(y_i | \mathbf{x}_i, \boldsymbol{\beta}, \delta) \pi(\delta). \end{aligned} \quad (6.6)$$

Using the posterior distribution given in (6.6), we can use several Bayesian mechanisms in order to estimate the parameters. A *Gibbs-sampling algorithm* with *Metropolis-Hasting* step inside is useful in this case due to because the conditional posterior for  $\boldsymbol{\beta}$  and  $\delta$  do not have closed form.

## 6.4 L-Logistic mixed median regression (LLMR) model

The model described in previous section does not incorporates intra-cluster correlations. Thus, in order to extend the regression models based on L-Logistic distribution to model correlated proportional data, we specify the L-Logistic mixed-effects model, LLMR model. Thus, to specify the model, consider a sequence  $\mathbf{Y}_i = (Y_{i1}, Y_{i2}, \dots, Y_{in_i})$ ,  $Y_{ij} \in (0, 1)$ , on the  $i$ th sample unit taken at  $n_i$  subject. Considering a vector  $\mathbf{x}_{ij}$  of explanatory variable on the  $i$ th sample unit at the  $j$ th subject,  $j \leq n_i$ , the main issue can be about a model for  $\mathbf{Y}_i$  that incorporates intra-cluster correlation. A way to deal with this issue is via random effects formulation, in which a  $d$ -dimensional vector  $\mathbf{b}_i$  of random effects is considered for subject  $i$ . Let's assume that, given  $\mathbf{b}_i$ , the  $Y_{ij}$ 's are conditionally independent with median  $m_{ij} = \text{med}(y_{ij} | \mathbf{b}_i)$ , and

$$(Y_{ij} | \mathbf{b}_i) \sim LL(m_{ij}, \varphi), \quad (6.7)$$

in which the conditional median is related to the linear predictor

$$\eta_{ij} = g(m_{ij}) = \mathbf{x}_{ij}^T \boldsymbol{\beta} + \mathbf{v}_{ij}^T \mathbf{b}_i, \quad (6.8)$$

where  $\mathbf{v}_{ij}$  is a set of explanatory variables associated with the random effect,  $\boldsymbol{\beta}$  is a  $q$ -dimensional vector of fixed effects, and  $g(\cdot)$  is a link function. The model specification is completed by assuming

$$\mathbf{b}_i \sim N_d(\mathbf{0}, \mathbf{D}), \quad (6.9)$$

for  $i = 1, \dots, m$ . The parameters  $\boldsymbol{\beta}$  and the matrix  $\mathbf{D}$  are unknown.

### Bayesian estimation for LLMR model

Under assumption of conditional independence among the components of  $\mathbf{b} = (\mathbf{b}_1^T, \dots, \mathbf{b}_m^T)^T$  (given its parameters), the joint density function of  $\mathbf{Y} = (\mathbf{Y}_1^T, \dots, \mathbf{Y}_m^T)^T$  and  $\mathbf{b}$  is given by

$$f(\mathbf{y}, \mathbf{b} | \mathbf{X}, \mathbf{V}, \boldsymbol{\beta}, \varphi, \mathbf{D}) = \prod_{i=1}^m \prod_{j=1}^{n_i} f(y_{ij} | \mathbf{x}_{ij}, \mathbf{v}_{ij}, \boldsymbol{\beta}, \varphi, \mathbf{b}_i) f(\mathbf{b}_i | \mathbf{0}, \mathbf{D}), \quad (6.10)$$

where  $\mathbf{X} = (\mathbf{x}_1, \dots, \mathbf{x}_m)$  and  $\mathbf{V} = (\mathbf{v}_1, \dots, \mathbf{v}_m)$  contain the explanatory variables,  $f(y_{ij} | \mathbf{x}_{ij}, \mathbf{v}_{ij}, \boldsymbol{\beta}, \varphi, \mathbf{b}_i)$  is the density of the L-Logistic distribution with parameters  $m_{ij} = g^{-1}(\mathbf{x}_{ij}^T \boldsymbol{\beta} + \mathbf{v}_{ij}^T \mathbf{b}_i)$  and  $\varphi$ , and  $f(\mathbf{b}_i | \mathbf{0}, \mathbf{D})$  is the density of the normal distribution.

As in previous section, for the parameter  $\varphi$  we consider the transformation  $\log(\varphi) = -\delta$ ,  $\delta \in \mathbb{R}$ , and the normal prior for  $\delta$  such that  $\delta \sim N(0, 100)$ . For the fixed effect, we adopt the multivariate normal distribution as a prior,  $\boldsymbol{\beta} \sim N_q(\mathbf{0}, \boldsymbol{\Sigma})$ . Finally, the inverse Wishart distributions is adopted as prior for the covariance matrices of the random effects,  $\mathbf{D} \sim IW_q(\gamma/2, \mathbf{W}/2)$ . Thus, the augmented jointly posterior density for  $\mathbf{b}$ ,  $\boldsymbol{\beta}$  and  $\delta$  is given by

$$\pi(\mathbf{b}, \boldsymbol{\beta}, \delta | \mathbf{y}, \mathbf{X}, \mathbf{V}, \mathbf{D}) \propto \left[ \prod_{i=1}^m \prod_{j=1}^{n_i} f(y_{ij} | \mathbf{x}_{ij}, \mathbf{v}_{ij}, \boldsymbol{\beta}, \delta, \mathbf{b}_i) \pi(\boldsymbol{\beta}) \pi(\delta) \right] \prod_{i=1}^m f(\mathbf{b}_i | \mathbf{0}, \mathbf{D}) \pi(\mathbf{D}), \quad (6.11)$$

where  $\delta = -\log(\varphi)$ . Therefore, the full conditional posterior distributions for  $\boldsymbol{\beta}$ ,  $\delta$ , and  $\mathbf{b}_i$ ,  $i = 1, \dots, m$ , can be written as

$$\begin{aligned} \pi(\boldsymbol{\beta} | \mathbf{y}, \mathbf{X}, \mathbf{V}, \delta, \mathbf{b}) &\propto \prod_{i=1}^m \prod_{j=1}^{n_i} f(y_{ij} | \mathbf{x}_{ij}, \mathbf{v}_{ij}, \boldsymbol{\beta}, \delta, \mathbf{b}_i) \pi(\boldsymbol{\beta}), \\ \pi(\delta | \mathbf{y}, \mathbf{X}, \mathbf{V}, \boldsymbol{\beta}, \mathbf{b}) &\propto \prod_{i=1}^m \prod_{j=1}^{n_i} f(y_{ij} | \mathbf{x}_{ij}, \mathbf{v}_{ij}, \boldsymbol{\beta}, \delta, \mathbf{b}_i) \pi(\delta), \\ \pi(\mathbf{b}_i | \mathbf{y}, \mathbf{X}, \mathbf{V}, \boldsymbol{\beta}, \delta) &\propto \prod_{j=1}^{n_i} f(y_{ij} | \mathbf{x}_{ij}, \mathbf{v}_{ij}, \boldsymbol{\beta}, \delta, \mathbf{b}_i) f(\mathbf{b}_i | \mathbf{0}, \mathbf{D}) \pi(\mathbf{D}), \text{ for } i = 1, \dots, m, \end{aligned} \quad (6.12)$$

where  $\pi(\boldsymbol{\beta})$  and  $\pi(\delta)$  are the prior distributions for  $\boldsymbol{\beta}$  and  $\delta$ , respectively. The full conditional

posterior distribution of  $\mathbf{D}$  is given by

$$\begin{aligned}
\pi(\mathbf{D}|\mathbf{y}, \mathbf{X}, \mathbf{V}, \boldsymbol{\beta}, \delta, \mathbf{b}) &\propto \prod_{i=1}^m f(\mathbf{b}_i | \mathbf{0}, \mathbf{D}) \pi(\mathbf{D}) \\
&\propto \mathbf{D}^{-m/2} \exp \left\{ -\frac{1}{2} \sum_{i=1}^m \mathbf{b}_i^T \mathbf{D}^{-1} \mathbf{b}_i \right\} \times \\
&\quad |\mathbf{D}^{-1}|^{\frac{\gamma-r-1}{2}} \exp \left\{ -\frac{1}{2} \text{tr} [\mathbf{W} \mathbf{D}^{-1}] \right\} \\
&\propto \mathbf{D}^{-m/2} \exp \left\{ -\frac{1}{2} \text{tr} [(\sum_{i=1}^m \mathbf{b}_i \mathbf{b}_i^T) \mathbf{D}^{-1}] \right\} \times \\
&\quad |\mathbf{D}^{-1}|^{\frac{\gamma-r-1}{2}} \exp \left\{ -\frac{1}{2} \text{tr} [\mathbf{W} \mathbf{D}^{-1}] \right\} \\
&\propto |\mathbf{D}^{-1}|^{\frac{m+\gamma-r-1}{2}} \exp \left\{ -\frac{1}{2} \text{tr} [(\mathbf{W} + \sum_{i=1}^m \mathbf{b}_i \mathbf{b}_i^T) \mathbf{D}^{-1}] \right\},
\end{aligned} \tag{6.13}$$

where  $r = \dim(\mathbf{D})$ . It follows from Equation (6.13) that the full conditional posterior of  $\mathbf{D}$  must be a inverse Wishart distribution.

With this information, as said before, it is possible to use several Bayesian mechanisms in order to estimate the parameters. Here, we adopt a *Gibbs-sampling algorithm* with *Metropolis-Hasting* step due to because the conditional posterior for  $\boldsymbol{\beta}$  and  $\delta$  does not have closed form.

