University of São Paulo
“Luiz de Queiroz” College of Agriculture

Advances on the Birnbaum-Saunders distribution

Luiz Ricardo Nakamura

Thesis presented to obtain the degree of Doctor in Science.
Area: Statistics and Agricultural Experimentation

Piracicaba
2016

Luiz Ricardo Nakamura
Degree in Statistics

Advances on the Birnbaum-Saunders distribution

versão revisada de acordo com a resolução CoPGr 6018 de 2011

Advisor:
Prof. Dr. **ROSELI APARECIDA LEANDRO**

Thesis presented to obtain the degree of Doctor in Science.
Area: Statistics and Agricultural Experimentation

Piracicaba
2016

Dados Internacionais de Catalogação na Publicação
DIVISÃO DE BIBLIOTECA - DIBD/ESALQ/USP

Nakamura, Luiz Ricardo
Advances on the Birnbaum-Saunders distribution / Luiz Ricardo Nakamura. - - versão revisada de acordo com a resolução CoPGr 6018 de 2011. - - Piracicaba, 2016.
94 p. : il.

Tese (Doutorado) - - Escola Superior de Agricultura "Luiz de Queiroz".

1. Distribuição Birnbaum-Saunders generalizada 2. GAMLSS 3. Modelos aditivos generalizados 4. Regressão não-paramétrica 5. Software R 6. Splines penalizados I. Título

CDD 519.532
N163a

DEDICATION

To my parents,
Edmilson Akiyoshi Nakamura *and* **Neide Henriques Novo Nakamura**

To them,
I lovingly dedicate this work.

ACKNOWLEDGMENTS

I would like to express my gratitude to all those who gave me the possibility to complete this thesis, especially my parents Edmilson Akiyoshi Nakamura and Neide Henrique Novo Nakamura and my brother Cézar Augusto Nakamura, for their love and supporting me throughout my life.

To my girlfriend Thaís Santarsiere, for her huge patience throughout the whole process of my Doctorate life. I really apologize for my extreme bad mood, thank you so much for understanding me.

To my adviser, Prof. Dr. Roseli Aparecida Leandro, for the continuous support of my Doctorate, for her patience, motivation, enthusiasm and immense knowledge.

To Prof. Dr. Cristian Marcelo Villegas Lobos, for his scientific contribution, intellectual input and especially for participating in the whole process of the Doctorate training. I really appreciate your friendship.

To Prof. Dr. Mikis Stasinopoulos and Prof. Dr. Robert Rigby from London Metropolitan University, for receiving me on my internship period and for all the support during my stay in London.

To Prof. Dr. Clarice Garcia Borges Demétrio and Prof. Dr. Taciana Villela Savian, at ESALQ/USP, for their valuable guidance and friendship.

To Dr. Vlasios Voudouris and Prof. Dr. Robert Gilchrist for their scientific contribution and intellectual input.

To my special friends, Ana Julia Righetto, Andrea Isabel Estévez Garcia, Camila Rodrigues Gonçalves Cerqueira, Natalie Verónika Rondinel Mendoza, Pedro Henrique Ramos Cerqueira, Renan Mercuri Pinto, Rodrigo Rossetto Pescim and Thiago Gentil Ramires who were always willing to help, and also, for providing me a more enjoyable life in the last years.

To my friends from London, Abu Hossain, Celia Garcia, Fernanda De Bastiani, Roberta Freezor, Silvia Laiz, Svetlomir Stoyanov and Veera Snellman for their help, support and friendship helping me to have a more pleasure life during my stay in the UK.

To my friends from Piracicaba and ESALQ/USP, Alan Winkler, Cássio Dessotti, Cristiano Dela Piccolla, Djair Frade, Edilan Quaresma, Ezequiel López, Fernando Mayer, Guilherme Biz, João Vitor Teodoro, Lucas Cunha, Simone Grego, Talita Tanaka and Wagner Wolff.

To the staff of the Department of Exact Science at ESALQ/USP, the secretaries Luciane Brajão, Mayara Segatto, Rosni Pinto and Solange Sabadin, in special to the computer technicians Eduardo Bonilha and Jorge Wiendl.

This work was fully supported by CAPES, Coordenação de Aperfeiçoamento de Pessoal de Nível Superior, Brazil.

SUMMARY

RESUMO

Avanços na distribuição Birnbaum-Saunders

A distribuição Birnbaum-Saunders (BS) é o modelo mais popular utilizado para descrever processos de fadiga. Ao longo dos anos, essa distribuição vem recebendo aplicações nas mais diversas áreas, demandando assim algumas extensões mais flexíveis para resolver problemas mais complexos. Uma das extensões mais conhecidas na literatura é a família de distribuições Birnbaum-Saunders generalizada (GBS), que inclui as distribuições Birnbaum-Saunders caso-especial (BS-SC) e Birnbaum-Saunders t generalizada (BSGT) como modelos especiais. Embora a distribuição BS-SC tenha sido previamente desenvolvida na literatura, nunca foi estudada mais profundamente e, assim, nesta tese, um estudo bayesiano é desenvolvido acerca da mesma além de um novo gerador de números aleatórios dessa distribuição ser apresentado. Adicionalmente, um modelo de regressão baseado na distribuição BSGT é desenvolvido utilizando-se os modelos aditivos generalizados para locação, escala e forma (GAMLSS), os quais apresentam grande flexibilidade tanto para a assimetria como para a curtose. Uma nova extensão da distribuição BS também é apresentada, denominada família de distribuições Birnbaum-Saunders potência (BSP), que contém inúmeros casos especiais ou limites já publicados na literatura, incluindo a família GBS. A principal característica desta nova família é que ela é capaz de produzir formas tanto uni como bimodais dependendo do valor de seus parâmetros. Esta nova família também é introduzida na estrutura dos modelos GAMLSS para fornecer uma ferramenta capaz de modelar todos os parâmetros da distribuição como funções lineares e/ou não-lineares suavizadas de variáveis explicativas. Ao longo desta tese são apresentadas cinco diferentes aplicações em conjuntos de dados reais para ilustrar os resultados teóricos obtidos.

Palavras-chave: Distribuição Birnbaum-Saunders generalizada; GAMLSS; Modelos aditivos generalizados; Regressão não paramétrica; Linguagem R; Splines penalizados

ABSTRACT

Advances on the Birnbaum-Saunders distribution

The Birnbaum-Saunders (BS) distribution is the most popular model used to describe lifetime process under fatigue. Throughout the years, this distribution has received a wide ranging of applications, demanding some more flexible extensions to solve more complex problems. One of the most well-known extensions of the BS distribution is the generalized Birnbaum-Saunders (GBS) family of distributions that includes the Birnbaum-Saunders special-case (BS-SC) and the Birnbaum-Saunders generalized t (BSGT) models as special cases. Although the BS-SC distribution was previously developed in the literature, it was never deeply studied and hence, in this thesis, we provide a full Bayesian study and develop a tool to generate random numbers from this distribution. Further, we develop a very flexible regression model, that admits different degrees of skewness and kurtosis, based on the BSGT distribution using the generalized additive models for location, scale and shape (GAMLSS) framework. We also introduce a new extension of the BS distribution called the Birnbaum-Saunders power (BSP) family of distributions, which contains several special or limiting cases already published in the literature, including the GBS family. The main feature of the new family is that it can produce both unimodal and bimodal shapes depending on its parameter values. We also introduce this new family of distributions into the GAMLSS framework, in order to model any or all the parameters of the distribution using parametric linear and/or nonparametric smooth functions of explanatory variables. Throughout this thesis we present five different applications in real data sets in order to illustrate the developed theoretical results.

Keywords: Generalized additive models; Generalized Birnbaum-Saunders distribution; GAMLSS; Non-parametric regression; Penalized splines; R software

1 INTRODUCTION

In the last decades, several new distributions, especially continuous univariate ones, and models are being created in order to solve increasingly more complex problems. Among them, the Birnbaum-Saunders (BS) (BIRNBAUM; SAUNDERS, 1969) distribution is the most popular model used to describe lifetime process under fatigue. Throughout the years, the BS distribution is being applied in different fields apart from the fatigue problems (PESCIM et al., 2014), demanding some extensions in order to provide more flexible models. Díaz-García and Leiva (2005) proposed the generalized Birnbaum-Saunders (GBS) family of distributions based on elliptical distributions, e.g. student's t, Cauchy and Laplace distribution. Díaz-García and Domínguez-Molina (2006) presented the three-parameter BS distribution. Later, Vilca and Leiva (2006) developed a BS model based on skew normal distributions. Vilca et al. (2010) and Castillo et al. (2011) developed the epsilon-skew BS distribution. More recently, Cordeiro and Lemonte (2011) and Pescim et al. (2014) defined the beta BS and the Kummer beta BS distributions, respectively.

In several practical applications, the response variable Y is affected by one or more explanatory variables. In the cases where the response variable follows a non-Gaussian distribution that belongs to the exponential family, the generalized linear model (GLM) (NELDER; WEDDERBURN, 1972) is one of the most used models in the literature, being applied in different problems, such as in Demétrio et al. (2007) who evaluated possible factors that affects conception rates in lactating holstein cows; Mascarin et al. (2010) and Urbano et al. (2013) studied the relationship between larva mortality at different virus concentrations at different temperatures; among others.

Another very common approach to explain a response variable using explanatory variables is the generalized additive models (GAM) (HASTIE; TIBSHIRANI, 1990). As in the GLM, this model allows that the response variable follow any distribution that belongs to the exponential family. In GAM, the mean μ of the response variable is modelled as parameteric and/or non-parametric functions, e.g. penalized splines (EILERS; MARX, 1996), of the explanatory variables. Some applications using this approach can be seen in Pearce et al. (2011) who presented a study about the relationship of air pollutant concentration with variables as temperature, atmospheric pressure and radiation; Pullenayegum et al. (2013) performed a study of health utilities among patients with diabetes; McKeown and Sneddon (2014) measured patients' emotions over time; among others.

The main assumption of the previous models is that the response variable Y must follow a distribution which belongs to the exponential family. Since the BS distribution and its extensions do not belong to the exponential family, usually their regression models are based on the log location-scale models (LAWLESS, 2003). Among them, Rieck and Nedelman (1991) and

Barros et al. (2008) proposed the log-Birnbaum-Saunders and log-Birnbaum-Saunders-t regression models, respectively. However, as in the GLM and GAM, we are just allowed to model the mean μ of the response variable as a function of explanatory variables. Such properties may be an issue when we are modelling very complex data sets, specially the ones involving highly skew data and excess kurtosis.

In order to provide more flexibility to the usual models, Rigby and Stasinopoulos (2005) developed the generalized additive models for location, scale and shape (GAMLSS). A very flexible class of univariate semi-parametric regression models which allows any parameter (not only the location parameter) from a given distribution (that does not necessarily belong to the exponential family) to be modelled as parametric and/or additive nonparametric smooth functions of explanatory variables. This approach is receiving great deal of attention in recent years since any distribution can be used to model a given response variable Y, including high skewed, platykurtic and leptokurtic shapes, as functions of a set of covariates. For instance, GAMLSS framework was used to study the projections of production of natural gas in Voudouris et al. (2014), to study bacterial cellulose production from agro-industrial waste in Hernández et al. (2015), to model proportion data including 0 and 1 in Hossain et al. (2016), among others.

This present thesis has as main objective to provide some advances on the BS distribution and it is organized in four different chapters as follows. In Chapter 2, we discuss the estimation of the Birnbaum-Saunders special-case (BS-SC) distribution under the Bayesian framework, proposing a new method based on Markov chain Monte Carlo (MCMC) to generate observations for the given distribution. In Section 2.1 we give a very brief review of the Birnbaum-Saunders (BS) distribution, linking it with the generalized Birnbaum-Saunders (GBS) family of distributions, presenting some of the works available in the literature and motivating the use of the BS-SC distribution. Section 2.2 provides a study about some properties of the BS-SC distribution. In Section 2.3 we discuss about prior distribution and posterior analysis. We conduct a simulation study generating observations from the BS-SC distribution using the proposed random number generator through the MCMC method in Section 2.4. An application is presented in Section 2.5. Some concluding remarks are addressed in Section 2.6.

In Chapter 3, we provide a regression model for the Birnbaum-Saunders generalized t (BSGT) (Genç, 2013) distribution, which is a special case of the GBS family, based on the GAMLSS framework. In Section 3.1 we cite some works available in the literature using the BS distribution and some of its extensions, providing a motivation for the use of the new regression model. Section 3.2 presents a review of the BSGT distribution, showing some of its special and/or limiting cases. In Section 3.3 we introduce the BSGT distribution into the GAMLSS framework. A simulation study is performed in Section 3.4. We provide an application in Section 3.5. Section 3.6 ends Chapter 3 with some concluding remarks.

In Chapter 4, we define a new family of distributions called the Birnbaum-Saunders power (BSP) distribution. Section 4.1 introduces the new family of distributions. In Section 4.2 we present some of its special and/or limiting cases. The method of maximum likelihood is discussed in Section 4.3 in order to obtain estimates for the BSP parameters. Section 4.4 visits the GAMLSS framework again, presenting a new R package specially developed for fitting BSP models. We show the great flexibility of the new BSP distributions in an application available in Section 4.5. We end this chapter with some concluding remarks in Section 4.6.

Finally, in Chapter 5 we focus on the bimodality shapes of the BSP family of distributions. In Section 5.1 we present a motivation to use this family. Section 5.2 presents a brief review about the GAMLSS framework used to produce the bimodal regression model. A simulation study, including explanatory variables is presented in Section 5.3. Section 5.4 provides two real data set applications. Finally, Section 5.5 concludes the chapter with some comments.

References

BARROS, M.; PAULA, G.A.; LEIVA, V. A new class of survival regression models with heavy-tailed errors: robustness and diagnostics. **Lifetime Data Analysis**, Boston, v. 14, p. 316–332, 2008.