### **Application of the models LLR and LLMR to real data**

In this section, the MCMC algorithm for estimating parameters in L-Logistic mixed-effects model is applied to real data analyzed in Chapter 4. There, we consider the Human Development Index of municipalities, MHDI, of Northeast region and São Paulo state in Brasil, 2010 year. The model considered was a regression model where the response variable follows a mixture of Simplex distributions. The variable MHDI (described in Chapter 3) was taken as response and the PPPM (proportion of poor people per municipality, described in Chapter 4) was taken as the covariate. Here, we analyze these data sets using quantile regression models with and without random effect. Our interest lies in the relationship between the MHDI variable and the PPPM variable. Since in Northeast region in Brazil the municipalities are grouped in 9 states and we considered, additionally, the São Paulo state, we have 10 State to be analyzed. The model with random effect, LLMR, for the median of the MHDI was formulated as follows:

$$\begin{aligned}
(Y_{ij} | \mathbf{b}_i) &\sim LL(m_{ij}, \varphi), \\
\text{logit}(m_{ij}) &= \beta_0 + \beta_1 x_{ij} + b_{0i}, \\
\log(\varphi) &= -\delta,
\end{aligned} \tag{6.14}$$

for  $j = 1, \dots, n_i$  municipalities and  $i = 1, \dots, 10$  states, where  $x_{ij}$  is a proportion of poor people in the  $i$ th municipality of  $j$ th state. In this case, we introduce a random intercept associate with the state. A L-Logistic regression model, as described in the previous section, was considered for



the same data. That is, we also considered the LLR model as follows:

$$\begin{aligned} Y_i &\sim LL(m_i, \varphi), \\ \text{logit}(m_i) &= \beta_0 + \beta_1 x_i, \\ \log(\varphi) &= -\delta, \end{aligned} \quad (6.15)$$

for all  $i = 1, \dots, 2439$  municipalities.

Here, we consider prior independence between the parameters such that  $\beta_r \sim N(0, 100)$ , for  $r = 0, 1$ ,  $\delta \sim N(0, 100)$  and  $D \sim \text{gamma}(1/20, 1/20)$  with  $b_{0i} \sim N(0, D)$ , for  $i = 1, \dots, 10$ .

The estimates and 95% HDP intervals to the parameter of the LLRM and LLR models, applied to MHDI data, are shown in Table 17 where we can see that the amplitudes of the 95% HDP intervals are small. Based on the results shown in this table, we can conclude that, for both models, the MHPI has negative effect on the median of the MHDI, as expected. Additionally, the variance of random intercept in the LLMR model is quite close to zero, that is, the effect of the state seems low when compared with the effect of the covariate. It is due to we have just one full region where the states have similar characteristics. For the same reason, the estimated values for the random effect are all close to zero. Table 17, also shown the model comparison using the criteria presented in Chapter 5.4. Based on this criteria, we can conclude that there is no significant difference between the considered models, for more information about the obtaining of this critérios for the LLMR model see, for example, [Bayes, Bazán and Castro \(2017\)](#). Thus, after a selection based on parsimonious criteria we can conclude that the LLR model is better than LLMR model for this application, and then a random intercept is not necessary in the model.

Table 17 – Posterior mean and credibility intervals of the estimated parameters for the MHDI data, and model comparison between the LLR and LLMR models.

	<b>L-Logistic Regression Model</b>	<b>L-Logistic Mixed Regression Model</b>
Parameter	Estimate (95% HPD)	Estimate (95% HPD)
$\beta_0$	1.35 (1.34, 1.36)	1.35, (1.33, 1.37)
$\beta_1$	-1.47 (-1.49, -1.45)	-1.47 (-1.50, -1.44)
$-\delta$	-2.77 (-2.73, -2.80)	-2.76(-2.72, -2.80)
$D$	-	0.012 (0.003, 0.025)
Estimated values for the random effect for each State		
São Paulo	-	1.2e-04 ( -2.6e-02, 2.4e-02)
Alagoas	-	-1.8e-04 (-2.5e-02, 2.3e-02)
Bahia	-	-8.0e-05 ( -2.6e-02, 2.5e-02)
Ceará	-	-3.9e-04 ( -2.7e-02, 2.3e-02)
Maranhão	-	-1.7e-04 ( -2.5e-02, 2.2e-02)
Paraíba	-	6.7e-05 (-2.4e-02, 2.5e-02)
Pernambuco	-	-3.6e-04 ( -2.5e-02, 2.4e-02)
Piauí	-	-5.0e-05 ( -2.5e-02, 2.5e-02)
Rio Grande do Norte	-	9.2e-05 ( -2.5e-02, 2.4e-02)
Sergipe	-	3.6e-04 ( -2.3e-02, 2.4e-02)
Criteria for model comparison		
WAIC	5530.481	5530.27
EAIC	-11051.836	-11020.86
EBIC	-11034.438	-10997.66
DIC	-11054.814	-11040.41

## 6.5 Mixture of L-Logistic mixed-effect models

Now, we present a new model based on a mixture of the LLMR model presented in the previous section to longitudinal data. We call this model of mixture of L-Logistic mixed median regression (MLLMR) model. For the model formulation, consider a sequence  $\mathbf{Y}_i = (Y_{i1}, Y_{i2}, \dots, Y_{in_i})$ ,  $Y_{ij} \in (0, 1)$ , on the  $i$ th sample unit taken at  $n_i$  points at the time (assumed equally spaced) and a vector  $\mathbf{x}_{ij}$  of explanatory variable on the  $i$ th sample unit at the  $j$ th time point, here we suppose  $n_1 = \dots = n_m = n$ . For the model formulation, suppose that, given  $\mathbf{b}_i$ , each  $Y_{ij}$  follows a finite mixture of L-Logistic distribution, i.e.,

$$(Y_{ij}|\mathbf{b}_i) \sim \sum_{c=1}^{k_j} \omega_{jc} LL(m_{ijc}, \varphi_c), \quad (6.16)$$

in which the conditional median is related to the linear predictor such that

$$\begin{aligned} \eta_{ijc} = g(m_{ijc}) &= \mathbf{x}_{ij}^T \boldsymbol{\beta}_c + \mathbf{v}_{ij}^T \mathbf{b}_i, \\ \log(\varphi_c) &= -\delta. \end{aligned} \quad (6.17)$$

for  $c = 1, \dots, k_j$ ,  $1 \leq k_j \leq k_{max}$  for  $j = 1, \dots, m$  with  $\delta_c \in \mathbb{R}$ . The density function of  $\mathbf{Y} = (\mathbf{Y}_1^T, \dots, \mathbf{Y}_m^T)^T$  given  $\mathbf{b}$  can be written as

$$f(\mathbf{y} | \boldsymbol{\beta}^*, \boldsymbol{\varphi}^*, \boldsymbol{\omega}^*, \mathbf{b}) = \prod_{i=1}^m \prod_{j=1}^{n_i} \sum_{c=1}^{k_j} \omega_{jc} f(y_{ij} | \mathbf{x}_{ij}, \mathbf{v}_{ij}, \boldsymbol{\beta}_c, \delta_c, \mathbf{b}_i), \quad (6.18)$$

where  $\boldsymbol{\beta}^* = (\boldsymbol{\beta}_1, \dots, \boldsymbol{\beta}_k)$ ,  $\boldsymbol{\delta}^* = (\delta_1, \dots, \delta_k)$ ,  $\boldsymbol{\omega}^* = (\boldsymbol{\omega}_1, \dots, \boldsymbol{\omega}_n)$ , with  $\boldsymbol{\omega}_j = (\omega_1, \dots, \omega_{k_j})$ , and  $f(\cdot)$  is the pdf of L-Logistic distribution. Note that, the random effect  $\mathbf{b} = (\mathbf{b}_1, \dots, \mathbf{b}_n)$  is the same for all components of the model. That is, the mixture is conditioned on  $\mathbf{b}$ .

### Bayesian inference to MLLMR model

In order to simplify the inference process, consider a unobserved random vector  $Z_{ij} = (Z_{ij1}, \dots, Z_{ijk_j})$  such that  $Z_{ijc} = 1$  if  $y_{ij}$  belongs to the  $c$ th mixture component and  $Z_{ijc} = 0$  otherwise,  $i = 1, \dots, m$ ,  $j = 1, \dots, n$ . Thus, the augmented data likelihood,  $(\mathbf{Y}, Z, \mathbf{b})$ ,  $Z = (Z_{11}, \dots, Z_{1m}, Z_{21}, \dots, Z_{2m}, \dots, Z_{n1}, \dots, Z_{nm})$ , is given by

$$L(\boldsymbol{\beta}^*, \boldsymbol{\delta}^*, \boldsymbol{\omega}^*, \mathbf{b} | \mathbf{y}, \mathbf{X}, \mathbf{V}, Z, \mathbf{D}) = \prod_{i=1}^m \prod_{j=1}^{n_i} \prod_{c=1}^{k_j} [\omega_{jc} f(y_{ij} | \mathbf{x}_{ij}, \mathbf{v}_{ij}, \boldsymbol{\beta}_c, \delta_c, \mathbf{b}_i)]^{Z_{ijc}} f(\mathbf{b}_i | \mathbf{0}, \mathbf{D}). \quad (6.19)$$

Let's assume that  $y_{ij}$  is assigned to component  $c$  of the  $i$ th cluster,  $Z_{ijc} = 1$ , then, given  $\mathbf{b}$ , we have

$$L(\boldsymbol{\beta}^*, \boldsymbol{\delta}^* | \mathbf{y}, \mathbf{X}, \mathbf{V}, Z, \mathbf{D}, \mathbf{b}) = \prod_{i,j \in \{i,j: Z_{ijc}=1\}} f(y_{ij} | \mathbf{x}_{ij}, \mathbf{v}_{ij}, \boldsymbol{\beta}_c, \delta_c, \mathbf{b}_i), \quad (6.20)$$

for  $c = 1, \dots, k_{max}$ .

The  $kmax$  factors given by (6.20) can be combined with the priors for the parameters  $\boldsymbol{\beta}_c$  and  $\delta_c$ , for  $c = 1, \dots, kmax$ , leading to the its full conditional posterior distribution as follow:

$$\pi(\boldsymbol{\beta}_c | \mathbf{y}, \mathbf{X}, \mathbf{V}, Z, \boldsymbol{\delta}^*, \mathbf{b}, \mathbf{D}) \propto \prod_{i,j,c \in \{i,j,c: Z_{ijc}=1\}} f(y_{ij} | \mathbf{x}_{ij}, \mathbf{v}_{ij}, \boldsymbol{\beta}_c, \delta_c, \mathbf{b}_i) \pi(\boldsymbol{\beta}_c) \quad (6.21)$$

$$\pi(\delta_c | \mathbf{y}, \mathbf{X}, \mathbf{V}, Z, \boldsymbol{\delta}^*, \mathbf{b}, \mathbf{D}) \propto \prod_{i,j,c \in \{i,j,c: Z_{ijc}=1\}} f(y_{ij} | \mathbf{x}_{ij}, \mathbf{v}_{ij}, \boldsymbol{\beta}_c, \delta_c, \mathbf{b}_i) \pi(\delta_c) \quad (6.22)$$

for  $c = 1, \dots, kmax$ .