BIRNBAUM, Z. W.; SAUNDERS, S. C. A new family of life distributions. **Journal of Applied Probability**, Sheffield, v. 6, p. 319–327.

CASTILLO, N.O.; GÓMEZ, H.W.; BOLFARINE, H. Epsilon Birnbaum-Saunders distribution family: properties and inference. **Statistical Papers**, Berlin, v. 52, p. 871–883, 2011.

CORDEIRO, G. M.; LEMONTE, A. J. The beta Birnbaum-Saunders distribution: an improved distribution for fatigue life modeling. **Computational Statistics & Data Analysis**, Amsterdam, v. 55, p. 1445–1461, 2011.

DEMÉTRIO, D. G. B.; SANTOS, R. M.; DEMÉTRIO, C. G. B.; VASCONCELOS, J. L. M. Factors affecting conception rates following artificial insemination or embryo transfer in lactating holstein cows. **Journal of Dairy Science**, Champaign, v. 90, n. 11, p. 5073–5082, 2007.

DÍAZ-GARCÍA, J. A.; LEIVA, V. A new family of life distributions based on the elliptically contourned distribution. **Journal of Statistical Planning and Inference**, Amsterdam, v. 128, p. 445–457, 2005.

DÍAZ-GARCÍA, J.A.; DOMÍNGUEZ-MOLINA, J. R. Some generalisations of Birnbaum-Saunders and sinh-normal distributions. **International Mathematical Forum**, Ruse, v. 1, p. 1709-1727, 2006.

EILERS, P. H. C.; MARX, B. D. Flexible smoothing with B-splines and penalties. **Statistical Science**, Hayward, v. 11, p. 89–102, 1996.

GENÇ, A. I. The generalized T Birnbaum-Saunders family. **Statistics: A Journal of Theoretical and Applied Statistics**, Hampshire, v. 47, n. 3, p. 613–625, 2013.

HASTIE, T. J.; TIBSHIRANI, R. J. **Generalized Additive Models**. London: Chapman and Hall/CRC, 1990. 352 p.

HERNÁNDEZ, F.; TORRES, M.; ARTEAGA, L.; CASTRO, C. GAMLSS models applied in the treatment of agro-industrial waste. **Comunicaciones en Estadística**, Bogotá, v. 8, n. 2, p. 245–254, 2015.

HOSSAIN, A. RIGBY, R.; STASINOPOULOS, M.; ENEA, M. Centile estimation for a proportion response variable. **Statistics in Medicine**, Chichester, v. 35, p. 895–904, 2016.

LAWLESS, J. F. **Statistical Models and Methods for Lifetime Data**. 2nd ed. New York: Wiley, 2003. 630 p.

MASCARIN, G. M.; ALVES, S. B.; RAMPELOTTI-FERREIRA, F. T.; URBANO, M. R.; DEMÉTRIO, C. G. B.; DELALIBERA JR., I. Potential of a granulovirus isolate to control *Phthorimaea operculella* (Lepidoptera: Gelechiidae). **Biocontrol**, Dordrecht, v. 55, n. 5, p. 657–671, 2010.

NELDER, J. A.; WEDDERBURN, R. W. M. Generalized linear models. **Journal of the Royal Statistical Society – Series A (General)**, London, v. 135, p. 370–384, 1972.

PESCIM, R. R.; CORDEIRO, G. M.; NADARAJAH, S.; DEMÉTRIO, C. G. B.; ORTEGA, E. M. M. The Kummer beta Birnbaum-Saunders: an alternative fatigue life distribution. **Hacettepe Journal of Mathematics and Statistics**, Ankara, v. 43, p. 473–510, 2014.

PEARCE, J. L.; BERINGER, J.; NICHOLLS, N.; HYNDMAN, R. J.; TAPPER, N. J. Quantifying the influence of local meteorology on air quality using generalized additive models. **Atmospheric Environment**, Oxford, UK, v. 45, n. 6, p. 1328–1336, 2011.

PULLENAYEGUM, E. M.; WONG, H. S.; CHILDS, A. Generalized additive models for the analysis of EQ-5D utility Data. **Medical Decision Making**, Thousand Oaks, v. 33, n. 2, p. 244–251, 2013.

RIECK, J. R; NEDELMAN, J. R. A log-linear model for the Birnbaum-Saunders distribution. **Technometrics**, Richmond, v. 33, p. 51–60, 1991.

RIGBY, R.A.; STASINOPOULOS, D.M. Generalized additive models for location, scale and shape. **Journal of the Royal Statistical Society – Series C (Applied Statistics)**, Oxford, UK, v. 54, p. 507–554, 2005.

URBANO, M. R.; HINDE, J.; DEMÉTRIO, C. G. B. Dose response models with natural mortality and random effects. **Journal of Agricultural, Biological, and Environmental Statistics**, Alexandria, VA, v. 18, n. 4, p. 594–610, 2013.

VILCA, F.; LEIVA, V. A new fatigue life model based on the family of skew-elliptical distributions. **Communications in Statistics: Theory and Methods**, New York, v. 35, p. 229-244, 2006.

VILCA, F.; SANHUEZA, A.; LEIVA, V.; CHRISTAKOS, G. An extended Birnbaum-Saunders model and its application in the study of environmental quality in Santiago, Chile. **Stochastic Environmental Research and Risk Assessment**, Berlin, v. 24, p. 771–782, 2010.

VOUDOURIS, V.; MATSUMOTO, K.; SEDGWICK, J.; RIGBY, R.; STASINOPOULOS, M.; JEFFERSON, M. Exploring the production of natural gas through the lenses of the ACEGES model. **Energy Policy**, Guildford, v. 64, p. 124–133, 2014.

2 BAYESIAN INFERENCES FOR THE BIRNBAUM-SAUNDERS SPECIAL-CASE DISTRIBUTION[1]

Abstract

In this paper, we discuss the estimation of the Birnbaum-Saunders Special-Case (BS-SC) distribution through the Bayesian approach considering its parameters independents, assuming gamma priors for both of them. As the full posterior conditionals do not have closed forms we use the Metropolis-Hastings algorithm to generate samples from the joint posterior distribution. We present a simulation study proposing the Markov chain Monte Carlo (MCMC) method as a random number generator, considering the cases where the BS-SC distribution has symmetric and asymmetric shapes. An application related to ozone concentration is presented in this paper using the described methodology.

Keywords: Generalized Birnbaum-Saunders distributions; Markov Chain Monte Carlo; Metropolis-Hastings algorithm; Random number generator

2.1 Introduction

The Birnbaum-Saunders (BS) distribution was developed to study problems of vibration in commercial aircraft that caused fatigue in the materials (BIRNBAUM; SAUNDERS, 1969). The authors used their knowledge of fatigue problems to build a new family of distributions, which models materials lifetime subject to dynamic loads. Through the years, the BS distribution has been widespread in many works, such as Rieck and Nedelman (1991) created a log-linear model for the BS distribution; Achcar (1993) introduced the Bayesian approach on the estimation of the BS parameters; Villegas et al. (2011) introduced the BS mixed models for censored data; Balakrishnan et al. (2011) presented mixtured models based on the BS distribution; among others.

A random variable T with parameters $\alpha > 0$ and $\beta > 0$, denoted by $T \sim BS(\alpha, \beta)$, is defined in terms of the Gaussian distribution as follows

$$T = \beta \left[\frac{\alpha Z}{2} + \sqrt{\left(\frac{\alpha Z}{2} \right)^2 + 1} \right]^2, \tag{2.1}$$

where $Z \sim N(0,1)$. Its probability density function (pdf) is given by

$$f_T(t) = \frac{t^{-3/2}(t+\beta)}{2\alpha\sqrt{2\pi\beta}} \exp\left\{ -\frac{1}{2\alpha^2} \left(\frac{t}{\beta} + \frac{\beta}{t} - 2 \right) \right\}, \ t > 0, \tag{2.2}$$

where $\alpha > 0$ is the shape parameter and $\beta > 0$ is the scale parameter and median of the distribution. As α grows, the BS distribution becomes positively asymmetrical, whereas when $\alpha \to 0$, the distribution becomes symmetric around β.

[1]Luiz Ricardo Nakamura, Roseli Aparecida Leandro (ESALQ/USP) and Cristian Villegas (ESALQ/USP). Revista Brasileira de Biometria, v. 34, n. 2, Forthcoming.

Later, Díaz-García and Leiva (2005) proposed a new family of distributions so-called the generalized Birnbaum-Saunders (GBS), defined in terms of elliptic distributions. Here, the assumption that $Z \sim N(0,1)$ from equation (4.1) is relaxed for any univariate symmetric distribution, i.e.

$$T = \beta \left[\frac{\alpha U}{2} + \sqrt{\left(\frac{\alpha U}{2}\right)^2 + 1} \right]^2 ,$$

where $U \sim S(0,1;g)$, g corresponds to the kernel of the pdf of symmetric distribution used and α and β are the same as presented in equation (2.2). Thus, it follows that a random variable T follows a GBS distribution, denoted by $T \sim GBS(\alpha, \beta; g)$, if its pdf is given by

$$f_T(t) = c \frac{t^{-3/2}(t+\beta)}{2\alpha\sqrt{\beta}} g \left[\frac{1}{\alpha^2} \left(\frac{t}{\beta} + \frac{\beta}{t} - 2 \right) \right], \; t > 0, \tag{2.3}$$

where c is the normalization constant and g corresponds to the kernel of the pdf of symmetric distribution to be used. In particular, when $U \sim N(0,1)$ we have the BS distribution. For instance, some other extensions of the BS distribution can be found in Vilca-Labra and Leiva (2006) who assumed that U could follow any skew elliptical distribution; Owen (2006) developed a three-parameter BS distribution; Gómez et al. (2009) introduced the generalized slash Birnbaum-Saunders family of distributions; Castillo et al. (2009) proposed a new extension based on the epsilon-skew symmetric distributions; Guiraud et al. (2009) and Leiva et al. (2012) introduced a non-centrality parameter to the BS and BS-t distributions; among others.

Due to their properties and flexibility in modelling different types of data, the GBS distributions received wide attention in different research areas, e.g. Leiva et al. (2008) modelled the air pollutant concentration in Chile using the GBS distributions; Leiva et al. (2012) used the GBS distributions in the forestry sciences, modelling the diameter of trees; Marchant et al. (2013) utilized distributions from the GBS family on a financial dataset. Cancho et al. (2010) present the only study using Bayesian approach on a GBS distribution besides the standard BS.

One of the GBS distributions that are not explored in the literature is the Birnbaum-Saunders special-case (BS-SC) distribution, also proposed by Díaz-García and Leiva (2005), which has as baseline the special-case distribution. For further information, see (GUPTA; VARGA, 1993). The BS-SC model has heavier tails than the classic BS distribution and so could be used in cases where there are only a few observations on the extremes of the distribution. Also, since the BS-SC has heavy tails we can compare it with the BS-t distribution but, in a Bayesian approach, the BS-SC distribution is easier to fit considering that the BS-t model has a degree of freedom parameter (ν) which is somewhat not very easy to estimate.

In this paper we consider the Bayesian inference as a tool for parameter estimation of the BS-SC distribution. This approach was chosen since the distribution has only its first moment and therefore becomes the natural choice for the inference process, since in the classical approach some asymptotic assumptions are violated and thus the estimates are not reliable. The modelling of uncertainty on shape (α) and scale (β) parameters, considered independent in this work, was performed by gamma prior distributions due to their parametric space. Since the full conditional posterior distributions do not have closed form, the Metropolis-Hastings (HASTINGS, 1970) was used to obtain samples of the joint posterior distribution, and hence the Bayesian estimates.

For the simulation study, we propose the generation of data from BS-SC distribution to be performed by MCMC-based algorithms (as Metropolis-Hastings), since the quantile function of this model does not have a closed form. After data generation Bayesian estimates were obtained and compared. Finally, an application, comparing the BS, BS-t and BS-SC distributions, to a real dataset related to ozone concentration in New York city is presented in order to validate the inference process.

The rest of this paper is organized as follows. In Section 1.2, we define the BS-SC distribution, notation and structure, comparing it to the classic BS distribution. In Section 1.3, prior distribution and posterior analysis are described. In Section 1.4, we bring up the simulation study, with data generation and its estimates. In Section 1.5, an illustrative example based on real data is provided. Finally, Section 1.6 ends with some concluding remarks.

2.2 Birnbaum-Saunders Special-Case distribution

Let X be a random variable which follows a Special-Case (SC) distribution (GUPTA; VARGA, 1993), denoted by $X \sim SC(\mu, \sigma)$, so its pdf is given by

$$f_X(x) = \frac{2^{\frac{1}{2}}}{\pi\sigma} \left(1 + \left[\frac{(x-\mu)^2}{\sigma^2}\right]^2\right)^{-1}, \; x \in \mathbb{R} \tag{2.4}$$

where $\mu \in \mathbb{R}$ and $\sigma > 0$ are, respectively, location and scale parameters of the distribution.

The SC model is a symmetric distribution (GUPTA; VARGA, 1993) that has heavier tails than the Gaussian distribution and hence could be an interesting competitive model to it and to the BS-t distribution when there are some extreme values in the tails of the distribution. Further, the SC distribution allocates more information around its mode as we can see in Figure 2.1 that presents a comparison between the BS and BS-SC models for different values of σ^2.

The only moments that can be obtained for this distribution are the first and second one. For any $n \geq 3$, $\text{E}(X^n)$ does not exist since they diverge. Mean and variance of the SC distribution are given respectively by

$$\text{E}(X) = \mu \quad \text{and} \quad \text{Var}(X) = \sigma^2,$$

which are the same of the Gaussian distribution.

An extension of the BS distribution was proposed by Díaz-García and Leiva (2005), where they presented the family of generalized Birnbaum-Saunders (GBS) distributions, which pdf is expressed in (2.3). One particular case of the GBS distribution is the Birnbaum-Saunders special-case (BS-SC) distribution that is obtained writing the pdf (2.4) as equation (2.3).

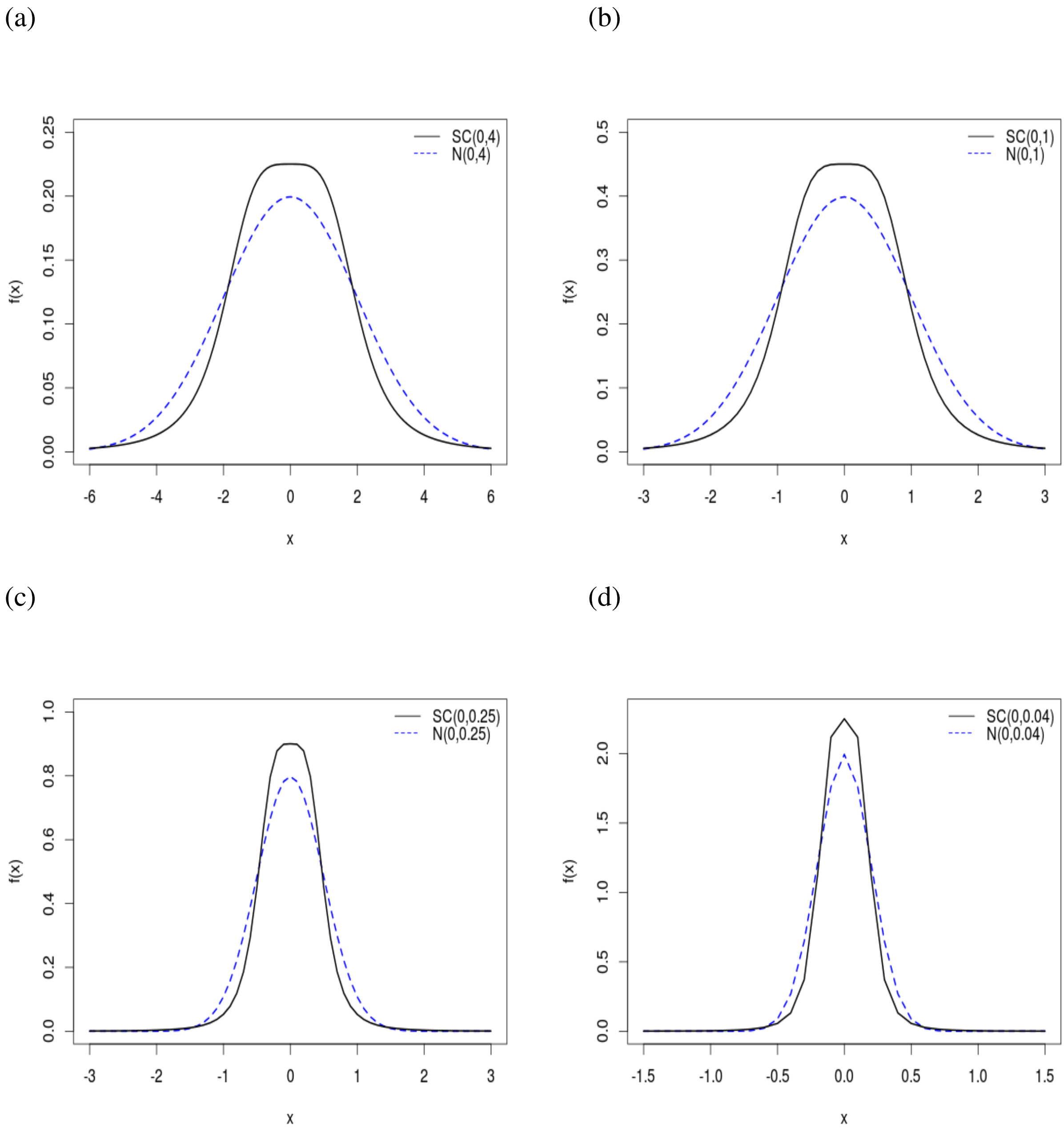


Figure 2.1 – Probability density functions for Gaussian and SC distributions with $\mu = 0$ and different values of σ: (a) $\sigma^2 = 4$; (b) $\sigma^2 = 1$; (c) $\sigma^2 = 0.25$; and (d) $\sigma^2 = 0.04$.

We say that a random variable T follows a BS-SC distribution, denoted as $T \sim BS - SC(\alpha, \beta)$, if its pdf is given by

$$f_T(t) = \frac{t^{-3/2}(t+\beta)}{\pi\alpha\sqrt{2\beta}} \left[1 + \frac{1}{\alpha^4}\left(\frac{t}{\beta} + \frac{\beta}{t} - 2\right)^2\right]^{-1}, \; t > 0,$$

where $\alpha > 0$ and $\beta > 0$ are the shape and scale parameters. If $T \sim BS - SC(\alpha, \beta)$, then $Y = aT \sim \mathscr{BS} - \mathscr{SC}(\alpha, a\beta)$ and $Y = T^{-1} \sim BS - SC(\alpha, \beta^{-1})$ (DÍAZ-GARCÍA; LEIVA, 2005).

According to theorem 3 from Díaz-García and Leiva (2005), the only moment that can be obtained of the BS-SC distribution is the first one, given by

$$\mathrm{E}(T) = \beta\left(1 + \frac{\alpha^2}{2}\right),$$

which is exactly equal to the first moment of the classic BS distribution (BIRNBAUM; SAUNDERS, 1969).

The pdf behavior of a random variable $T \sim BS - SC(\alpha, \beta)$ is quite similar to BS pdf (Figure 2.2). Graphically, the main difference between the BS-SC and BS distributions as expected from the comparison between the SC and Gaussian distribution, comes from the fact that the first one has heavier extreme tails than the second one. Also, we can observe that the BS-SC distribution allocates more observations around its mode than the BS distribution when $0 < \alpha < 1$. These two main differences make the BS-SC more attractive than the classic BS distribution in cases where some extreme values are observed on the tails of the distribution.

2.3 Prior distribution and posterior analysis

Let $T_1, \ldots, T_n$ be independent and identically distributed random variables, where $T_i \sim BS-\ SC(\alpha, \beta)$, $i = 1, \ldots, n$. A useful reparametrization for the classic BS distribution is $\lambda = \alpha^{-2}$ since we can take a conditionally conjugate gamma prior for λ. Here we use the same reparametrization although the conjugate property is not valid in our case. Thus, setting $\lambda = \alpha^{-2}$, the BS-SC likelihood function, without normalization constant, can be written as

$$L(\lambda, \beta|\mathscr{D}) \propto \frac{\lambda^{n/2}}{\beta^{n/2}} \frac{\prod_{i=1}^{n} t_i^{-3/2}\,(t_i + \beta)}{\prod_{i=1}^{n}\left[1 + \lambda^2\left(\frac{t_i}{\beta} + \frac{\beta}{t_i} - 2\right)^2\right]},\ \lambda > 0, \tag{2.5}$$

where $\mathscr{D}$ denotes the data.

The uncertainty of the parameters λ and β, considered to have independent prior distributions, is described as

$$\pi(\lambda) \propto \lambda^{a-1}\mathrm{e}^{-b\lambda}, \quad \lambda > 0, \tag{2.6}$$

and

$$\pi(\beta) \propto \beta^{c-1}\mathrm{e}^{-d\beta}, \quad \beta > 0, \tag{2.7}$$

i.e., we used the gamma distribution with shape and rate hyperparameters a and b, respectively, for λ and gamma distribution with shape and rate hyperparamenters c and d, respectively, for β.

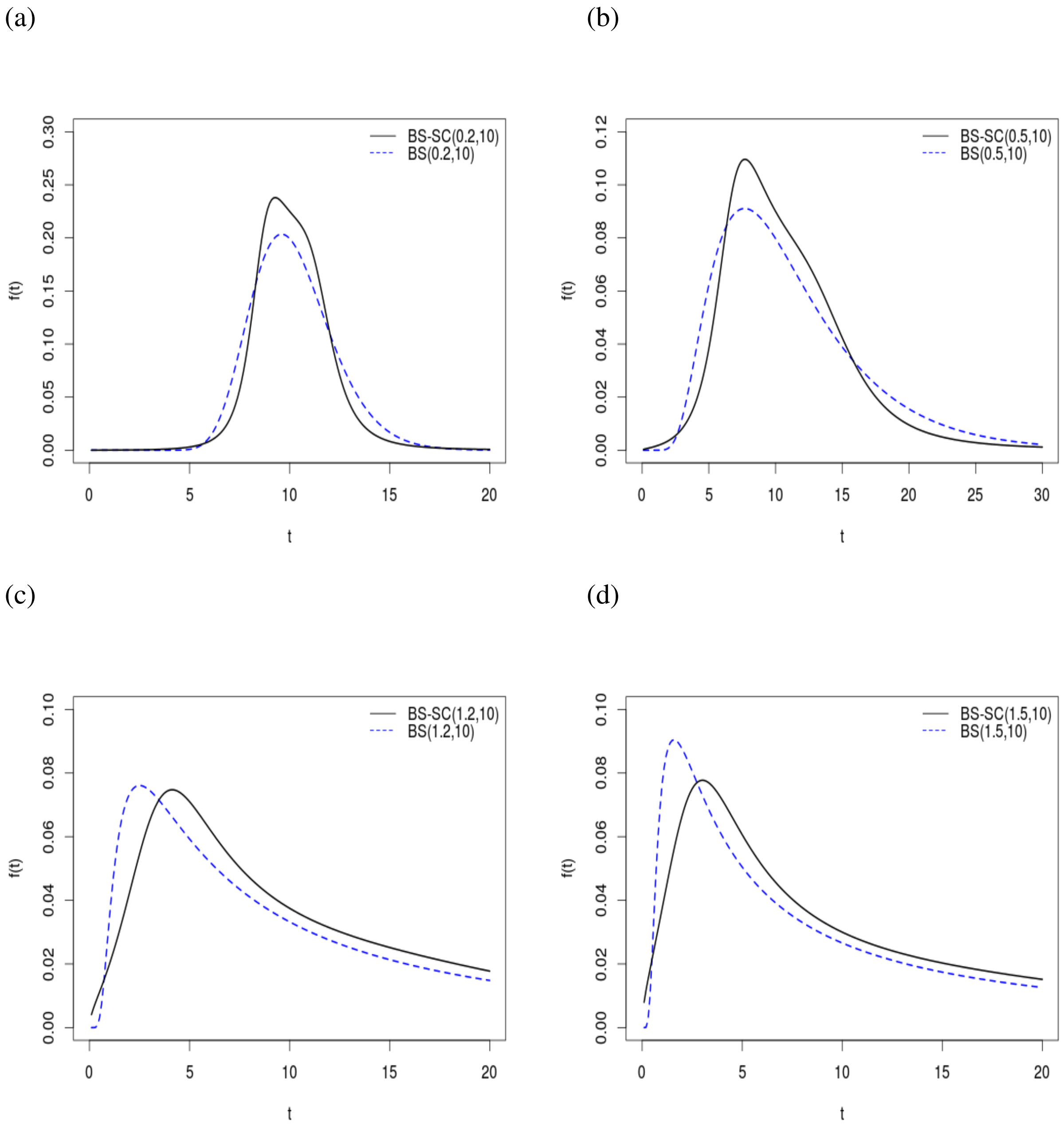


Figure 2.2 – Probability density functions for BS and BS-SC distributions with $\beta = 10$ and different values of α: (a) $\alpha = 0.2$; (b) $\alpha = 0.5$; (c) $\alpha = 1.2$; and (d) $\alpha = 1.5$.

Combining the information from data in equation (4.4), with the prior information from equations (2.6) and (2.7), we obtain the joint posterior density function of (λ,β), i.e.

$$\pi(\lambda, \beta|\mathscr{D}) \propto \frac{\lambda^{\frac{n}{2}+a-1}\mathrm{e}^{-b\lambda-d\beta}}{\beta^{\frac{n}{2}-c+1}} \frac{\prod_{i=1}^{n} t_i^{-3/2}\left(t_i+\beta\right)}{\prod_{i=1}^{n}\left[1+\lambda^2\left(\frac{t_i}{\beta}+\frac{\beta}{t_i}-2\right)^2\right]}. \tag{2.8}$$

Therefore, the marginal posterior distributions are easily obtained from equation (2.8) as follows

$$\pi(\lambda|\beta,\mathscr{D}) \propto \frac{\lambda^{\frac{n}{2}+a-1}\mathrm{e}^{-b\lambda}}{\prod_{i=1}^{n}\left[1+\lambda^2\left(\frac{t_i}{\beta}+\frac{\beta}{t_i}-2\right)^2\right]}$$

and

$$\pi(\beta|\lambda,\mathscr{D}) \propto \beta^{c-\frac{n}{2}-1}\mathrm{e}^{-d\beta}\frac{\prod_{i=1}^{n} t_i^{-\frac{3}{2}}\left(t_i+\beta\right)}{\prod_{i=1}^{n}\left[1+\lambda^2\left(\frac{t_i}{\beta}+\frac{\beta}{t_i}-2\right)^2\right]}.$$

We can observe that the marginal posterior distributions do not have closed form and, thus, to acquire samples from the joint posterior distribution the Metropolis-Hastings (HASTINGS, 1970) algorithm will be used.