The full conditional posterior distribution for  $\mathbf{b}_i$  is given by

$$\pi(\mathbf{b}_i | \mathbf{y}, \mathbf{X}, \mathbf{V}, Z, \boldsymbol{\beta}^*, \boldsymbol{\delta}^*, \mathbf{D}) \propto \prod_{j,c \in \{j,c: Z_{ijc}=1\}} f(y_{ij} | \mathbf{x}_{ij}, \mathbf{v}_{ij}, \boldsymbol{\beta}_c, \delta_c, \mathbf{b}_i) f(\mathbf{b}_i | \mathbf{0}, \mathbf{D}) \pi(\mathbf{D}). \quad (6.23)$$

Here, we also consider a Dirichlet prior with parameter vector  $\mathbf{v} = (v_1, \dots, v_{k_j})$  for each vector of the weights  $\boldsymbol{\omega}_j = (\omega_{j1}, \dots, \omega_{jk_j})$ , and Multinomial distribution for the vectors  $Z_{ij}$ , for  $j = 1, \dots, n$  and  $i = 1, \dots, m$ . Thus, the full conditional posterior for  $\boldsymbol{\omega}_j$  is given by

$$\pi(\boldsymbol{\omega}_j | \mathbf{y}, \mathbf{X}, \mathbf{V}, Z, \mathbf{b}, \boldsymbol{\beta}^*, \boldsymbol{\delta}^*, \mathbf{D}) \propto \prod_{c=1}^{k_j} \omega_{jc}^{\eta_{jc} + v_c - 1}, \quad (6.24)$$

where  $\eta_{jc} = \sum_{i=1}^m Z_{ijc}$ . Note that, for each  $j = 1, \dots, n$ ,  $\sum_{c=1}^{k_j} Z_{ijc} = 1$  then we suppose each random vector  $Z_{ij} = (Z_{ij1}, \dots, Z_{ijk_j})$  independently distributed according to a multinomial distribution with parameters 1 and  $\boldsymbol{\omega}_j = (P(Z_{ij1} = 1 | \boldsymbol{\omega}_j, k_j) \dots P(Z_{ijk_j} = 1 | \boldsymbol{\omega}_j, k_j))$ , for  $i = 1, \dots, m$ . Then

$$\begin{aligned} P(Z_{ijc} = 1 | \mathbf{y}, \mathbf{X}, \mathbf{V}, \mathbf{b}, \mathbf{D}, \boldsymbol{\beta}^*, \boldsymbol{\delta}^*, \boldsymbol{\omega}^*) &\propto P(Z_{ijc} = 1 | \boldsymbol{\omega}_j, k_j) f(y_{ij} | Z_{ijc} = 1, \mathbf{x}_{ij}, \mathbf{v}_{ij}, \boldsymbol{\beta}_c, \delta_c, \mathbf{b}_i, \mathbf{D}, \boldsymbol{\omega}_j) \\ &= \omega_{jc} f(y_{ij} | \mathbf{x}_{ij}, \mathbf{v}_{ij}, \boldsymbol{\beta}_c, \delta_c, \mathbf{b}_i, \mathbf{D}), \end{aligned} \quad (6.25)$$

for  $c = 1, \dots, k_j$  and  $i = 1, \dots, m$ .

A summary of the *Gibbs sampling* algorithm for simulate samples from the posterior distribution is given in Algorithm 3. For updating the parameters in  $\boldsymbol{\beta}^*$ ,  $\boldsymbol{\delta}^*$  and  $\mathbf{b}$  we can use a *Metropolis-Hastings* step in the algorithm.

---

**Algorithm 3** – Algorithm for simulate samples from the posterior distribution of the parameters of the mixture of mixed L-Logistic regression models

---

1. Initialize choosing  $\boldsymbol{\omega}_j^{(t)} = \boldsymbol{\omega}_j^{(0)}$ ,  $\boldsymbol{\beta}_c^{(t)} = \boldsymbol{\beta}_c^{(0)}$ ,  $\boldsymbol{\delta}_c^{(t)} = \boldsymbol{\delta}_c^{(0)}$ ,  $\mathbf{b}_i^{(t)} = \mathbf{b}_i^{(0)}$ ,  $D^{(t)} = D^{(0)}$  and  $Z_{ijc}^{(t)} = Z_{ijc}^{(0)}$ , for  $j = 1, \dots, n$ ,  $i = 1, \dots, m$  and  $c = 1, \dots, k_j$ .
  2. For  $t = 0, 1, 2, \dots$  repeat
    - a) For  $j = 1, \dots, n$  do.
      - i. For  $i = 1, \dots, m$  draw  $Z_{ij}^{(t+1)} \sim \text{Multinomial}(1, \pi_{ij1}^{(t)}, \dots, \pi_{ijk_j}^{(t)})$ , wherein
 
$$\pi_{ijc}^{(t)} \propto \omega_{jc} f(y_{ij} | \mathbf{x}_{ij}, \mathbf{v}_{ij}, \boldsymbol{\beta}_c^{(t)}, \boldsymbol{\delta}_c^{(t)}, \mathbf{b}_i, \mathbf{D})$$
    - b) For  $j = 1, \dots, n$  generate  $\boldsymbol{\omega}_j^{(t+1)}$  from the distribution given by (6.24).
    - c) For  $i = 1, \dots, m$  do.
      - i. Generate  $\mathbf{b}_i^{(t+1)}$  from the distribution given by (6.23).
    - d) For  $c = 1, \dots, k_{max}$ 
      - i. Generate  $\boldsymbol{\beta}_c^{(t+1)}$  from the distribution given by (6.21).
      - ii. Generate  $\boldsymbol{\delta}_c^{(t+1)}$  from the distribution given by (6.22)
- 

### **Application of MLLMR model to simulated data**

In this section we present an application to simulated data. Here, we also consider prior independence between the parameters in which  $\beta_{cr} \sim N(0, 100)$ ,  $\delta_c \sim N(0, 100)$  and  $D \sim \text{gamma}(1/20, 1/20)$  with  $b_{0i} \sim N(0, D)$ , for  $r = 0, 1, \dots, q$ ,  $c = 1, \dots, k_j$ ,  $j = 1, \dots, n$  and  $i = 1, \dots, m$ . For the weight of the mixture we consider  $\boldsymbol{\omega}_j \sim \text{Dirichlet}(1, \dots, 1)$ , for  $j = 1, \dots, n$ .

The simulated data were generated from a mixture of L-Logistic distribution with random-effect, such that

$$(Y_{ij} | b_{0i}) \sim \sum_{c=1}^{k_j} \omega_{jc} LL(m_{ijc}, \varphi_c), \quad (6.26)$$

for  $i = 1, \dots, 100$ ,  $j = 1, \dots, 6$  and  $c = 1, \dots, k_j$ , in which

$$\begin{aligned} \text{logit}(m_{ijc}) &= \beta_{c0} + \beta_{c1}x_{ij} + \beta_{c2}t_j + b_{i0}, \\ \log(\varphi_c) &= -\delta_c. \end{aligned} \quad (6.27)$$

The fixed values for  $k_j$ , for  $j = 1, \dots, 6$ , were 2, 3, 3, 3, 2, 2, respectively. The  $x_{ij}$  were generated independently from  $N(0, 1)$ , for  $j = 1, \dots, 5$  and  $i = 1, \dots, 100$ , and the random effects  $b_{0i}$  were generated independently from  $N(0, D)$  with the true  $D = 1$ . Table 18 shows the true values of the regression coefficients and the parameters  $\delta_c$ . Finally, the random vectors of indicators variables in  $Z$  were generated from a multinomial distribution with probabilities given by true values of  $\boldsymbol{\omega}_j$ , also shown in Table 18. In this table, we can see the posterior mean and 95% HPD intervals for the parameters of the MLLMR model from the fitting to the simulated data. The results of the application to simulated data were satisfactory. Figure 18 shows the samples from the posterior distribution of the parameters of the MLLMR model, considering the simulated data, where we

can observe the fast convergence of the algorithm to the real values of the parameters, which are represented for the black line.

Table 18 – Posterior mean and 95% HPD intervals for the parameters of MLLMR model applied to simulated data.

Real value	Posterior mean and 95% HPD intervals		
	Component 1	Component 2	Component 3
$\beta_0 = (-1.0, -2.5, -1.8)$	-0.89(-1.19, -0.60)	-2.47(-2.91, -2.10)	-1.90(-2.47,-1.29)
$\beta_1 = (2.3, 1.0, 1.6)$	2.34(2.21, 2.49)	1.03(0.85, 1.21)	1.59(1.43, 1.74)
$\beta_2 = (1.0, 0.5, 1.9)$	0.94(0.87, 1.01)	0.50(0.41, 0.60)	1.95(1.76, 2.12)
$\delta = (0.7, 0.2, 1.0)$	0.62(0.47, 0.76)	0.13(0.02, 0.24)	0.96(0.74, 1.16)
$\omega_1 = (0.40, 0.60)$	0.47 (0.31, 0.62)	0.53(0.38, 0.69)	-
$\omega_2 = (0.45, 0.25, 0.30)$	0.54 (0.34, 0.75)	0.21 (0.12, 0.33)	0.24 (0.06, 0.42)
$\omega_3 = (0.35, 0.32, 0.33)$	0.25 (0.11, 0.40)	0.32 (0.21, 0.42)	0.43 (0.28, 0.57)
$\omega_4 = (0.40, 0.20, 0.40)$	0.44 (0.32, 0.56)	0.22 (0.13, 0.32)	0.34 (0.23, 0.44)
$\omega_5 = (0.50, 0.50)$	0.57 (0.45, 0.68)	0.43 (0.32, 0.55)	-
$\omega_6 = (0.30, 0.70)$	0.29 (0.19, 0.39)	0.71 (0.61, 0.81)	-
$D = 1$	0.91(0.59, 1.26)		