2.4 Simulation study

Since the quantile function of the BS-SC distribution has no closed form, we propose in this paper the data generation of this model via Metropolis-Hastings algorithm. The steps to obtain the observations are described below:

- Step 1: Establish an initial value for the start of the algorithm, denoted by $y^{(0)}$;
- Step 2: $y^{(i+1)} = y^{(i)}$, where $y^{(i)}$, $i = 0, \ldots, M-1$, is the new sample of the chain;
- Step 3: Generate a new candidate y_{new} from a proposal distribution $g(y)$;
- Step 4: Generate u from an Uniform(0,1);
- Step 5: If $u > \frac{f(y^{(i)})}{f(y_{\text{new}})}\frac{g(y_{\text{new}})}{g(y^{(i)})}$ we should keep the observation $y^{(i)}$, otherwise $y^{(i)} = y_{\text{new}}$;
- Step 6: Repeat Steps 2 to 5 until a certain number of observations M is obtained.

It is noteworthy that the acceptance rate should be maintained between 25% and 45%, considering that a low acceptance rate may indicate that the sample values are in the distribution tails, while a high acceptance rate may indicate that the values are being sampled only from regions with high probability density.

In this study we generate four different scenarios with the BS-SC model, covering cases where the shape of the distribution is near symmetrical ($\alpha = 0.2$) or very asymmetrical ($\alpha = 1.5$):

- Scenario 1: $BS-SC(\alpha = 0.2, \beta = 1.5)$
- Scenario 2: $BS-SC(\alpha = 0.2, \beta = 0.2)$

Table 2.1 – Average estimates and the associated Monte Carlo errors for the Bayesian approach of the simulation from the BS-SC distribution with different values of α and β

Empirical distribution	Number of observations	Prior 1		Prior 2	
		$\hat{\alpha}$	$\hat{\beta}$	$\hat{\alpha}$	$\hat{\beta}$
	15	0.2144 (0.0009)	1.5037 (0.0200)	0.2086 (0.0009)	1.5031 (0.0011)
$\alpha = 0.2$	20	0.2106 (0.0007)	1.5038 (0.0013)	0.2065 (0.0007)	1.5031 (0.0009)
$\beta = 1.5$	30	0.2056 (0.0005)	1.5024 (0.0007)	0.2030 (0.0005)	1.5022 (0.0007)
	50	0.2036 (0.0004)	1.5007 (0.0005)	0.2022 (0.0003)	1.5006 (0.0005)
	100	0.2018 (0.0002)	1.5006 (0.0004)	0.2012 (0.0002)	1.5006 (0.0004)
	15	0.2121 (0.0011)	0.2004 (0.0002)	0.2064 (0.0009)	0.2004 (0.0001)
$\alpha = 0.2$	20	0.2120 (0.0008)	0.2004 (0.0001)	0.2072 (0.0007)	0.2004 (0.0001)
$\beta = 0.2$	30	0.2059 (0.0005)	0.2009 (0.0001)	0.2036 (0.0005)	0.2001 (0.0001)
	50	0.2040 (0.0004)	0.2002 (<0.0001)	0.2026 (0.0003)	0.2001 (<0.0001)
	100	0.2021 (0.0003)	0.2001 (<0.0001)	0.2014 (0.0002)	0.2000 (<0.0001)
	15	1.5591 (0.0142)	1.5812 (0.0236)	1.5260 (0.0204)	1.5115 (0.0055)
$\alpha = 1.5$	20	1.5572 (0.0106)	1.5675 (0.0191)	1.5260 (0.0156)	1.5123 (0.0050)
$\beta = 1.5$	30	1.5263 (0.0063)	1.5596 (0.0122)	1.5178 (0.0160)	1.5109 (0.0043)
	50	1.5201 (0.0043)	1.5399 (0.0067)	1.5209 (0.0063)	1.5100 (0.0035)
	100	1.5156 (0.0031)	1.5256 (0.0031)	1.5141 (0.0030)	1.5092 (0.0026)
	15	1.5942 (0.0117)	0.2198 (0.0017)	1.5348 (0.0205)	0.2012 (0.0024)
$\alpha = 1.5$	20	1.5428 (0.0066)	0.2150 (0.0011)	1.5122 (0.0192)	0.2011 (0.0021)
$\beta = 0.2$	30	1.5404 (0.0046)	0.2102 (0.0007)	1.5109 (0.0110)	0.2007 (0.0020)
	50	1.5129 (0.0059)	0.2039 (0.0005)	1.5096 (0.0061)	0.2002 (0.0016)
	100	1.5028 (0.0029)	0.2012 (0.0003)	1.5022 (0.0030)	0.2005 (0.0009)

- Scenario 3: $BS - SC(\alpha = 1.5, \beta = 1.5)$
- Scenario 4: $BS - SC(\alpha = 1.5, \beta = 0.2)$

For each scenario we used five different sample sizes ($n_1 = 15$, $n_2 = 20$, $n_3 = 30$, $n_4 = 50$ and $n_5 = 100$) and generated 1,000 datasets. In computing the Bayesian estimates we ran 50,000 iterations, with a burn-in=10,000 and thin=10. For prior information we have used two different gamma priors: i) Prior 1 is a non-informative prior with hyperparameters $a = b = c = d = 0.01$; and ii) Prior 2 is an informative prior in which the hyper-parameters was chosen in such a way that the prior mean became the expected value of the corresponding parameter. All the simulation study was performed on R software (R CORE TEAM, 2013) in a HP Proliant M530e Gen8 computer. Table 2.1 presents the posterior mean for both parameters, α and β, obtained from the Bayesian methods, as well as their Monte Carlo errors (in parentheses) for both priors. Clearly the posterior means that were calculated are really close to the real simulated values, indicating that both simulation and inference processes are satisfactory. Further, as expected, the informative prior (Prior 2) outperformed the non-informative prior (Prior 1), especially when the distribution is asymmetrical ($\alpha = 1.5$) with a low sample size.

2.5 Application

In this section we illustrate the proposed methodology to estimate the parameters of the BS-SC distribution in a real dataset that refers to the ozone concentration in New York city in 1973.

This dataset is available on `lattice` package in `R` under the name `environmental` and further details can be obtained in Bruntz et al. (1974).

Non-informative prior distributions for the parameters α and β of the BS, BS-SC and BS-t distributions, considered to be independent, were used to obtain the Bayesian estimates ($\lambda = 1/\alpha^2 \sim Gamma(0.01, 0.01)$ and $\beta \sim Gamma(0.01, 0.01)$). Moreover, for the BS-$t$ distribution it was considered the uniform distribution as a prior distribution for the inverse of ν, i.e., $1/\nu \sim U(0.1, 0.5)$ that is somewhat informative but it was necessary in order to obtain the convergence for all three parameters of the model.

Two chains were generated for each model (Figure 2.3 presents the ones related to the BS-SC distribution) by Metropolis-Hastings algorithm with 50,000 iterations, where the first 10,000 were discarded as a burn-in and it was used a thin of 10 in this case. Both chains converged according to the Gelman & Rubin criterion (GELMAN; RUBIN, 1992). Furthermore, the auto-correlation of the parameters is well controlled. Therefore, according to these indications, there is no problem on the posterior statistics.

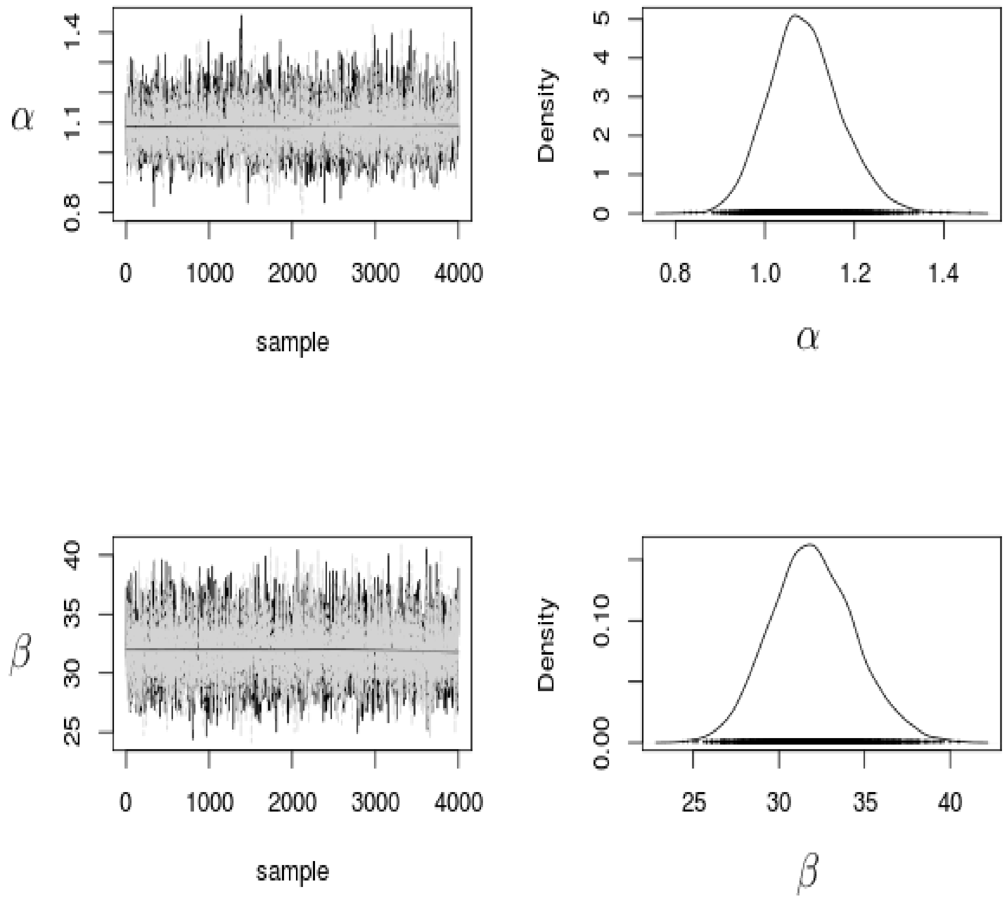


Figure 2.3 – History of generated chains and their densities of the parameters α and β from BS-SC distribution, for the ozone concentration dataset.

Table 2.2 provides posterior means, standard deviations and the 95% highest posterior density (HPD) credible intervals of the parameters of the BS-SC, BS and BS-t distributions. Moreover, Table 2.2 displays the deviance information criterion (DIC) value in order to compare these models (smaller values of DIC provide better fit), see (CARLIN; LOUIS, 2009). DIC was used since it is the most common goodness-of-fit measure in Bayesian analysis (GELMAN et al., 2013). We can observe that the parameters standard deviations for the distributions are not numerically high when compared to the posterior mean itself, excepting for ν that is actually expected. Furthermore, the HPD amplitude is not high, indicating that the parameters estimates are satisfactory (Table 2.2). Finally, we can say that the BS-SC distribution is the best model since it returned smaller value of DIC (2551.243). The fit of the BS-SC, BS and BS-t distributions, using Bayesian approach, to the dataset in study, can be seen on Figure 2.4.

Table 2.2 – Posterior means, standard deviations and 95% HPD credible intervals of parameters from the BS-SC distribution of the ozone concentration dataset

	Parameter	Estimate	Standard deviation	Lower HPD (95%)	Upper HPD (95%)	DIC
BS-SC	α	1.0880	0.0808	0.9376	1.2531	2551.243
	β	31.9780	2.4781	27.3799	37.0678	
BS	α	0.9994	0.0690	0.8701	1.1379	2631.518
	β	27.9995	2.3373	23.7091	32.9264	
BS-t	α	0.8235	0.0737	0.6808	0.9642	2592.069
	β	31.0800	2.5860	26.1370	36.2266	
	ν	8.4160	7.4744	2.5073	15.1702	

2.6 Concluding remarks

In this paper we presented the Bayesian inference as an alternative to be used in parameters estimation of the Birnbaum-Saunders Special-Case distribution since only the first moment of this distribution can be obtained, and then the frequentist approach should be avoided as some asymptotic properties are violated. We showed that there is no closed conditional posterior distributions when the gamma distribution – intuitively assumed due to the parametric spaces – with independent parameters is assumed as a prior distribution and, thus, the Metropolis-Hastings algorithm is required to generate the MCMC samples. However, as elucidated in the simulation study and in the real dataset application, the estimates for parameters α and β obtained by this approach are satisfactory. Furthermore, we showed that it is possible to use the Metropolis-Hastings algorithm for the simulation of BS-SC data in an accurate way and it possibly could be used in any model. We presented one application related to the ozone concentration in New York city showing that, despite the similarity between the BS-SC distribution and the BS standard model, the BS-SC distribution fitted better according to the deviance information criterion. Finally, the Bayesian methodology applied to this work, on estimation and data simulation, and on problems involving BS-SC distribution was shown to be extremely efficient and

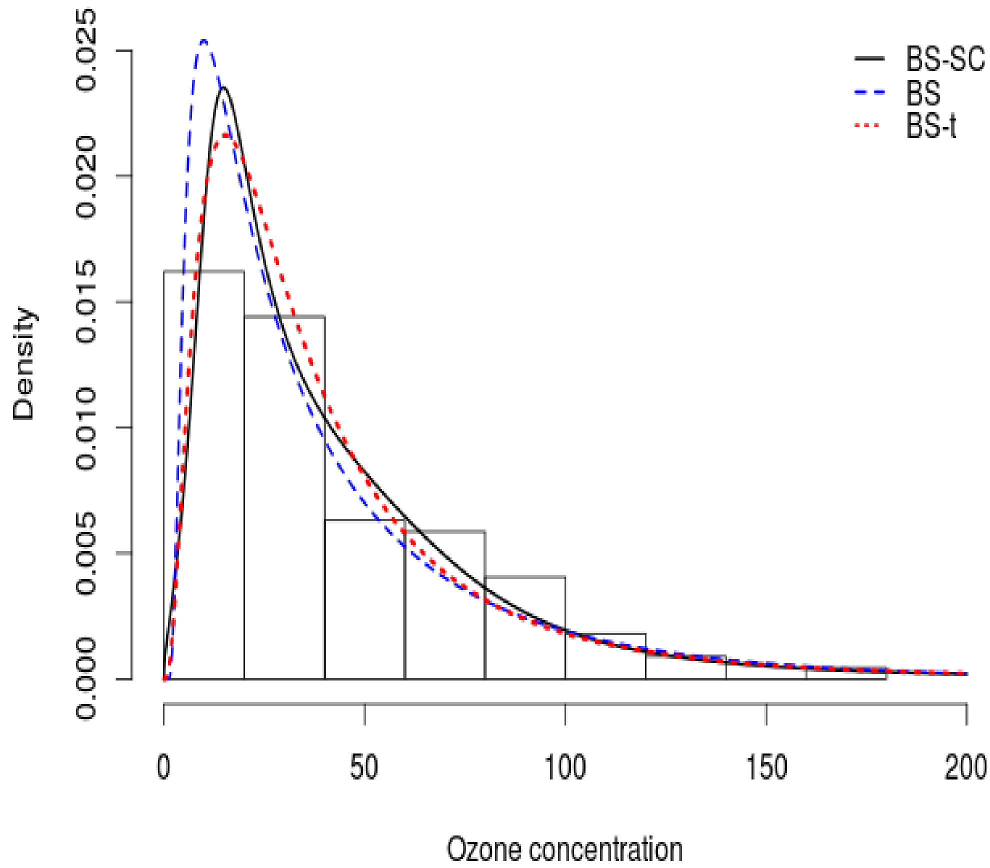


Figure 2.4 – Histogram of the ozone dataset and the fitted curve from BS-SC and BS distributions.

interesting.