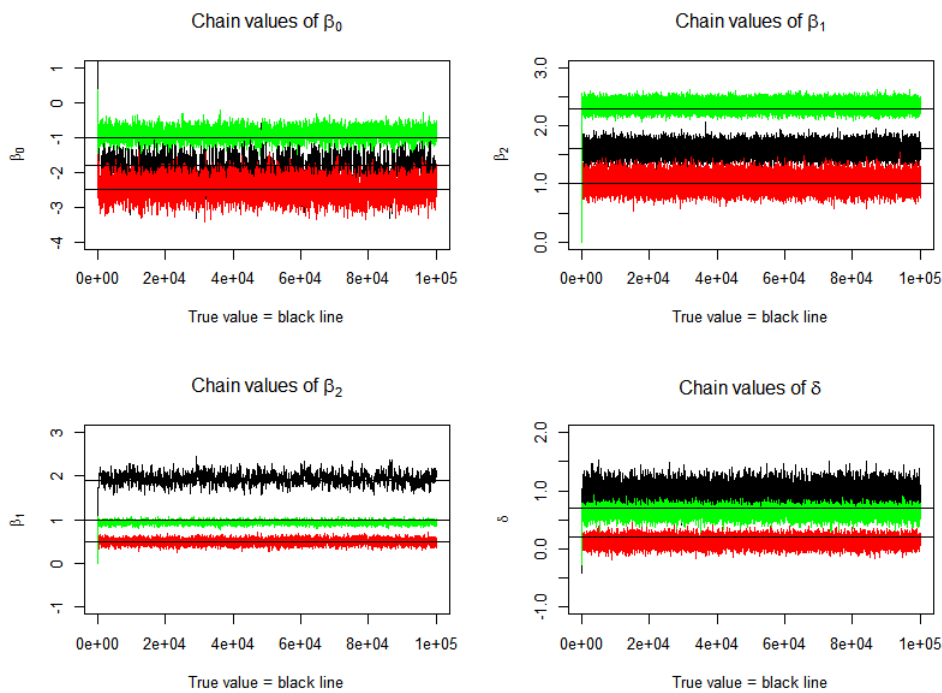


Figure 18 – Chais values for the parameters of MLLMR model considering the simulated data where the values of the parameters of componente 1 are in green, values of the parameters of component 2 are in black and the values of parameters of component 3 are in red.

### Application of MLLMR model to real data

The MLLMR model was applied to real data presented in the Chapter 2, where we consider the votes obtained by the Workers' Party in five presidential elections from 1994 to 2010 in Brazil. In the Chapter 2, these data sets were analyzed individually. Here, we consider

that this data can be modeled as a longitudinal data. Thus, we fit the MLLMR model to all data set presented in the Chapter 2, considering the proportion of the votes as a response. The model considered was:

$$\begin{aligned} (Y_{ij}|b_{0i}) &\sim \sum_{c=1}^{k_j} \omega_{jc} LL(m_{ijc}, \varphi_c), \\ \text{logit}(m_{ijc}) &= \beta_{c0} + \beta_{c1}t_j + b_{i0}, \\ \log(\varphi_c) &= -\delta_c, \end{aligned} \quad (6.28)$$

for  $i = 1, \dots, 75, j = 1, \dots, 5$  and  $c = 1, \dots, k_j$  with  $k_j \in \{1, 2, 2, 2, 1\}$ . The time  $t$  is relate to the year of election, being the time 1 the first election considered, year 1994, and so on.

The posterior mean and 95% HPD intervals for the parameters of the model are presented in Table 19. For this analysis, we consider 1, 2, 2, 2 and 1 components for the data of the years 1994, 1998, 2002, 2006 and 2010, respectively. However, Table 19 shows the estimated values for the parameters vector  $\omega_{1,2}$   $\omega_{1,3}$  and  $\omega_{1,4}$  very close to zero, giving evidence that the mixture may has disappeared with the inclusion of the time covariate. Figure 19 shows the samples of posterior distribution for data of percentage of votes.

Table 19 – Posterior mean and 95% HPD intervals for the parameters of MLLMR model applied to data of votes percentage.

Parameters	Posterior mean and 95% HPD intervals	
	Component 1	Component 2
$\beta_0$	-1.79 (-1.92, -1.66)	-2.81(-2.99, -2.60)
$\beta_1$	0.35 (0.32, 0.38)	0.65 (0.59, 0.71)
$\delta$	1.48 (1.33, 1.61)	1.43 (1.32, 1.52)
$\omega_1$	-	-
$\omega_2$	0.026(0.001,0.060)	0.974 (0.940, 0.999)
$\omega_3$	0.026 (0.001 0.060)	0.974 (0.940, 0.100)
$\omega_4$	0.026 (0.001, 0.061)	0.974(0.940, 0.100)
$\omega_5$	-	-
$D$	0.043 (0.018, 0.072)	

## 6.6 Remarks

In this chapter we develop a extension of the L-Logistic regression (LLR) model in order to include random effect by considering L-Logistic mixed regression (LLMR) and Mixture of L-Logistic mixed regression (MLLMR) models. Applications of the models are still in development, however we can see that this type of models can be promissory.

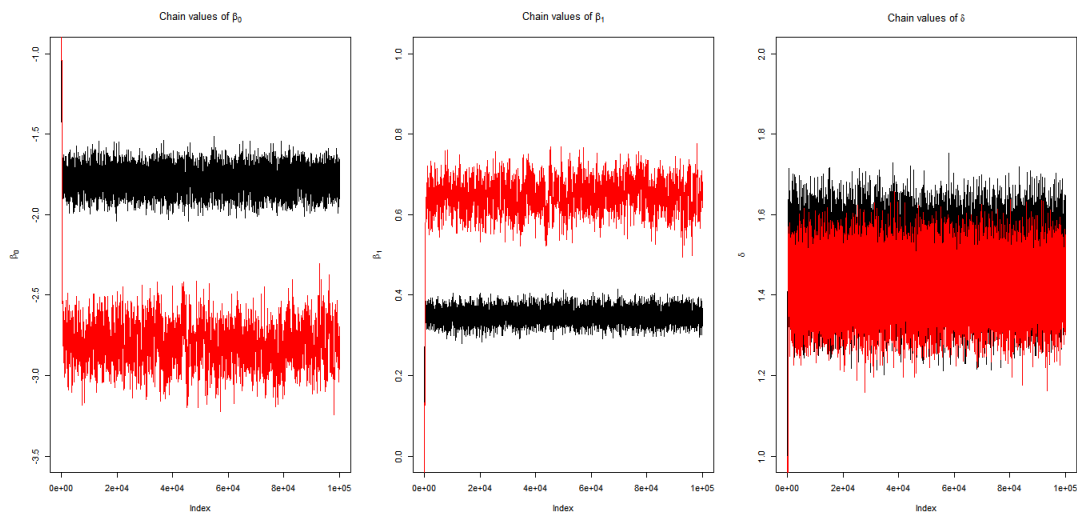


Figure 19 – Chain values for the parameters of MLLMR model considering the data of percentage of votes, where the values of the parameters of component 1 are in black and values of the parameters of component 2 are in red.





---

# CONTRIBUTIONS AND FUTURE DEVELOPMENTS

---

---

This Chapter is dedicated to give a description of the contributions of the thesis, and future development.

---

## 7.1 Contributions

In this thesis, we investigate two bounded continuous distributions addressing some features in regression analysis that not were studied in the literature, which are particularly important in the statistical analysis. These development were motivated by the data analysis presented in the Chapter 2. The Chapter 3 presents the first bounded distribution investigated called mixture of Simplex distributions that can be seen as a generalization of the Simplex distribution. If the mixture of Simplex distributions have just one component then we have a single Simplex distribution. We present the inference in the mixture of Simplex distribution under Bayesian approach for the general case where the number of component in the mixture model is unknown. An analysis to simulated data sets from a mixture of Simplex distributions with 2 and 3 components were conducted. For these applications, we found that the method provides a good estimation of the number of components, as well as of the other parameters of the model since the estimated values lie close to the real values of the parameters. In the Chapter 4 we generalize this model introducing covariates obtained a mixture of Simplex regression model, and then we apply it to data where the response variable is a proportion. The second bounded continuous distribution investigated in Chapter 5 and 6 is the L-Logistic distribution. The parameterization of this distribution presented in this work has interesting properties that can be useful to model data in the unit interval since covariates can be included in the model as a function of median, which

is a parameter of this distribution. A R package was included in the Specifically in the Chapter 6, we explore a mixture of L-Logistic distribution with mixed-effect for the longitudinal data. For inference, for all models treated in this thesis, we adopt a Bayesian approach using a Markov chain Monte Carlo (MCMC) method for the modeling framework. The issues of model fitting are addressed by means of a hybrid algorithm that combines Metropolis-Hasting with Gibbs sampling algorithms implemented in R language. The functions that provide the probability density function, cumulative density function, quantile function and random generation for the L-Logistic distribution with parameters  $m$  and  $b$  are available in the CRAN of R program.

Based on description given before, we can cite the following contributions:

- development of mixture models for bounded response,
- development of regression models for limited response, i.e., a mixture of Simplex regression models, the L-Logistic regression model, the L-Logistic regression model with random effect, and the mixture of L-Logistic regression model with random effect.
- Bayesian inference for the models proposed in the different chapters of this thesis, including estimation, model comparison criteria and residual analysis,
- R codes to estimation, convergence analysis, posterior analysis and inference which must be available online, and provision of the "llogistic" package.

## 7.2 Future development

For future, the maximum likelihood estimation for the models presented in this thesis can be developed. In addition, a *Reversible-jump* algorithm can be implemented to estimate a number of component for a mixture of L-Logistic distribution when the number of component in the mixture is unknown. Additionally, some other extensions for this work can include diagnostics and residual analysis for all regression models treated here. The Package in R of the L-Logistic distribution contain just basic functions and can be improved. Relate to the Chapter 2, we intend to develop a distribution from the Weibull distribution to bounded data through some transformation. Specifically, we intend to apply the resulting distribution from this transformation to political data analyzed in the Chapter 2.

## BIBLIOGRAPHY

---

---

AL-AWADHI, F.; HURN, M.; JENNISON, C. Improving the acceptance rate of reversible jump MCMC proposals. **Statistics and Probability Letters**, v. 69, n. 2, p. 189–198, 2004. Citation on page [53](#).

ARNOLD, B. C.; GROENEVELD, R. A. Measuring skewness with respect to the mode. **The American Statistician**, Taylor & Francis, Ltd. on behalf of the American Statistical Association, v. 49, n. 1, p. 34–38, 1995. Citation on page [71](#).