References

ACHCAR, J. A. Inferences for the Birnbaum-Saunders fatigue life model using Bayesian methods. **Computational Statistics & Data Analysis**, Amsterdam, v. 15, p. 367–380, 1993.

BALAKRISHNAN, N.; GUPTA, R. C.; KUNDU, D.; LEIVA, V.; SANHUEZA, A. On some mixture models based on the Birnbaum-Saunders distribution and associated inference. **Journal of Statistical Planning and Inference**, Amsterdam, v. 141, p. 2175–2190, 2011.

BIRNBAUM, Z. W.; SAUNDERS, S. C. A new family of life distribution. **Journal of Applied Probability**, Sheffield, v. 6, p. 319–327, 1969.

BRUNTZ, S. M.; CLEVELAND, W. S.; KLEINER, B.; WARNER, J. L. The dependence of ambient ozone on solar radiation, wind, temperature, and mixing height. In: SYMPOSIUM ON ATMOSPHERIC DIFFUSION AND AIR POLLUTION, 1974, Boston. **Proceedings...** Boston: American Meterological Society, p. 125–128.

CANCHO, G. C.; ORTEGA, E. M. M.; PAULA, G. A. On estimation and influence diagnostics for log-Birnbaum-Saunders Student-t regression models: full Bayesian analysis. **Journal of Statistical Planning and Inference**, Amsterdam, v. 140, p. 2486–2496, 2010.

CARLIN, B. P.; LOUIS, T. A. **Bayesian methods for data analysis**. 3rd ed. Boca Raton: Chapman & Hall/CRC, 2009. 552 p.

DÍAZ-GARCÍA, J. A.; LEIVA, V. A new family of life distributions based on the elliptically contourned distribution. **Journal of Statistical Planning and Inference**, Amsterdam, v. 128, p. 445–457, 2005.

GELMAN, A.; RUBIN, D. B. Inference from iterative simulation using multiple sequences. **Statistical Science**, Hayward, v. 7, p. 457–472, 1992.

GUPTA, A. K.; VARGA, T. **Elliptically Contoured Models in Statistics**. Boston: Springer, 1993. 335 p.

HASTINGS, W. K. Monte Carlo sampling methods using Markov chains and their applications. **Biometrika**, Oxford, UK, v. 57, p. 97–109, 1970.

LEIVA, V.; BARROS, M.; PAULA, G. A.; SANHUEZA, A. Generalized Birnbaum-Saunders distributions applied to air pollutant concentration. **Environmetrics**, West Sussex, v. 19, p. 234–249, 2008.

MARCHANT, C.; BERTIN, K.; LEIVA, V.; SAULO, H. Generalized Birnbaum-Saunders kernel density estimators and an analysis of financial data. **Computational Statistics & Data Analysis**, Amsterdam, v. 63, p. 1–15, 2013.

R CORE TEAM. **R: A language and environment for statistical computing**. Vienna, 2013. Available at: <http://cran.r-project.org/>. Access: 01 jun. 2016.

RIECK, J. R.; NEDELMAN, J. R. A log-linear model for the Birnbaum-Saunders distribution. **Technometrics**, Richmond, VA, v. 33, p. 51–60, 1991.

VILLEGAS, C.; PAULA, G. A.; LEIVA, V. Birnbaum-Saunders mixed models for censored reliability data analysis. **IEEE Transactions on Reliability**, New York, v. 60, pp. 748–758, 2011.

3 MODELLING LOCATION, SCALE AND SHAPE PARAMETERS OF THE BIRNBAUM-SAUNDERS GENERALIZED t DISTRIBUTION[1]

Abstract

The Birnbaum-Saunders generalized t (BSGT) distribution is a very flexible family of distributions that admits different degrees of skewness and kurtosis and includes some important special or limiting cases available in the literature, such as the Birnbaum-Saunders and Birnbaum-Saunders t distributions. In this paper, we provide a regression type model to the BSGT distribution based on the generalized additive models for location, scale and shape (GAMLSS) framework. The resulting model has high flexibility and therefore a great potential to model the distribution parameters of response variables that present light or heavy tails, i.e. platykurtic or leptokurtic shapes, as functions of explanatory variables. For different parameter settings, some simulations are performed to investigate the behavior of the estimators. The potentiality of the new regression model is illustrated by means of a real motor vehicle insurance data set.

Keywords: Finance; GAMLSS; Generalized additive models; Penalized splines; Positively skewed data

3.1 Introduction

The Birnbaum-Saunders (BS) distribution is the most popular model used to describe the lifetime process under fatigue. It was proposed by Birnbaum and Saunders (1969) due to the problems of vibration in commercial aircraft that caused fatigue in the materials. This distribution is also known as the fatigue life distribution and can be used to represent failure time in various scenarios. As reported by Pescim et al. (2014), the BS distribution has received wide ranging applications in past years that include: Leiva et al. (2008) modelled the air pollutant concentration at ten monitoring stations located in different zones in Santiago, Chile; Podlaski (2008) modelled the diameter at breast height of near-natural complex structure silver fir (*Abies alba* Mill.) forests; Leiva et al. (2009) and Vilca et al. (2010) studied the level of water quality in Santiago, Chile, by modelling of hourly dissolved oxygen concentrations at four stations; Garcia-Papani et al. (2016) studied the phosphorus concentration in Cascavel, Brazil; among others.

Because of the widespread study and applications of the BS distribution, there is a need for new generalizations of this distribution. Díaz-García and Leiva (2005) proposed a family of generalized Birnbaum-Saunders (GBS) distributions based on countoured elliptical distributions such as Pearson VII and Student's t distributions, Vilca and Leiva (2006) introduced a BS model based on skew normal distributions. Gómez et al. (2009) extended the BS distribution from the slash-elliptic model. Vilca et al. (2010) and Castillo et al. (2011) developed the

[1]Luiz R. Nakamura, Robert A. Rigby (LMU), Dimitrios M. Stasinopoulos (LMU), Roseli A. Leandro (ESALQ/USP), Cristian Villegas (ESALQ/USP) and Rodrigo R. Pescim (UEL). Submitted.

epsilon-skew Birnbaum-Saunders distribution. More recently, Cordeiro and Lemonte (2011) and Pescim et al. (2014) defined the beta Birnbaum-Saunders and the Kummer beta Birnbaum-Saunders models, respectively.

Despite some of those BS generalized distributions induce asymmetry, symmetry and promote weight variation of the tail, they do not provide all these shapes in the same density function. A highly flexible model which admits light or heavy tails, shaper or flatter peaked shape and it has some important special and/or limiting cases, is the Birnbaum-Saunders generalized t (BSGT) distribution proposed by Genç (2013). This generalization of the BS distribution contains some models previously studied in the literature and, therefore, the BSGT enables to study and fit various types of data with different shapes by a unified approach.

In many practical applications, the responses are affected by explanatory variables such as the socioeconomics and school levels, blood pressure, cholesterol level, soil quality, climate, among many others. BS regression models are widely used to estimate the reliability or predict the durability of non-repairable copies of materials. Among them, Rieck and Nedelman (1991) proposed a log-linear regression model based on the BS distribution. Diagnostic analysis for the BS regression model were developed by Galea et al. (2004), Leiva et al. (2007) and Xie and Wei (2007), while the Bayesian inference was introduced by Tsionas (2001). Barros et al. (2008) proposed a class of lifetime regression models that includes the log-Birnbaum-Saunders-t (BS-t) regression models as special case. Furthermore, Lemonte and Cordeiro (2009) and Villegas et al. (2011) studied the BS nonlinear and BS mixed models, respectively. However, those BS regression models follow the same idea of many previous regression type models in the literature such as generalized linear models (NELDER; WEDDERBURN, 1972), generalized additive models (HASTIE; TIBSHIRANI, 1990) and log location-scale models (LAWLESS, 2003). These models use only the location parameter of the distribution of the response variable which is a major limitation since other parameters may need to be modelled.

In this context, Rigby and Stasinopoulos (2005) developed the generalized additive models for location, scale and shape (GAMLSS), a very general class of univariate regression models whose main advantage is that all parameters of a given distribution (that does not necessarily belong to the exponential family) can be modelled as parametric and/or additive non-parametric smooth functions of explanatory variables, which can lead to a simpler distribution for a given response variable Y, simplifying the interpretation of the problem in study. Within GAMLSS the shape of the conditional distribution of the response variable can vary according to the values of the explanatory variables, allowing great modelling flexibility.

In this paper, we introduce the BSGT distribution into the GAMLSS framework in order to provide a very flexible regression model for this family, modelling all of its four parameters using explanatory variables. The new regression model may be fitted to a data set with light or heavy tails, i.e. a platykurtic or leptokurtic response variable as, for example, the total claim amount of an insurance company. The rest of this paper is outlined as follows. Section 3.2 provides a brief review of the BSGT family of distributions. The BSGT is introduced into the

GAMLSS framework in Section 3.3. Section 3.4 shows a simulation study with different values of the parameters. A real data set application regarding insurance is provided in Section 3.5 to show the BSGT flexibility, and comparing it with well-known models. Section 3.6 ends the paper with some concluding remarks.

3.2 The BSGT distribution – a brief review

Díaz-García and Leiva (2005) proposed the GBS family of distributions defined by transformation from any random variable Z with symmetric distribution S (GUPTA; VARGA, 1993), with density given by

$$f_Z(z|\zeta,\phi,\boldsymbol{\delta}) = c\,K\left[\frac{(x-\zeta)^2}{\phi^2}\right],\ -\infty < z < \infty,$$

as

$$Y = \beta\left[\frac{\alpha Z}{2} + \sqrt{\left(\frac{\alpha Z}{2}\right)^2 + 1}\right]^2 \tag{3.1}$$

where $Z \sim S(\zeta = 0, \phi = 1, \boldsymbol{\delta})$, $\boldsymbol{\delta}$ correponds to the parameters inherited from the baseline distribution, c is the normalizing constant such that $f_Y(y)$ is a proper density, $K(\cdot)$ is the kernel of the density of Z, $\alpha > 0$ represents the shape parameter and $\beta > 0$ is the scale parameter and is also the median of the distribution. As $\alpha \to 0$, the GBS distribution becomes symmetrical around β, whereas when α grows the distribution becomes increasingly positively skewed. Its probability density function (pdf) can be expressed as

$$f_Y(y|\alpha,\beta,\boldsymbol{\delta}) = \frac{c}{2\,\alpha\,\beta^{\frac{1}{2}}}\,y^{-\frac{3}{2}}\,(y+\beta)\,K\left(\frac{1}{\alpha^2}\left[\frac{y}{\beta} + \frac{\beta}{y} - 2\right]\right),$$

for $y > 0$.

If Z has a generalized t (GT) distribution, $Z \sim GT(0,1,\nu,\tau)$, with pdf given by

$$f_Z(z|\nu,\tau) = \frac{\tau}{2\nu^{\frac{1}{\tau}}B\left(\frac{1}{\tau},\nu\right)\left(1+\frac{|z|^\tau}{\nu}\right)^{\nu+\frac{1}{\tau}}},$$

where $-\infty < z < \infty$, $\nu > 0$, $\tau > 0$ and $B(\cdot)$ is the beta function, then the random variable Y obtained from transformation (3.1) has a BSGT distribution with pdf given by

$$f_Y(y|\alpha,\beta,\nu,\tau) = \frac{\tau y^{-\frac{3}{2}}\,(y+\beta)}{4\,\alpha\,\beta^{\frac{1}{2}}\,\nu^{\frac{1}{\tau}}B\left(\frac{1}{\tau},\nu\right)}\left(1+\frac{1}{\nu\alpha^\tau}\left|\frac{y}{\beta}+\frac{\beta}{y}-2\right|^{\frac{\tau}{2}}\right)^{-\left(\nu+\frac{1}{\tau}\right)}, \tag{3.2}$$

where $y > 0$, $\alpha > 0$, $\beta > 0$, $\nu > 0$ and $\tau > 0$. As in (3.1), if the shape parameter $\alpha \to 0$, the distribution becomes near symmetrical around β and when α grows, the model becomes increasingly positively skewed; β is a scale parameter and is also the median of the distribution;

and ν and τ are the parameters related to the peak and tails of the distribution. Note that as $y \to \infty$ then $f_Y(y|\alpha,\beta,\nu,\tau) = O\left(y^{-\frac{\nu\tau}{2}-1}\right)$, the same order as a t distribution with $\nu\tau/2$ degrees of freedom. Hence, small values of the product of ν and τ result in a heavier upper tail. Similarly, larger values of the product of ν and τ will result in a lighter upper tail. Parameter τ also affects the peak of the distribution: $0 < \tau \leq 1$ results in a spiked peak in the distribution, with a sharp spike (i.e. infinite derivative) if $0 < \tau < 1$, while a larger τ results in an increasingly flatter peak.

Despite its flexibility that can combine symmetrical/asymmetrical shapes with light or heavy tails (i.e. leptokurtic or platykurtic densities), the BSGT model is important since it has some special or limiting cases already proposed in the literature such as the BS, BS-t, BS-Laplace, BS-Cauchy and BS-power exponential (BSPE) distributions, as displayed in Figure 3.1.

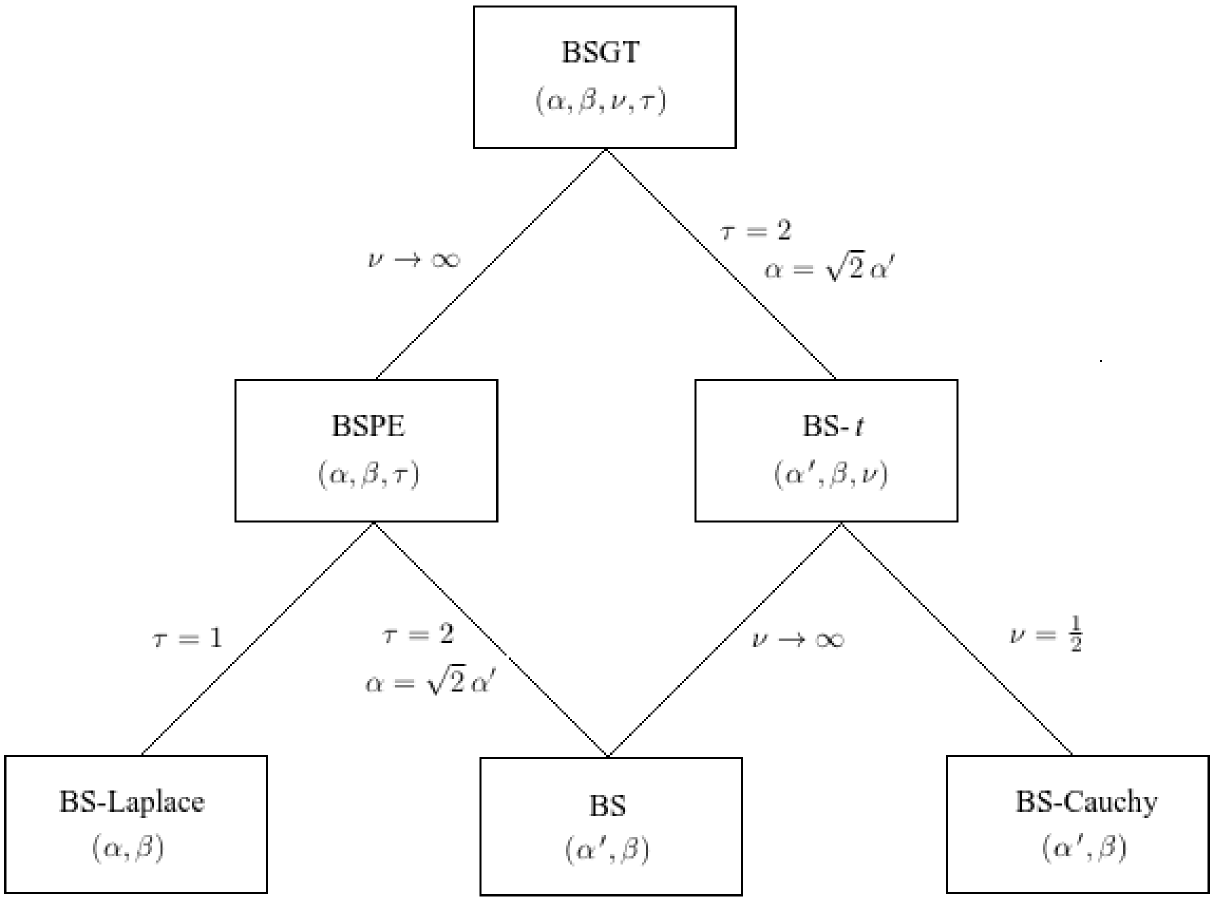


Figure 3.1 – Relationships of the BSGT special models

3.3 GAMLSS model for the BSGT distribution

GAMLSS are semi-parametric regression models that involve a distribution for the response variable (parametric part) and may involve non-parametric smoothing terms when modelling parameters of the distribution as functions of explanatory variables. The GAMLSS `R` implementation, called the `gamlss` package, includes distributions with up to four parameters that are commonly represented by μ for location, σ for scale and ν and τ for shape (RIGBY; STASINOPOULOS, 2005). Hence, for the BSGT distribution, we consider $\mu = \beta$, $\sigma = \alpha$, $\nu = \nu$ and $\tau = \tau$ to obey the established notation in GAMLSS framework in R software (STASINOPOULOS; RIGBY, 2007). Moreover, from this point, we say that a random variable Y follows a BSGT distribution, denoted by $Y \sim$ BSGT(μ, σ, ν, τ).

The GAMLSS model for the BSGT distribution assumes that conditional on its parameters ($\boldsymbol{\mu}$, $\boldsymbol{\sigma}$, $\boldsymbol{\nu}$ and $\boldsymbol{\tau}$), observations Y_i are independent BSGT$(\boldsymbol{\mu}, \boldsymbol{\sigma}, \boldsymbol{\nu}, \boldsymbol{\tau})$ variables with pdf given in (3.2), and can be expressed as

$$\begin{aligned}
g_1(\boldsymbol{\mu}) &= \eta_1 = \boldsymbol{X}_1\boldsymbol{\beta}_1 + \sum_{j=1}^{J_1} h_{j1}(\boldsymbol{x}_{j1}) \\
g_2(\boldsymbol{\sigma}) &= \eta_2 = \boldsymbol{X}_2\boldsymbol{\beta}_2 + \sum_{j=1}^{J_2} h_{j2}(\boldsymbol{x}_{j2}) \\
g_3(\boldsymbol{\nu}) &= \eta_3 = \boldsymbol{X}_3\boldsymbol{\beta}_3 + \sum_{j=1}^{J_3} h_{j3}(\boldsymbol{x}_{j3}) \\
g_4(\boldsymbol{\tau}) &= \eta_4 = \boldsymbol{X}_4\boldsymbol{\beta}_4 + \sum_{j=1}^{J_4} h_{j4}(\boldsymbol{x}_{j4}),
\end{aligned} \tag{3.3}$$

where $g_k(\cdot)$, $k = 1, 2, 3, 4$, are the link functions, $\boldsymbol{\beta}_k^\top = \left(\beta_{1k}, \ldots, \beta_{J'_k k}\right)$ denotes the parameter vector associated to explanatory variables with design matrix $\boldsymbol{X}_k$ and each h_{jk} function is a smooth non-parametric function of an explanatory variable $\boldsymbol{x}_{jk}$, being typically a smoothing spline (for more details, see e.g. HASTIE; TIBSHIRANI, 1990) or P-spline (EILERS; MARX, 1996).

3.3.1 Selecting the response variable distribution and diagnostics

Two different stages comprehend the strategy to fit a GAMLSS model: fitting and diagnostics. In the fitting stage, we fit different models using a generalized Akaike information criterion (GAIC, for more information, see VOUDOURIS et al., 2012) to compare them (the model with the smallest GAIC is selected). The Akaike information criterion (AIC) (AKAIKE, 1974) and Schwarz Bayesian criterion (SBC) (SCHWARZ, 1978) are special cases of the GAIC(k) when $k = 2$ and $k = \log(n)$, respectively.

In the diagnostic stage, we use the normalized (randomized) quantile residuals (DUNN; SMYTH, 1996) which are defined by

$$\hat{r}_i = \Phi^{-1}(\hat{u}_i),$$

where Φ^{-1} is the inverse cumulative distribution function of a standard normal variable and $\hat{u} = F_Y(y|\hat{\boldsymbol{\theta}})$ is the fitted cumulative distribution function. The main advantage of this type of residual is that its true values r_i, $i = 1, \ldots, n$ always have a standard normal distribution given the assumption that the model is correct, whatever the distribution of the response variable, i.e. if the model for the response variable is correct, the residuals have a standard normal distribution.

3.3.2 Selecting the explanatory variables

In order to select the explanatory variables for the BSGT model, we use a backward/forward algorithm implemented in `gamlss` package called `StepGAICAll.A`:

1) Step 1: select a model for μ using a forward GAIC selection procedure and fixing σ, ν and τ;

2) Step 2: select a model for σ using a forward GAIC selection procedure given the model for μ in Step 1 and fixing ν and τ;

3) Step 3: select a model for ν using a forward GAIC selection procedure given the models for μ and σ obtained in Steps 1 and 2, respectively, and fixing τ;

4) Step 4: select a model for τ using a forward GAIC selection procedure given the models for μ, σ and ν obtained in Steps 1, 2 and 3, respectively;

5) Step 5: perform a backward GAIC selection procedure to select a model for ν given the models for μ, σ and τ obtained from Steps 1, 2 and 4, respectively;

6) Step 6: perform a backward GAIC selection procedure to select a model for σ given the models for μ, ν and τ obtained from Steps 1, 5 and 4, respectively;

7) Step 7: perform a backward GAIC selection procedure to select a model for μ given the models for σ, ν and τ obtained from Steps 6, 5 and 4, respectively.

The resulting model may contain different terms for each of the parameters μ, σ, ν and τ.

3.3.3 Computational functions

In order to perform a simulation study with the BSGT distribution and a real data set application using the BSGT regression model, we implemented this family into the `gamlss` package in `R` (for more details about GAMLSS framework estimation processes, see Appendices A, B and C or RIGBY; STASINOPOULOS, 2005) and the following functions will be available in the `gamlss.dist` package:

1) `dBSGT()` gives the BSGT probability density function;

2) `pBSGT()` gives the BSGT cumulative distribution function (cdf);

3) `qBSGT()` gives the BSGT quantile function, i.e. inverse cdf; and

4) `rBSGT()` is the BSGT random number generator.

It is noteworthy that we can also fit the special and/or limiting cases of the BSGT distribution in `gamlss`, e.g. in order to fit a BS-t distribution, we use the following arguments within the main function `gamlss()`: `tau.fix=TRUE` and `tau.start=2` in order to fix $\tau = 2$ in BSGT(μ, σ, ν, τ) model.

3.4 Simulation study

We performed a simulation study generating 12 different scenarios with two sample sizes ($n = 100$ and $n = 500$) using the `rBSGT()` function. The scenarios were chosen in such a way that all possible density shapes could be covered, using as true parameter values $\mu = 50$ and

1) $\sigma = 0.5$ for near symmetrical (Table 3.1) and $\sigma = 1.5$ for very asymmetrical shapes (Table 3.2);

2) $\nu = 1.0$ for heavy-tailed and $\nu = 5.0$ for less heavy tailed;

3) $\tau = 1.5$, $\tau = 2.0$ and $\tau = 10.0$ since a low value of τ tends to a sharper peaked shape (leptokurtic), while a high value of τ tends to a flatter peaked shape (platykurtic).

The simulation study was performed in a HP Proliant M530e Gen8 Computer under a Debian Linux operating system. Tables 3.1 and 3.2 present the true simulated parameter values, average estimates (AE) and standard deviations (SD) for the estimated parameters for near symmetrical ($\sigma = 0.5$) and very asymmetrical ($\sigma = 1.5$) scenarios, respectively. The required numerical evaluations are implemented in `R` software through the `gamlss` function (STASINOPOULOS; RIGBY, 2007).

As expected, we observe (from Tables 3.1 and 3.2) that when $n = 500$ we obtain closer estimates compared to the true generating value and the SD values decrease. Moreover, we can note that the estimates of parameters ν and τ are slightly more imprecise than μ and σ which could be happening since they are often highly correlated. Also, the distribution of the parameter estimators of ν and τ are highly positively skewed.