ATKINSON, K. **An introduction to numerical analysis**. [S.l.]: Wiley, 2008. Citation on page [37](#).

BARNDORFF-NIELSEN, O.; JORGENSEN, B. Some parametric models on the simplex. **Journal of Multivariate Analysis**, v. 39, n. 1, p. 106 – 116, 1991. Citations on pages [23](#) e [42](#).

BAYES, C.; BAZÁN, J. L.; GARCÍA, C. A new robust regression model for proportions. **Statistics and Its Interface**, v. 7, n. 4, p. 841–866, 2012. Citation on page [79](#).

BAYES, C. L.; BAZÁN, J. L.; CASTRO, M. de. A quantile parametric mixed regression model for bounded response variables. **Statistics and Its Interface**, v. 10, n. 3, p. 483–493, 2017. Citation on page [95](#).

BAYES, L. C.; BAZÁN, J. L.; CASTRO, M. A quantile parametric mixed regression model for bounded response variables. **Statistics and Its Interface**, v. 10, n. 3, p. 483–493, 2017. Citations on pages [23](#) e [66](#).

BERDUFI, D. Statistical detection of vote count fraud (2009 albanian parliamentary election and benford's law). **Mediterranean Journal of Social Sciences**, v. 5, p. 755–772, 2014. Citations on pages [24](#) e [29](#).

BERKHOF, J.; MECHELEN, I. V.; GELMAN, A. A bayesian approach to the selection and testing of mixture models. v. 13, p. 423–442, 2003. Citation on page [35](#).

BOHN, S. R. Social policy and vote in brazil bolsa familia and the shifts in lula's electoral base. v. 46, p. 54–79, 2011. Citation on page [28](#).

BONAT, W. H.; RIBEIRO, P. J.; ZEVIANI, W. M. Likelihood analysis for a class of beta mixed models. **Journal of Applied Statistics**, v. 42, n. 2, p. 252–266, 2015. Citation on page [90](#).

BOUGUILA, N.; ELGUEBALY, T. A fully Bayesian model based on reversible jump MCMC and finite beta mixtures for clustering. **Expert Syst. Appl.**, v. 39, n. 5, p. 5946–5959, 2012. Citations on pages [23](#), [32](#), [42](#), [45](#), [46](#), [48](#) e [57](#).

BOUGUILA, N.; ZIOU, D.; MONGA, E. Practical Bayesian estimation of a finite beta mixture through gibbs sampling and its applications. **Statistics and Computing**, v. 16, n. 2, p. 215–225, 2006. Citations on pages [23](#) e [42](#).

- BRYS, G.; HUBERT, M.; STRUYF, A. A comparison of some new measures of skewness. In: DUTTER, R.; FILZMOSER, P.; GATHER, U.; ROUSSEEUW, P. J. (Ed.). **Developments in Robust Statistics**. [S.l.]: Springer-Verlag, 2003. p. 98–113. Citation on page 71.
- BUCKLEY, J. Estimation of models with beta-distributed dependent variables: A replication and extension of paolino's study. **Political Analysis**, Cambridge University Press, v. 11, n. 2, p. 204–205, 2003. Citations on pages 23 e 66.
- CAI, Y.; JIANG, T. Estimation of non-crossing quantile regression curves. **Australian and New Zealand Journal of Statistics**, v. 57, n. 1, p. 139–162, 2015. Citation on page 88.
- CHEN, M.; SHAO, Q.; IBRAHIM, J. **Monte Carlo Methods in Bayesian Computation**. [S.l.]: Springer New York, 2000. Citation on page 35.
- CHIB, S.; JELIAZKOV, I. Marginal likelihood from the metropolis-hastings output. **Journal of the American Statistical Association**, v. 96, p. 270–281, 2001. Citations on pages 35 e 59.
- CIFUENTES, M.; SEMBAJWE, G.; TAK, S.; GORE, R.; KRIEBEL, D.; PUNNETT, L. The association of major depressive episodes with income inequality and the human development index. **Social Science and Medicine**, v. 67, n. 4, p. 529 – 539, 2008. Citations on pages 23 e 42.
- CUFF, V.; LEWIS, A.; MILLER, S. The weibull distribution and benford's law. **MSP**, v. 8, n. 5, p. 859–874, 2014. Citations on pages 24 e 29.
- CUFF, V.; LEWIS, A.; MILLER, S. J. The effective use of benford's law to assist in detecting fraud in accounting data. **Journal of Forensic Accounting**, v. 5, p. 17–34, 2014. Citations on pages 24 e 29.
- DIEBOLT, J.; ROBERT, C. P. Estimation of finite mixture distributions through Bayesian sampling. **Journal of the Royal Statistical Society. Series B**, Blackwell Publishing for the Royal Statistical Society, v. 56, n. 2, p. 363–375, 1994. Citation on page 42.
- ESCOBAR, M. D.; WEST, M. Bayesian density estimation and inference using mixtures. **Journal of the American Statistical Association**, v. 90, p. 577–588, 1995. Citation on page 36.
- ESTATÍSTICA, D. d. E. e. R. Fundação Instituto Brasileiro de Geografia e. **Pesquisa nacional por amostra de domicílios, PNAD.: Síntese de indicadores**. [S.l.]: IBGE, 2014. Citation on page 58.
- FARIA, S.; GONÇALVES, F. Financial data modeling by Poisson mixture regression. **Journal of Applied Statistics**, Taylor & Francis, v. 40, n. 10, p. 2150–2162, 2013. Citation on page 42.
- FERRARI, S.; CRIBARI-NETO, F. Beta regression for modelling rates and proportions. **Journal of Applied Statistics**, Taylor & Francis, v. 31, n. 7, p. 799–815, 2004. Citations on pages 23, 66, 82 e 84.
- FEWSTER, R. M. A simple explanation of benford's law. **The American Statistician**, v. 63, n. 1, p. 26–32, 2009. Citation on page 29.
- Figuroa-Zúñiga, J. I.; Arellano-Valle, R. B.; FERRARI, S. L. P. Mixed beta regression: A Bayesian perspective. **Computational Statistics & Data Analysis**, Elsevier Science Publishers B. V., v. 61, n. 1, p. 137–147, 2013. Citations on pages 77 e 89.

GELFAND, A. E.; SMITH, A. F. M. Sampling-based approaches to calculating marginal densities. **Journal of the American Statistical Association**, v. 85, n. 410, p. 398–409, 1990. Citation on page 42.

GELMAN, A.; CARLIN, J.; STERN, H.; DUNSON, D.; VEHTARI, A.; RUBIN, D. **Bayesian Data Analysis**. Third edition. Philadelphia: Taylor & Francis, 2013. (Chapman & Hall/CRC Texts in Statistical Science). Citations on pages 74, 75, 76 e 86.

GEWEKE, J. Evaluating the accuracy of sampling-based approaches to calculating posterior moments. In: BERNARDO, J. M.; BERGER, J.; DAWID, A. P.; SMITH, J. F. M. (Ed.). **Bayesian Statistics**. [S.l.]: Oxford University Press, 1992. p. 169–193. Citations on pages 37 e 85.

GOLDFELD, S. M.; QUANDT, R. E. A markov model for switching regressions. **Journal of Econometrics**, v. 1, n. 1, p. 3 – 15, 1973. Citation on page 56.

GÓMEZ-DÉNIZ, E.; SORDO, M. A.; CALDERÍN-OJEDA, E. The log-lindley distribution as an alternative to the beta regression model with applications in insurance. **Insurance: Mathematics and Economics**, v. 54, n. 1, p. 49–57, 2014. Citations on pages 23 e 66.

GREEN, P. J. Reversible jump Markov chain Monte Carlo computation and Bayesian model determination. **Biometrika**, v. 82, p. 711–732, 1995. Citations on pages 43, 46 e 47.

GREEN, P. J.; MIRA, A. Delayed rejection in reversible jump Metropolis-Hastings. **Biometrika**, v. 88, p. 1035–1053, 2001. Citation on page 53.

GUPTA, A.; NADARAJAH, S. **Handbook of Beta Distribution and Its Applications**. Philadelphia: Taylor & Francis, 2004. Citation on page 82.

HAO, L.; NAIMAN, D. **Quantile Regression**. New Jersey: SAGE Publications, 2007. Citation on page 83.

HOLMES, C. C.; JASRA, A.; STEPHENS, D. A. Markov chain Monte Carlo methods and the label switching problem in Bayesian mixture modeling. **Statistical Science**, v. 20, n. 1, p. 50–67, 2005. Citation on page 33.

HURN, M.; JUSTEL, A.; ROBERT, C. P. Estimating mixtures of regressions. **Journal of Computational and Graphical Statistics**, v. 12, n. 1, p. 55–79, 2003. Citation on page 56.

JOHNSON, N.; KOTZ, S.; BALAKRISHNAN, N. **Continuous Univariate Distributions**. Second. New York: John Wiley & Sons, 1994. Citation on page 70.

JOHNSON, N. L. Systems of frequency curves generated by methods of translation. **Biometrika**, Biometrika Trust, v. 36, n. 1/2, p. 149–176, 1949. Citations on pages 66 e 70.

JOHNSON, N. L.; TADIKAMALLA, P. R. Translated family of distribution. In: BALAKRISHNAN, N. (Ed.). **Handbook of the Logistic Distribution**. [S.l.]: Taylor & Francis, 1991, (Statistics: A Series of Textbooks and Monographs). chap. 8, p. 189–208. Citations on pages 24 e 66.

JONES, K.; JOHNSTON, R. People, places and regions: Exploring the use of multi-level modelling in the analysis of electoral data. **British Journal of Political Science**, v. 22, p. 323–360, 1992. Citation on page 28.

JONES, M. Kumaraswamy's distribution: A beta-type distribution with some tractability advantages. **Statistical Methodology**, v. 6, n. 1, p. 70 – 81, 2009. Citation on page 23.

JØRGENSEN, B. **The Theory of Dispersion Models**. [S.l.]: Taylor & Francis, 1997. (Monographs on Statistics and Applied Probability). Citation on page 44.

KASS, R. E.; RAFTERY, A. E. Bayes factors. **Journal of the american statistical association**, v. 90, p. 773–795, 1995. Citation on page 37.

KOENKER, R.; BASSETT, J. G. Regression quantiles. **Econometrica**, JSTOR, v. 46, n. 1, p. 33–50, 1978. Citation on page 66.

LEMONTE, A. J.; BAZÁN, J. L. New class of johnson sb distributions and its associated regression model for rates and proportions. **Biometrical Journal**, v. 58, n. 4, p. 727–746, 2016. ISSN 1521-4036. Citations on pages 23 e 66.

LEMONTE, A. J.; BAZÁN, J. L. New class of johnson sb distributions and its associated regression model for rates and proportions. **Biometrical Journal**, v. 58, n. 4, p. 727–746, 2016. Citation on page 87.

LÓPEZ, F. O. A bayesian approach to parameter estimation in simplex regression model: A comparison with beta regression. **Revista Colombiana de Estadística**, v. 36, n. 1, p. 1–21, 2013. Citations on pages 23, 24 e 42.

MARIN, J.-M.; MENGERSEN, K. L.; ROBERT, C. Bayesian modelling and inference on mixtures of distributions. In: DEY, D.; RAO, C. (Ed.). **Handbook of Statistics: Volume 25**. [S.l.]: Elsevier, 2005. Citations on pages 31 e 34.

MARTIN, A. D.; QUINN, K. M.; PARK, J. H. MCMCpack: Markov chain Monte Carlo in R. **Journal of Statistical Software**, v. 42, n. 9, p. 22, 2011. Citations on pages 39, 61 e 78.

MCDONALD, J.; RANSOM, M. The generalized beta distribution as a model for the distribution of income: Estimation of related measures of inequality. In: **Modeling Income Distributions and Lorenz Curves**. [S.l.]: Springer New York, 2008, (Economic Studies in Equality, Social Exclusion and Well-Being, v. 5). p. 147–166. Citations on pages 23 e 42.

MCLACHLAN, G.; PEEL, D. **Finite Mixture Models**. [S.l.]: Wiley, 2004. (Wiley series in probability and statistics: Applied probability and statistics). Citations on pages 30, 42, 43 e 53.

MENEGUELLO, R. Electoral behaviour in brazil; the 1994 presidential elections. **International social science journal**, p. 627–641, 1995. Citation on page 28.

MOORS, J. J. A. A quantile alternative for kurtosis. **Journal of the Royal Statistical Society, Series B.**, v. 37, n. 1, p. 25–32, 1988. Citation on page 72.

PAZ, R. F.; BAZAN, J. L. **llogistic: The L-Logistic Distribution**. [S.l.], 2017. R package version 1.0.0. Available: <<https://CRAN.R-project.org/package=llogistic>>. Citations on pages 69 e 91.

PAZ, R. F.; BAZÁN, J. L.; ELHER, R. A Weibull mixture model for the votes of a Brazilian political party. In: **EBEB: Interdisciplinary Bayesian Statistics**. [S.l.]: Springer, 2015, (Springer Proceedings in Mathematics and Statistics, v. 118). p. 229–241. Citations on pages 24, 42 e 59.

PAZ, R. F.; BAZÁN, J. L.; MILAN, L. A. Bayesian estimation for a mixture of simplex distributions with an unknown number of components: Hdi analysis in brazil. **Journal of Applied Statistics**, v. 0, n. 0, p. 1–14, 2015. Citation on page 24.

PAZ, R. F.; BAZÁN, J. L.; MILAN, L. A. Bayesian estimation for a mixture of simplex distributions with an unknown number of components: Hdi analysis in brazil. **Journal of Applied Statistics**, v. 0, n. 0, p. 1–14, 2015. Citation on page 66.

PNUD; IPEA; FJP. **Atlas do Desenvolvimento Humano no Brasil**. Brasilia, Brazil: PNUD, 2013. Disponíble in: <http://www.atlasbrasil.org.br/2013/pt/>. Citations on pages 42, 51 e 82.

R Development Core Team. **R: A Language and Environment for Statistical Computing**. Vienna, Austria, 2016. ISBN 3-900051-07-0. Available: [<http://www.r-project.org/>](http://www.r-project.org/). Citations on pages 37, 48 e 74.

RICHARDSON, S.; GREEN, P. J. On bayesian analysis of mixtures with an unknown number of components (with discussion). **Journal of the Royal Statistical Society: Series B (Statistical Methodology)**, v. 59, n. 4, p. 731–792, 1997. Citations on pages 34, 42, 43, 46 e 47.

ROBERT, C. P.; CASELLA, G. **Monte Carlo Statistical Methods (Springer Texts in Statistics)**. [S.l.]: Springer-Verlag New York, Inc., 2005. Citation on page 33.

ROSS, S. M. **Simulation, Fourth Edition**. Orlando, FL, USA: Academic Press, Inc., 2006. ISBN 0125980639. Citation on page 46.

SMITHSON, M.; VERKUILEN, J. A better lemon squeezer? maximum-likelihood regression with beta-distributed dependent variables. **Psychological Methods**, American Psychological Association, v. 11, n. 1, p. 54, 2006. Citations on pages 67, 84, 86 e 88.

SONG, P. X.-K.; TAN, M. A marginal models for longitudinal continuous proportional data. **Biometrics**, v. 56, n. 2, p. 496–502, 2000. Citations on pages 23, 24 e 42.

SPIEGELHALTER, D. J.; BEST, N. G.; CARLIN, B. P.; LINDE, A. van der. Bayesian measures of model complexity and fit. p. 583–639, 2002. Citation on page 59.

STENSHOLT, E. Beta distributions in a simplex and impartial anonymous cultures. **Mathematical Social Sciences**, v. 37, n. 1, p. 45–57, 1999. Citations on pages 23 e 42.

STEPHENS, M. Bayesian analysis of mixture models with an unknown number of components - an alternative to reversible jump methods. **The Annals of Statistics**, The Institute of Mathematical Statistics, v. 28, n. 1, p. 40–74, 2000. Citation on page 34.

TADIKAMALLA, P. R.; JOHNSON, N. L. Systems of frequency curves generated by transformations of logistic variables. **Biometrika**, v. 69, n. 2, p. 461, 1982. Citations on pages 24, 65, 66, 69, 70, 87 e 90.

\_\_\_\_\_. Tables to facilitate fitting tadikamalla and johnson's lb distributions. **Communications in Statistics - Simulation and Computation**, v. 19, n. 4, p. 1201–1229, 1990. Citations on pages 24 e 66.

TANNER, M. A.; WONG, W. H. The calculation of posterior distributions by data augmentation. **Journal of the American Statistical Association**, v. 82, n. 398, p. 528–540, 1987. Citation on page 42.

TSIONAS, E. G. Bayesian analysis of finite mixtures of weibull distributions. **Communications in Statistics - Theory and Methods**, v. 31, n. 1, p. 37–48, 2002. Citation on page [30](#).

VIELE, K.; TONG, B. Modeling with mixtures of linear regressions. **Statistics and Computing**, v. 12, n. 4, p. 315–330, 2002. Citation on page [58](#).

WANG, M.; RENNOLLS, K. Tree diameter distribution modelling: introducing the logit-logistic distribution. **Canadian Journal of Forest Research**, v. 35, n. 6, p. 1305–1313, 2005. Citations on pages [66](#) e [69](#).



---

## PROCEDURE FOR SIMULATE SAMPLE FROM A MIXTURE OF SIMPLEX DISTRIBUTIONS WITH UNKNOWN NUMBER OF COMPONENT

---

---

This appendix address the summarize the MCMC technique mentioned in the Chapter 3 by giving a description of the Gibbs sampling algorithm used to sample from the joint probability distribution. The *Gibbs sampling* algorithm is used combined with *Metropolis-Hastings* (MR and reversible-jump) algorithm for obtain the sample of the posterior distribution of parameters  $(\boldsymbol{\theta}, \boldsymbol{\omega}, k)$  and  $Z_i$ , for  $i = 1, \dots, N$ . The Algorithm 4 show a scheme of the procedure.

---

**Algorithm 4** – Algorithm for simulate samples from the posterior joint probability distribution of the parameters of mixture of simplex

1. Initialize choosing  $k^{(t)} = k^{(0)}$ ,  $\boldsymbol{\omega}^{(t)} = \boldsymbol{\omega}^{(0)}$ ,  $\boldsymbol{\mu}_j^{(t)} = \boldsymbol{\mu}_j^{(0)}$ ,  $(\sigma_j^2)^{(t)} = (\sigma_j^2)^{(0)}$  and  $Z_{ij}^{(t)} = Z_{ij}^{(0)}$ , for  $i = 1, \dots, n$   $j = 1, \dots, k^{(t)}$ .
2. For  $t = 0, 1, 2, \dots$  repeat
  - a) For  $i = 1, \dots, n$  draw  $Z_i^{(t+1)} \sim \text{Multinomial}(1, \boldsymbol{\pi}_{i1}^{(t)}, \dots, \boldsymbol{\pi}_{ik^{(t)}}^{(t)})$ , wherein
$$\boldsymbol{\pi}_{ij}^{(t)} = P(Z_{ij}^{(t)} = 1 | y_i, \boldsymbol{\mu}_j^{(t)}, (\sigma_j^2)^{(t)}) \propto \omega_j \mathcal{S}(y_i | \boldsymbol{\mu}_j^{(t)}, (\sigma_j^2)^{(t)})$$
  - b) Generate  $\boldsymbol{\omega}^{(t+1)}$  from the distribution given by (3.10).
  - c) For  $j = 1, \dots, k^{(t)}$  do
    - i. Generate  $\phi_j^{(t+1)}$  from the distribution given by (3.8) and do  $(\sigma_j^2)^{(t+1)} = 1/\phi_j^{(t+1)}$ .
    - ii. For updating  $\boldsymbol{\mu}_j$  a *Metropolis-hastings* step is done, then
      - generate  $\boldsymbol{\mu}'_j \sim \text{Beta}(\boldsymbol{\delta}^{(t)}, \boldsymbol{\eta}^{(t)})$  where  $(\boldsymbol{\delta}^{(t)}, \boldsymbol{\eta}^{(t)})$  is computed in A.
      - Compute
$$\alpha(\boldsymbol{\mu}_j^{(t)}, \boldsymbol{\mu}'_j) = \min \left\{ 1, \frac{P(\boldsymbol{\mu}'_j | \mathbf{y}, Z, (\sigma_j^2)^{(t+1)}) \text{Be}(\boldsymbol{\mu}_j^{(t)} | \boldsymbol{\delta}^{(t+1)}, \boldsymbol{\eta}^{(t+1)})}{P(\boldsymbol{\mu}_j^{(t)} | \mathbf{y}, Z, (\sigma_j^2)^{(t+1)}) \text{Be}(\boldsymbol{\mu}'_j | \boldsymbol{\delta}^{(t)}, \boldsymbol{\eta}^{(t)})} \right\}$$
where  $\text{Be}(x|\cdot)$  is the density of beta distribution evaluated at  $x$ .
      - Generate  $u \sim \text{Uniform}(0, 1)$
      - If  $\alpha(\boldsymbol{\mu}_j^{(t)}, \boldsymbol{\mu}'_j) < u$  then  $\boldsymbol{\mu}_j^{(t+1)} = \boldsymbol{\mu}'_j$  else  $\boldsymbol{\mu}_j^{(t+1)} = \boldsymbol{\mu}_j^{(t)}$ .
  - d) For updating  $k^{(t)}$ , merge two component of the mixture into one or splinting one into two by using reversible-jump step.