Table 3.1 – Real parameter value, average estimates (AE) and standard deviations (SD) based on 1,000 simulations of the near symmetrical version of the BSGT distribution

		$n = 100$		$n = 500$	
Parameters	True value	AE	SD	AE	SD
		Scenario 1			
μ	50	50.040	2.616	50.000	1.065
σ	0.5	0.461	0.116	0.490	0.050
ν	1.0	1.272	0.986	1.130	0.487
τ	1.5	1.775	1.176	1.526	0.349
		Scenario 2			
μ	50	50.160	2.373	49.940	1.091
σ	0.5	0.469	0.089	0.494	0.036
ν	1.0	1.266	1.028	1.145	0.539
τ	2.0	2.319	1.459	2.053	0.523
		Scenario 3			
μ	50	49.950	1.220	50.000	0.507
σ	0.5	0.482	0.037	0.499	0.018
ν	1.0	1.030	1.996	1.232	0.953
τ	10.0	12.02	6.040	10.900	4.111
		Scenario 4			
μ	50	50.060	2.305	50.040	0.993
σ	0.5	0.494	0.078	0.5012	0.042
ν	5.0	3.790	3.640	5.263	3.277
τ	1.5	1.857	0.722	1.578	0.301
		Scenario 5			
μ	50	50.020	1.943	50.020	0.865
σ	0.5	0.484	0.059	0.5002	0.031
ν	5.0	3.695	3.749	5.388	3.659
τ	2.0	2.606	1.259	2.115	0.438
		Scenario 6			
μ	50	50.020	0.981	50.010	0.379
σ	0.5	0.473	0.032	0.494	0.013
ν	5.0	3.047	4.068	4.406	4.012
τ	10.0	10.570	3.351	11.530	2.868

Table 3.2 – Real parameter value, average estimates (AE) and standard deviations (SD) based on 1,000 simulations of the very asymmetrical version of the BSGT distribution

		$n = 100$		$n = 500$	
Parameters	True value	AE	SD	AE	SD
Scenario 7					
μ	50	51.010	8.381	50.100	3.198
σ	1.5	1.389	0.345	1.478	0.153
ν	1.0	1.245	0.970	1.151	0.502
τ	1.5	1.718	1.081	1.511	0.326
Scenario 8					
μ	50	50.240	7.263	50.120	3.061
σ	1.5	1.415	0.273	1.487	0.106
ν	1.0	1.177	1.004	1.153	0.603
τ	2.0	2.418	1.540	2.051	0.492
Scenario 9					
μ	50	50.100	3.068	50.040	1.222
σ	1.5	1.444	0.113	1.494	0.052
ν	1.0	1.056	1.872	1.322	1.288
τ	10.0	13.360	7.911	10.790	3.944
Scenario 10					
μ	50	50.42	6.671	50.09	2.753
σ	1.5	1.476	0.230	1.498	0.124
ν	5.0	4.339	4.229	5.090	3.061
τ	1.5	1.919	0.931	1.577	0.303
Scenario 11					
μ	50	50.14	5.430	49.95	2.258
σ	1.5	1.455	0.173	1.496	0.089
ν	5.0	3.815	4.008	4.965	3.120
τ	2.0	2.570	1.099	2.118	0.416
Scenario 12					
μ	50	50.05	2.678	50.020	1.011
σ	1.5	1.419	0.097	1.482	0.040
ν	5.0	3.041	8.159	4.732	5.874
τ	10.0	13.850	8.780	12.720	5.166

3.5 Application: motor vehicle insurance data

In this Section, we illustrate the usefulness of the BSGT regression model, using the GAMLSS framework, to the total claim amount (response variable, Y) from motor vehicle insurance policies over a twelve-month period in 2004–2005 (DE JONG; HELLER, 2008, p. 15). The original data set was composed of approximately 68,000 policies, but here, we used only those with at least one claim (totalling 3,911 policies). Using this reduced data set, Y ranges from 1.09 to 55,720.00, with mean=2,145.00, median=844.70, standard deviation=3,765.86, skewness=4.74 and kurtosis=38.58.

Since Y is a very positively skewed variable we used four different distributions besides the BSGT distribution which are possible suitable candidates for the response variable: the Box-Cox t (BCTo), generalized gamma (GG), inverse Gaussian (IG) distributions that are already available in `gamlss.dist` package and the BS distributions which is a special case of the BSGT distribution. The covariates used to build the models in order to explain Y are displayed in Table 3.3.

Table 3.3 – Covariates of the motor vehicle insurance data

Variable	Type	Range
Vehicle value (X_1)	Quantitative	\$0–\$139,000
Number of claims (X_2)	Factor	1, 2, 3, 4
Automobile manufacturing company (X_3)	Factor	A, B, C, D
Vehicle age (X_4)	Factor	1, 2, 3, 4 (1 is recent)
Driver gender (X_5)	Factor	male, female
Driver's area of residence (X_6)	Factor	A, B, C, D, E, F
Age band of policy holder (X_7)	Factor	1, 2, 3, 4, 5, 6 (1 is the youngest)
Amount of exposure during the year (X_8)	Quantitative	0–1

Here, we replaced X_1 by $X_1^* = \log(X_1 + 1)$ to modify the high skewness exhibited by this variable. After some previous analysis, we excluded six observations that presented $X_1 = 0$, i.e. the vehicles with value equals zero and the only two observations with $X_2 = 4$, i.e. when there were four claims, since they were considered as outliers. Finally, we fitted several regression models using the backward/forward algorithm available in Section 3.3. Moreover, we considered a P-spline (pb; for more details, see EILERS; MARX, 1996) in both quantitative covariates (X_1^* and X_8). Appropriate link functions for each of the parameters were chosen in all five distributions: when a distribution parameter θ has range $-\infty < \theta < \infty$, we used the identity link function, whereas, when $\theta > 0$ the logarithm link function was adopted.

A backward/forward selection of explanatory terms as described in Section 3.3 was performed for all parameters through `stepGAICAll.A` function in `gamlss` package (STASINOPOULOS; RIGBY, 2007) and values of global deviance (GD), Akaike information criterion

(AIC) and Schwarz information criterion (SBC) were computed in order to compare all fitted models. Table 3.4 displays those statistics from the best fitted models for each used distribution, and so, the BSGT regression model could be chosen as the more suitable model since it returned the smallest GD, AIC and SBC values (65,534.1, 65,607.1 and 65,836.0, respectively).

Table 3.4 – Statistics from the best fitted models for each used distribution

Model	GD	AIC	SBC
BSGT	65,534.1	65,607.1	65,836.0
BCTo	65,903.2	65,972.9	66,191.6
GG	65,953.8	66,019.4	66,225.2
IG	66,638.5	66,684.9	66,830.6
BS	66,205.5	66,271.5	66,478.6

The final and best model from the BSGT distribution under the GAMLSS framework (3.3) is given by

$$\begin{aligned}\log(\mu) &= 7.796 - 0.094X_1^* + 0.759(\text{if } X_2 = 2) + 1.231(\text{if } X_2 = 3)\\ &+ 0.208(\text{if } X_6 = B) + 0.215(\text{if } X_6 = C) + 0.148(\text{if } X_6 = D)\\ &+ 0.390(\text{if } X_6 = E) + 0.473(\text{if } X_6 = F) - 0.563X_8,\end{aligned}$$

$$\begin{aligned}\log(\sigma) &= 1.344 + h_{12}(X_1^*) - 0.306(\text{if } X_2 = 2) - 0.610(\text{if } X_2 = 3)\\ &+ 0.160(\text{if } X_6 = B) + 0.157(\text{if } X_6 = C) + 0.098(\text{if } X_6 = D)\\ &+ 0.129(\text{if } X_6 = E) + 0.129(\text{if } X_6 = F) - 0.397X_8,\end{aligned}$$

$$\begin{aligned}\log(\nu) &= -3.948 + 0.003(\text{if } X_4 = 2) + 0.073(\text{if } X_4 = 3) + 0.166(\text{if } X_4 = 4)\\ &- 0.086(\text{if } X_6 = B) - 0.262(\text{if } X_6 = C) - 0.442(\text{if } X_6 = D)\\ &- 0.635(\text{if } X_6 = E) - 0.856(\text{if } X_6 = F)\end{aligned}$$

and

$$\begin{aligned}\log(\tau) &= 4.747 + 0.449(\text{if } X_6 = B) + 0.475(\text{if } X_6 = C) + 1.009(\text{if } X_6 = D)\\ &+ 1.185(\text{if } X_6 = E) + 1.099(\text{if } X_6 = F). \qquad (3.4)\end{aligned}$$

We can note that four covariates were considered on the location parameter μ in the final BSGT model and both of the quantitative ones did not require any smoothing function. From the model for the median μ in (3.4), we observe that the higher is the vehicle value the lower is the median total claim amount which is somewhat unexpected. The same occurs with the exposure during the year. From the other two covariates considered in μ, we can say analyzing Figures 3.2(a) and (b) that the greater is the number of claims, greater will be the median total

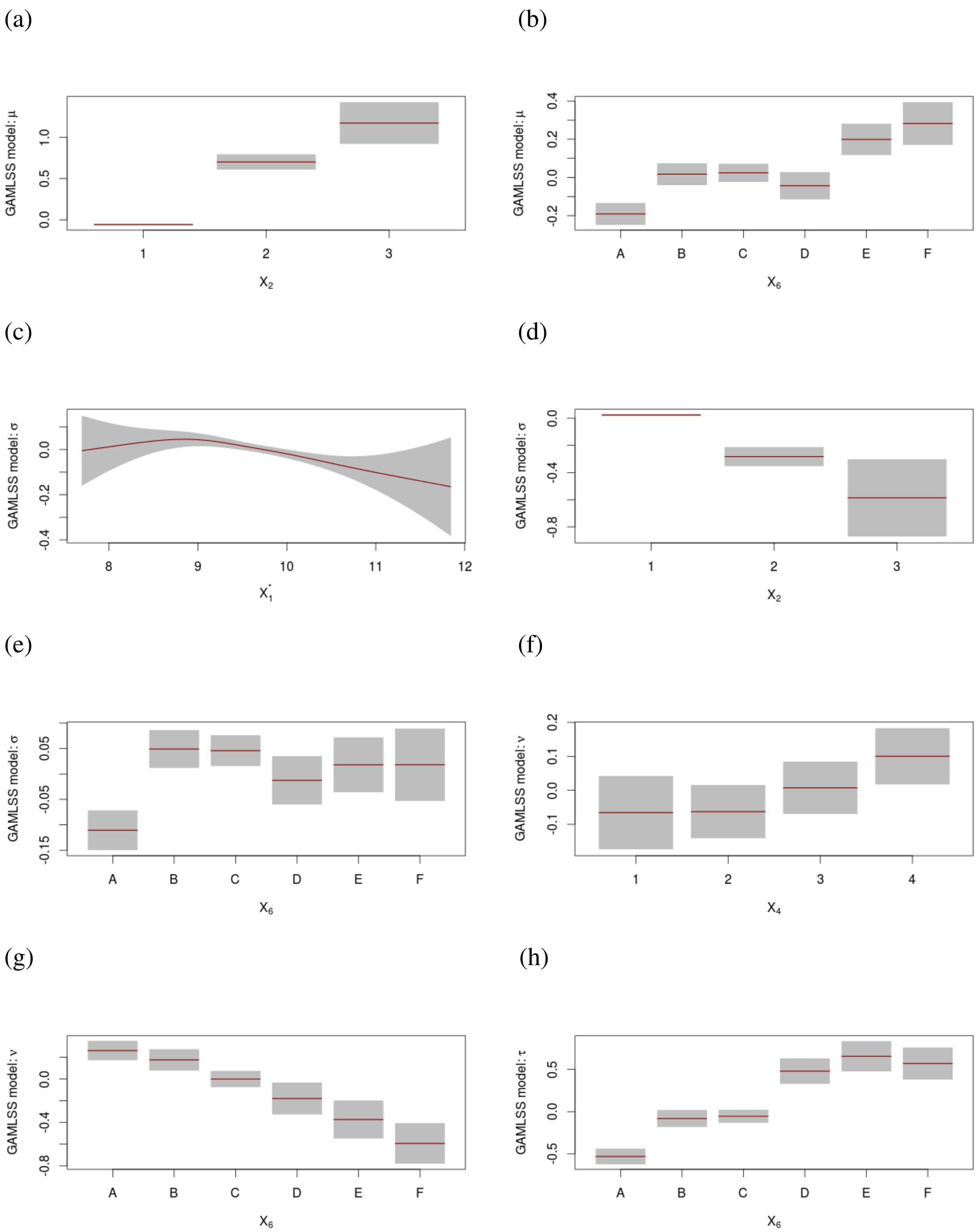


Figure 3.2 – Regression terms for parameter (a) and (b) μ; (c), (d) and (e) σ; (f) and (g) ν; and (h) τ. Note that linear relationships were omitted

claim amount and that people living in areas E and F tend to have higher median claim amounts, respectively.

As it was observed in model (3.4), we just need a smoothing function (h_{12}) to model the covariate X_1^* in σ. This relationship is showed in Figure 3.2(c) and we can note that for lower vehicle values there is a positive effect on the dispersion and after a certain point this relation becomes negative. Figures 3.2(d) and (e) present the relationship between the number of claims and driver's area of residence, respectively, with the dispersion. Further, the exposure during the year has a negative linear effect on dispersion. Figures 3.2(f)–(h) represent the relationships between selected covariates with the tails of the distribution.

Finally, the histogram and Q-Q plot of the normalized quantile residuals (DUNN; SMYTH, 1996) of model (3.4) are displayed in Figure 3.3. Figure 3.3(a), apart from one outlier, show us that the residuals adequately follow a normal distribution. Figure 3.3(b) confirms this outlier and also shows that there are a few points off the line in the high end of the range, but in general, the BSGT regression model provides a good fit to these data.

(a) (b)

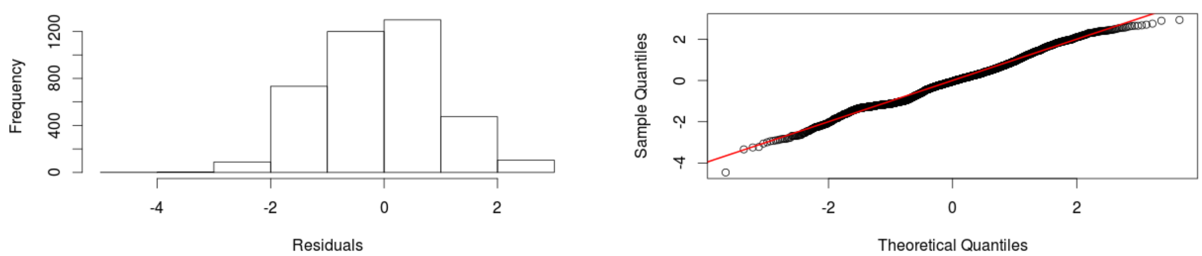


Figure 3.3 – (a) Histogram and (b) Q-Q plot of the normalized quantile residuals from the BSGT fitted regression model

3.6 Concluding remarks

In this paper, we used the Birnbaum-Saunders generalized t (BSGT) distribution proposed by Genç (2013) which admits light or heavy tails, shaper or flatter peaked shape and it has some important special cases studied in the literature. Based on this distribution, we proposed a BSGT regression model using the flexibility of the GAMLSS framework (RIGBY; STASINOPOULOS, 2005). The new regression model can be used as an alternative to model light and heavy-tailed response variables as parametric and/or additive non-parametric smooth functions of explanatory variables. Hence, this extended regression model is very flexible in many practical situations. Moreover, we conducted a simulation study using 12 different scenarios in order to cover all possible BSGT density shapes: near symmetrical and very asymmetrical, light and heavy-tailed (i.e. platykurtic and leptokurtic). We also discussed model checking analysis using the normalized quantile residuals in the new regression model fitted to a real data. An application to insurance data set demonstrated that it can be used quite effectively to provide better fits than others flexible regression models.