## Proposal distribution

We observed in the simulation process that the convergence of algorithm *Gibbs sampling* is affected by the acceptance of MR step in (2(c)ii). In order to improve the acceptance rate of  $\boldsymbol{\mu}'_j$  ( $j = 1, \dots, k^{(t)}$  and  $t = 0, 1, 2, \dots$ ), in the *Metropolis-Hastings* step (2(c)ii) of the algorithm, we adopt a beta distribution as the proposal distribution where the parameters  $\boldsymbol{\delta}^{(t)}$  and  $\boldsymbol{\eta}^{(t)}$  are obtained by solving

$$\begin{cases} \boldsymbol{\mu}_j^{(t)} = \frac{\boldsymbol{\delta}^{(t)}}{\boldsymbol{\delta}^{(t)} + \boldsymbol{\eta}^{(t)}} \\ \boldsymbol{\psi}^{(t)} = \frac{\boldsymbol{\delta}^{(t)} \boldsymbol{\eta}^{(t)}}{(\boldsymbol{\delta}^{(t)} + \boldsymbol{\eta}^{(t)})^2 (\boldsymbol{\delta}^{(t)} + \boldsymbol{\eta}^{(t)} + 1)} \end{cases} \quad (\text{A.1})$$

where  $\mu_j^{(t)}$  and  $\psi^{(t)}$  is the mean and variance of beta distribution with parameters  $\delta^{(t)} > 0$  and  $\delta^{(t)} > 0$ . Then we have

$$\begin{cases} \eta^{(t)} = \delta^{(t)} \left( \frac{1}{\mu_j^{(t)}} - 1 \right) \\ \delta^{(t)} = (\mu_j^{(t)})^2 \left( \frac{1 - \mu_j^{(t)}}{\psi^{(t)}} - \frac{1}{\mu_j^{(t)}} \right) \end{cases} \quad (\text{A.2})$$

The positivity of  $\delta^{(t)}$  and  $\delta^{(t)}$  is secured by making  $\psi^{(t)} < \mu^{(t)}(1 - \mu^{(t)})$  leading to  $\psi^{(t)} = \mu_j^{(t)}(1 - \mu_j^{(t)}) \times \tau$ , with  $0 < \tau < 1$ .



---

## PROOFS OF PROPERTIES OF THE L-LOGISTIC AND RESULTS FOR PRIOR SENSITIVITY ANALYSIS

---

This chapter provide proofs of the properties of L-logistic distribution presented in Chapter 5.3. Moreover, results of the preliminary study of the prior sensitivity analysis in Chapter 5.5 are presented and commented.

---

### Proofs of properties of the L-logistic in Section 5.3

For all cases, let  $Y \sim LL(m, b)$ , with pdf and cdf give by

$$f(y|m, b) = \frac{b(1-m)^b m^b y^{b-1} (1-y)^{b-1}}{[(1-m)^b y^b + m^b (1-y)^b]^2}, \quad 0 < y < 1, 0 < m < 1, b > 0. \quad (\text{B.1})$$

and

$$F_Y(y|m, b) = \left( 1 + \left( \frac{m(1-y)}{y(1-m)} \right)^b \right)^{-1}, \quad 0 < y < 1, \quad (\text{B.2})$$

### Proof of Property 5.3.1

Consider the following transformation  $Z = b \log \left( \frac{Y(1-m)}{m(1-Y)} \right)$ . Then the pdf of  $Z$  is given by

$$\begin{aligned}
 f_Z(z) &= f_Y \left( \left( 1 + e^{-z/b} \left( \frac{1-m}{m} \right) \right)^{-1} \right) \left| \frac{\partial \left( 1 + e^{-z/b} \left( \frac{1-m}{m} \right) \right)^{-1}}{\partial z} \right| \\
 &= \frac{b(1-m)^b m^b \left( \frac{1}{1 + \left( \frac{1-m}{m} \right) e^{-z/b}} \right)^{b-1} \left( \frac{\left( \frac{1-m}{m} \right) e^{-z/b}}{1 + \left( \frac{1-m}{m} \right) e^{-z/b}} \right)^{b-1}}{\left[ (1-m)^b \left( \frac{1}{1 + \left( \frac{1-m}{m} \right) e^{-z/b}} \right)^b + m^b \left( \frac{\left( \frac{1-m}{m} \right) e^{-z/b}}{1 + \left( \frac{1-m}{m} \right) e^{-z/b}} \right)^b \right]^2} \left( \frac{\left( \frac{1-m}{m} \right) e^{-z/b}}{b \left[ 1 + \left( \frac{1-m}{m} \right) e^{-z/b} \right]^2} \right) \\
 &= \frac{(1-m)^b m^b \left( \left( \frac{1-m}{m} \right) e^{-z/b} \right)^{b-1}}{\left[ (1-m)^b + m^b \left( \frac{1-m}{m} \right)^b e^{-z} \right]^2} \left( \left( \frac{1-m}{m} \right) e^{-z/b} \right) \\
 &= \frac{(1-m)^b m^b (1-m)^{b-1} e^{-z(b-1)/b}}{m^{b-1} \left[ (1-m)^b + (1-m)^b e^{-z} \right]^2} \left( \frac{1-m}{m} \right) e^{-z/b} = \frac{e^{-z}}{[1+e^{-z}]^2} \\
 &\Rightarrow f_Z(z) = \frac{e^z}{[1+e^z]^2} I_{\mathbb{R}}(z),
 \end{aligned}$$

which is the corresponding pdf of the standard logistic distribution.

### Proof of Property 5.3.2

Consider the reparameterization  $m = \frac{e^{-\frac{\delta}{b}}}{1+e^{-\frac{\delta}{b}}}$  with  $\delta > 0$ . Then the pdf and cdf of the L-logistic distribution with parameter vector  $(\delta, b)$  is obtained as

$$\begin{aligned}
 f(y|\delta, b) &= \frac{b \left( 1 - \frac{e^{-\frac{\delta}{b}}}{1+e^{-\frac{\delta}{b}}} \right)^b \left( \frac{e^{-\frac{\delta}{b}}}{1+e^{-\frac{\delta}{b}}} \right)^b y^{b-1} (1-y)^{b-1}}{\left[ \left( 1 - \frac{e^{-\frac{\delta}{b}}}{1+e^{-\frac{\delta}{b}}} \right)^b y^b + \left( \frac{e^{-\frac{\delta}{b}}}{1+e^{-\frac{\delta}{b}}} \right)^b (1-y)^b \right]^2} \\
 &= \frac{b \left( \frac{1}{1+e^{-\frac{\delta}{b}}} \right)^b \left( \frac{e^{-\frac{\delta}{b}}}{1+e^{-\frac{\delta}{b}}} \right)^b y^{b-1} (1-y)^{b-1}}{\left[ \left( \frac{1}{1+e^{-\frac{\delta}{b}}} \right)^b y^b + \left( \frac{e^{-\frac{\delta}{b}}}{1+e^{-\frac{\delta}{b}}} \right)^b (1-y)^b \right]^2} \\
 &= \frac{be^{-\delta} y^{b-1} (1-y)^{b-1}}{[y^b + e^{-\delta} (1-y)^b]^2} = \frac{be^{\delta} y^{b-1} (1-y)^{b-1}}{[y^b e^{\delta} + (1-y)^b]^2}
 \end{aligned}$$

and

$$\begin{aligned}
 F_Y(y|\delta, b) &= \left( 1 + \left( \frac{\frac{e^{-\frac{\delta}{b}}}{1+e^{-\frac{\delta}{b}}} (1-y)}{y \left( \frac{1}{1+e^{-\frac{\delta}{b}}} \right)} \right)^b \right)^{-1} \\
 &= \left( 1 + \left( \frac{e^{-\frac{\delta}{b}} (1-y)}{y} \right)^b \right)^{-1} \\
 &= \left( 1 + e^{-\delta} \left( \frac{1-y}{y} \right)^b \right)^{-1}.
 \end{aligned}$$

### **Proof of Property 5.3.3**

Consider the reparameterization  $\alpha = \frac{1}{1 + \left(\frac{m}{1-m}\right)^b}$ ,  $\alpha \in (0, 1)$ . After a few simple algebraic manipulations we obtain the corresponding pdf and cdf of L-logistic distribution with parameter vector  $(\alpha, b)$  given by

$$f(y|\alpha, b) = \frac{b\alpha(1-\alpha)y^{b-1}(1-y)^{b-1}}{[y^b\alpha + (1-y)^b(1-\alpha)]^2}, \quad 0 < y < 1,$$

and

$$F_Y(y|\alpha, b) = \left(1 + \left(\frac{1-\alpha}{\alpha}\right) \left(\frac{1-y}{y}\right)^b\right)^{-1}, \quad 0 < y < 1.$$

### **Proof of Property 5.3.4**

Give  $Y \sim LL(\delta, b)$ , using the inverse of the transformation give in Property 3.2 and next the Property 3.1, we can easily obtain the pdf of the distribution of  $Z' = \delta + b \log\left(\frac{Y}{1-Y}\right)$  as being the pdf of the standard logistic distribution.

### **Proof of Property 5.3.5**

Defining the transformation  $Y = X(d-c) + c$ ,  $c, d \in \mathbb{R}$ , with  $X \sim LL(m, b)$ , we obtain easily the correspondent pdf of  $Y$ .

### **Proof of Property 5.3.6**

In order to find the mode  $y_0$  of the L-logistic distribution, we can take the derivative of the pdf given in (B.1), with respect to  $y$  as

$$\frac{\partial f(y|m, b)}{\partial y} = \frac{b((1-m)m)^b((1-y)y)^{-2+b} \{m^b(1-y)^b(-1+b+2y) - (1-m)^b(1+b-2y)y^b\}}{(m^b(1-y)^b + (1-m)^by^b)^3}.$$

Thus, making  $\frac{\partial f(y_0|m, b)}{\partial y} = 0$  we obtain  $\Leftrightarrow \{m^b(1-y)^b(-1+b+2y) - (1-m)^b(1+b-2y)y^b\} = 0$ . Therefore, the mode  $y_0$  when  $b > 1$  is a solution of the equation

$$\left(\frac{1-m}{m}\right)^b = \left(\frac{1-y_0}{y_0}\right)^b \frac{b+2y_0-1}{b-2y_0+1}.$$

For  $b \leq 1$ , the pdf is convex and does not have mode.

### **Proof of Property 5.3.7**

To show that the pdf of L-logistic distribution is symmetric when  $m = 0.5$  whatever the value of parameter  $b$ , is sufficient to show that  $f(0.5 - y) = f(0.5 + y)$ . Using the pdf given in (B.1) is

easy to show that

$$\begin{aligned}
 f(0.5 - y) &= \frac{b(1-0.5)^b 0.5^b (0.5-y)^{b-1} (1-0.5+y)^{b-1}}{[(1-0.5)^b (0.5-y)^b + 0.5^b (1-0.5+y)^b]^2} \\
 &= \frac{0.5^{2b} b(0.5-y)^{b-1} (0.5+y)^{b-1}}{0.5^{2b} [(0.5-y)^b + (0.5+y)^b]^2} \\
 &= \frac{b(0.5^2 - y^2)^{b-1}}{[(0.5-y)^b + (0.5+y)^b]^2} \\
 &= f(0.5 + y).
 \end{aligned}$$

### Proof of Property 5.3.8

Using the definition of the moments to (B.1) we have

$$E[Y^t] = \int_0^1 y^t \frac{b(1-m)^b m^b y^{b-1} (1-y)^{b-1}}{[(1-m)^b y^b + m^b (1-y)^b]^2} dy. \quad (\text{B.3})$$

By the transformation proposed in the Property 5.3.1, we have

$$E[Y^t] = \int_{-\infty}^{\infty} \left( \frac{1}{1 + \left(\frac{1-m}{m}\right) e^{-\frac{z}{b}}} \right)^t \frac{e^z}{[1+e^z]^2} dz.$$

Now, considering the transformation  $v = \frac{e^z}{1+e^z}$  (then  $z = \log\left(\frac{v}{1-v}\right)$  and  $dv = \frac{e^z}{(1+e^z)^2} dz$ .) and after a few simple algebraic manipulations, we obtain the expected result

$$E[Y^t] = \int_0^1 \left[ 1 + \left(\frac{1-v}{v}\right)^{1/b} \left(\frac{1-m}{m}\right) \right]^{-t} dv.$$

## Preliminary prior sensitivity analysis in Chapter 5.5

Here, we present a simulation study made to evaluate the sensitivity of different choices of prior distributions for parameter  $b$  of the L-logistic distribution. The five different prior distributions shown in Table T were considered for this study, which are assumed independent of the parameter  $m$  (with a unit uniform prior for  $m$ ). For simulate the data sets from the L-logistic distribution, we considered the values for  $m$  and  $b$  as follows:  $b \in \{0.5, 1, 5\}$  and  $m \in \{0.2, 0.5, 0.9\}$ , leading to nine scenarios or pairs of parameters, corresponding to nine models simulated. The size of the each data set was  $n = 100$ . Table T shows the values of WAIC, EAIC, EBIC, and DIC for the fitted models, considering the different prior distributions under study.

Table T – Statistics for model comparison, prior distributions for parameter  $b$  and the true value of the parameters of the L-logistic distribution used to simulate the data sets.

Parameter ( $m, b$ )	Prior	Criteria			
		WAIC	EAIC	EBIC	DIC
...					



Table T – Continued

Parameter ( $m, b$ )	Prior	Criteria			
		WAIC	EAIC	EBIC	DIC
(0.2, 0.5)	A, $b \sim \text{Gamma}(0.001, 0.001)$	41.76	-77.48	10.62	-82.68
	B, $b = U^2, U \sim \text{Uniform}(0, 100)$	41.76	-77.48	10.62	-82.69
	C, $\log(b) = L, L \sim \text{St}(10, 0, 2)$	41.78	-77.52	10.59	-82.75
	D, $b \sim \text{Gamma}(2.5, 1)$	41.79	-77.46	10.64	-82.65
	E, $b \sim \text{Gamma}(50, 1)$	35.83	-64.60	23.51	-56.91
(0.2, 1)	A, $b \sim \text{Gamma}(0.001, 0.001)$	31.95	-58.06	30.05	-63.48
	B, $b = U^2, U \sim \text{Uniform}(0, 100)$	31.96	-58.00	30.11	-63.35
	C, $\log(b) = L, L \sim \text{St}(10, 2)$	31.97	-58.07	30.04	-63.50
	D, $b \sim \text{Gamma}(2.5, 1)$	31.97	-58.03	30.08	-63.42
	E, $b \sim \text{Gamma}(50, 1)$	26.13	-45.28	42.83	-37.92
(0.2, 5)	A, $b \sim \text{Gamma}(0.001, 0.001)$	160.48	-314.94	-226.83	-320.15
	B, $b = U^2, U \sim \text{Uniform}(0, 100)$	160.48	-315.03	-226.93	-320.34
	C, $\log(b) = L, L \sim \text{St}(10, 2)$	160.49	-315.01	-226.90	-320.29
	D, $b \sim \text{Gamma}(2.5, 1)$	160.37	-314.85	-226.74	-319.97
	E, $b \sim \text{Gamma}(50, 1)$	156.06	-305.21	-217.11	-300.70
(0.5, 0.5)	A, $b \sim G(0.001, 0.001)$	25.89	-45.84	42.26	-51.16
	B, $b = U^2, U \sim \text{Uniform}(0, 100)$	25.91	-45.85	42.26	-51.18
	C, $\log(b) = L, L \sim \text{St}(10, 2)$	25.92	-45.85	42.26	-51.17
	D, $b \sim \text{Gamma}(2.5, 1)$	25.92	-45.86	42.24	-51.21
	E, $b \sim \text{Gamma}(50, 1)$	19.93	-32.76	55.34	-25.00
(0.5, 1)	A, $b \sim \text{Gamma}(0.001, 0.001)$	1.63	2.63	90.74	-2.73
	B, $b = U^2, U \sim \text{Uniform}(0, 100)$	1.65	2.67	90.77	-2.67
	C, $\log(b) = L, L \sim \text{St}(10, 2)$	1.66	2.62	90.73	-2.75
	D, $b \sim \text{Gamma}(2.5, 1)$	1.66	2.63	90.73	-2.74
	E, $b \sim \text{Gamma}(50, 1)$	-4.10	15.13	103.24	22.27
(0.5, 5)	A, $b \sim \text{Gamma}(0.001, 0.001)$	117.12	-228.30	-140.19	-233.62
	B, $b = U^2, U \sim \text{Uniform}(0, 100)$	117.12	-228.28	-140.17	-233.58
	C, $\log(b) = L, L \sim \text{St}(10, 2)$	117.13	-228.28	-140.18	-233.59
	D, $b \sim \text{Gamma}(2.5, 1)$	116.99	-228.09	-139.98	-233.20
	E, $b \sim \text{Gamma}(50, 1)$	112.63	-218.35	-130.25	-213.73
(0.9, 0.5)	A, $b \sim \text{Gamma}(0.001, 0.001)$	91.03	-176.01	-87.90	-181.16
	B, $b = U^2, U \sim \text{Uniform}(0, 100)$	91.05	-176.08	-87.98	-181.31
	C, $\log(b) = L, L \sim \text{St}(10, 2)$	91.07	-176.14	-88.03	-181.42
	D, $b \sim \text{Gamma}(2.5, 1)$	91.06	-176.04	-87.94	-181.23
	E, $b \sim \text{Gamma}(50, 1)$	84.98	-162.96	-74.85	-155.07
(0.9, 1)	A, $b \sim \text{Gamma}(0.001, 0.001)$	84.18	-162.36	-74.25	-167.55
	B, $b = U^2, U \sim \text{Uniform}(0, 100)$	84.20	-162.38	-74.28	-167.61
	C, $\log(b) = L, L \sim \text{St}(10, 2)$	84.24	-162.42	-74.32	-167.68
	D, $b \sim \text{Gamma}(2.5, 1)$	84.20	-162.33	-74.23	-167.501
	E, $b \sim \text{Gamma}(50, 1)$	78.54	-150.04	-61.94	-142.923

...

Table T – Continued

Parameter ( $m, b$ )	Prior	Criteria			
		WAIC	EAIC	EBIC	DIC
(0.9, 5)	A, $b \sim \text{Gamma}(0.001, 0.001)$ 9	218.35	-430.82	-342.71	-436.20
	B, $b = U^2, U \sim \text{Uniform}(0, 100)$	218.36	-430.84	-342.74	-436.24
	C, $\log(b) = L, L \sim \text{St}(10, 2)$	218.36	-430.67	-342.56	-435.90
	D, $b \sim \text{Gamma}(2.5, 1)$	218.25	-430.70	-342.59	-435.96
	E, $b \sim \text{Gamma}(50, 1)$	213.94	-421.05	-332.95	-416.66

As we can see in Table T, for all the simulated datasets in the nine scenarios, the values of WAIC, EAIC, EBIC, and DIC are quite close, showing no significant difference, giving evidence that the estimated models provide almost the same fit. Thus, for these cases, the posterior distribution does not seem to be sensitive with respect to the specification of these prior distributions. Additionally, when the prior E is used the estimated model achieves the worst fit among all models fitted with the other priors.