References

AKAIKE, H. A new look at the statistical model identification. **IEEE Transactions on Automatic Control**, New York, v. 19, p. 716–723, 1974.

BARROS, M.; PAULA, G.A.; LEIVA, V. A new class of survival regression models with heavy-tailed errors: robustness and diagnostics. **Lifetime data analysis**, Boston, v. 14, p. 316–332, 2008.

BIRNBAUM, Z.W.; SAUNDERS, S.C. (1969). A new family of life distributions. **Journal of Applied Probability**, Sheffield, v. 6, p. 319–327, 1969.

CASTILLO, N.O.; GÓMEZ, H.W.; BOLFARINE, H. Epsilon Birnbaum-Saunders distribution family: properties and inference. **Statistical Papers**, Berlin, v. 52, p. 871–883, 2011.

CORDEIRO, G.M.; LEMONTE, A.J. The beta Birnbaum-Saunders distribution: An improved distribution for fatigue life modeling. **Computational Statistics & Data Analysis**, Amsterdam, v. 55, p. 1445–1461, 2011.

DÍAZ-GARCÍA J.A.; LEIVA, V. A new family of life distributions based on the elliptically contoured distributions. **Journal of Statistical Planning and Inference**, Amsterdam, v. 128, p. 445–457, 2005.

DUNN, P.K.; SMYTH, G.K. Randomized quantile residuals. **Journal of Computational and Graphical Statistics**, Alexandria, VA, v. 5, p. 236–245, 1996.

EILERS, P.H.C.; MARX, B.D. Flexible smoothing with B-splines and penalties. **Statistical Science**, Hayward, v. 11, p. 89–102, 1996.

GALEA, M.; LEIVA, V.; PAULA, G.A. Influence diagnostics in log-Birnbaum-Saunders regression models. **Journal of Applied Statistics**, Sheffield, v. 31, p. 1049–1064, 2004.

GARCIA-PAPANI, F.; URIBE-OPAZO, M.A.; LEIVA, V.; AYKROYD, R.G. Birnbaum-Saunders spatial modelling and diagnostics applied to agricultural engineering data. **Stochastic Environmental Research and Risk Assessment**, Berlin, 2016, Available at: <http://link.springer.com/article/10.1007%2Fs00477-015-1204-4>. Access: 01 jun. 2016.

GENÇ, A.I. The generalized T Birnbaum-Saunders family. **Statistics: A Journal of Theoretical and Applied Statistics**, Hampshire, v. 47, p. 613–625. 2013.

GÓMEZ, H.W.; OLIVARES-PACHECO, J.F.; BOLFARINE, H. An extension of the generalized Birnbaum-Saunders distribution. **Statistics and Probability Letters**, Amsterdam, v. 79, p. 331-338, 2009.

GUPTA, A.K.; VARGA, T. **Elliptically Contoured Models in Statistics**. Boston: Springer, 1993. 335 p.

HASTIE, T.J.; TIBSHIRANI, R.J. **Generalized Additive Models**. London: Chapman and Hall/CRC,1990. 352 p.

JONG, P. de; HELLER, G.Z. **Generalized Linear Models for Insurance Data**. Cambridge: Cambridge University Press. 2008. 196 p.

LAWLESS, J.F. **Statistical Models and Methods for Lifetime Data**. 2nd ed. New York: Wiley, 2003. 630p.

LEIVA, V.; BARROS, M.; PAULA, G.A.; GALEA, M. Influence diagnostics in log-Birnbaum-Saunders regression models with censored data. **Computational Statistics & Data Analysis**, Amsterdam, v. 79, p. 5694–5707, 2007.

LEIVA, V.; BARROS, M.; PAULA, G.A.; SANHUEZA, A. Generalized Birnbaum-Saunders distributions applied to air pollutant concentration. **Environmetrics**, West Sussex, v. 19, p. 235–249, 2008.

LEIVA, V.; SANHUEZA, A.; ANGULO, J. M. A length-biased version of the Birnbaum-Saunders distribution with application in water quality. **Stochastic Environmental Research and Risk Assessment**, Berlin, v. 23, p. 299–307, 2009.

LEMONTE, A.J.; CORDEIRO, G.M. Birnbaum-Saunders nonlinear regression models. **Computational Statistics & Data Analysis**, Amsterdam, v. 53, p. 4441-4452, 2009.

MCDONALD, J.B.; NEWEY, W.K. Partially adaptive estimation of regression models via the generalized t distribution. **Econometric Theory**, New York, v. 4, p. 428–457, 1988.

NELDER, J.A.; WEDDERBURN, R.W.M. Generalized linear models. **Journal of the Royal Statistical Society – Series A (General)**, London, v. 135, p. 370–384, 1972.

PESCIM, R.R.; CORDEIRO, G.M.; NADARAJAH, S.; DEMÉTRIO, C.G.B.; ORTEGA, E.M.M. The Kummer beta Birnbaum-Saunders: An alternative fatigue life distribution. **Hacettepe Journal of Mathematics and Statistics**, Ankara, v. 43, 473–510, 2014.

PODLASKI, R. Characterization of diameter distribution data in near-natural forests using the Birnbaum-Saunders distribution. **Canadian Journal of Forest Research**, Ottawa, v. 38, p. 518–527, 2008.

RIECK, J.R.; NEDELMAN, J.R. A log-linear model for the Birnbaum-Saunders distribution. **Technometrics**, Richmond, VA, v. 33, p. 51–60, 1991.

RIGBY, R.A.; STASINOPOULOS, D.M. Generalized additive models for location, scale and shape, (with discussion). **Journal of the Royal Statistical Society – Series C (Applied Statistics)**, Oxford, UK, v. 54, p. 507–554, 2005.

SCHWARZ, G. Estimating the dimension of a model. **The Annals of Statistics**, Hayward, v. 6, p. 461–464, 1978.

STASINOPOULOS, D.M.; RIGBY, R.A. Generalized additive models for location scale and shape (GAMLSS) in R. **Journal of Statistical Software**, Los Angeles, v. 23, p. 1–10, 2007.

TSIONAS, E.G. Bayesian inference in Birnbaum-Saunders regression. **Communications in Statistics: Theory and Methods**, New York, v. 30, p. 179–193, 2001.

VILCA, F.; LEIVA, V. A new fatigue life model based on the family of skew-elliptical distributions. **Communications in Statistics: Theory and Methods**, New York, v. 35, p. 229–244, 2006.

VILCA, F.; SANHUEZA, A.; LEIVA, V.; CHRISTAKOS, G. An extended Birnbaum-Saunders model and its application in the study of environmental quality in Santiago, Chile. **Stochastic Environmental Research and Risk Assessment**, Berlin, v. 24, p. 771–782, 2010.

VILLEGAS, C.; PAULA, G.A.; LEIVA, V. Birnbaum-Saunders mixed models for censored reliability data analysis. **IEEE Transactions on Reliability**, New York, v. 748, p. 748–758, 2011.

VOUDOURIS, V.; GILCHRIST, R.; RIGBY, R.; SEDGWICK, J.; STASINOPOULOS, D. Modelling skewness and kurtosis with the BCPE density in GAMLSS. **Journal of Applied Statistics**, Sheffield, v. 39, p. 1279–1293, 2012.

XIE, F.C.; WEI, B.C. Diagnostics analysis for log-Birnbaum-Saunders regression models **Computational Statistics & Data Analysis**, Amsterdam, v. 51, p. 4692–4706, 2007.

4 A NEW EXTENSION OF THE BIRNBAUM-SAUNDERS DISTRIBUTION USING THE GAMLSS FRAMEWORK[1]

Abstract

In this paper, we introduce a new very flexible extension of the Birnbaum-Saunders (BS) distribution with up to six parameters, called the Birnbaum-Saunders power (BSP) distribution, which includes most of the BS type distributions already available in the literature. We provide a method for obtaining maximum likelihood estimators for its parameters and present some special cases of this new distribution family. We also introduce this new distribution into the generalized additive models for location, scale and shape (GAMLSS) framework, in order to model any or all the parameters of the distribution using parametric linear and/or non-parametric smooth functions of explanatory variables. A new generic package `gamlss.BSP` is created in `R` to fit the model. Finally, we present an application which relates the GAG concentration in children to age to illustrate the importance of the new family of distributions.

Keywords: Centile estimation; Generalized additive models; Penalized splines; R software; Regression; Skewed data

4.1 Introduction

The Birnbaum-Saunders (BS) distribution was developed by Birnbaum and Saunders (1969), motivated by problems of vibration in commercial aircraft that caused fatigue in materials, and became a very popular model to treat fatigue problems over the past few years. A positive random variable Y that follows a BS distribution with parameters $\sigma > 0$ and $\psi > 0$ is defined by the transformation

$$Y = \psi \left[\frac{Z}{2} + \sqrt{\left(\frac{Z}{2}\right)^2 + 1} \right]^2 ,$$

where $Z \sim N(0, \sigma^2)$ has a normal distribution with mean 0 and standard deviation σ. In the BS distribution, σ represents the shape parameter (denoted by α in the original parametrisation proposed by BIRNBAUM; SAUNDERS, 1969) and ψ is the scale parameter and also the median of the distribution. As $\sigma \to 0$, the BS distribution becomes symmetrical around ψ, whereas when σ grows the distribution becomes increasingly positively skewed.

In this paper, we introduce a new generalization of the BS distribution, called the Birnbaum-Saunders power (BSP) family of distributions, that admits different degrees of skewness and kurtosis and includes most of the Birnbaum-Saunders type distributions already available in the literature. Let Z follow any distribution on the real line, denoted by $Z \sim D(\boldsymbol{\theta})$, with parameter

[1]Luiz R. Nakamura, Robert A. Rigby (LMU), Dimitrios M. Stasinopoulos (LMU), Roseli A. Leandro (ESALQ/USP) and Cristian Villegas (ESALQ/USP). Submitted.

vector $\boldsymbol{\theta}$, and let

$$Y = \psi \left[\frac{Z}{2} + \sqrt{\left(\frac{Z}{2}\right)^2 + 1} \right]^{\xi}, \tag{4.1}$$

where $Y > 0$, then the distribution of Y is named here as the Birnbaum-Saunders power (BSP) distribution, where $\psi > 0$ is a scale parameter and $\xi > 0$ is a skewness parameter. For simplicity we will assume, from now on, that Z follows a distribution with up to four parameters, i.e. $\boldsymbol{\theta} = (\mu, \sigma, \nu, \tau)^{\top}$, where $-\infty < \mu < \infty$ is the location parameter, $\sigma > 0$ is the scale parameter and ν and τ are usually parameters related to the tails of the distribution of Z. However, after the transformation is performed, μ and σ are called non-centrality and shape parameters of Y respectively. Hence, the resulting BSP distribution for Y has up to six parameters and will be denoted as $Y \sim BSP(\psi, \xi, \mu, \sigma, \nu, \tau)$.

The following previous extensions of the BS distribution, proposed in the literature, due to the necessity of more flexible models than the standard BS, are special cases of (4.1). Letting $\xi = 2$ in (4.1) and assuming Z follows any symmetric distribution with parameter $\mu = 0$ gives the generalized Birnbaum-Saunders (GBS) family of distributions (DÍAZ-GARCÍA; LEIVA, 2005) for Y. The authors used eight different baseline distributions for Z: t, Pearson VII, Cauchy, special-case, Kotz type, Bessel, Laplace and logistic distributions. Later, Sanhueza et al. (2008) and Genç (2013) developed two other distributions for Y, which belong to the GBS family, using the power exponential and the generalized t distributions for Z respectively as the baseline models. Letting Z have a normal distribution with mean $\mu = 0$ in (4.1) gives the three-parameter BS distribution (DÍAZ-GARCÍA; DOMÍNGUEZ-MOLINA, 2006) for Y. Guiraud et al. (2009) and Leiva et al. (2012) included a non-centrality parameter μ in the normal and t distributions for Z, but used fixed $\xi = 2$ in (4.1), respectively.

The inverse of the one to one transformation (4.1) is given by

$$Z = \left[\left(\frac{Y}{\psi}\right)^{\frac{1}{\xi}} - \left(\frac{Y}{\psi}\right)^{-\frac{1}{\xi}} \right], \tag{4.2}$$

and hence if Y follows a BSP distribution then its probability density function (pdf) can be written as

$$f_Y(y|\psi, \xi; \boldsymbol{\theta}) = f_Z(z|\boldsymbol{\theta}) \left| \frac{dz}{dy} \right|, \; y > 0,$$

where $\boldsymbol{\theta}$ corresponds to the parameters inherited from the baseline distribution and

$$\frac{dz}{dy} = \frac{1}{y\xi} \left[\left(\frac{y}{\psi}\right)^{\frac{1}{\xi}} + \left(\frac{y}{\psi}\right)^{-\frac{1}{\xi}} \right].$$

Finally, from Equation (4.1), the exact 100α centile of Y, denoted by y_α and defined by $F_Y(y_\alpha) = p(Y \leq y_\alpha) = \alpha$, is given by

$$y_\alpha = \psi \left[\frac{z_\alpha}{2} + \sqrt{\left(\frac{z_\alpha}{2}\right)^2 + 1} \right]^{\xi}, \tag{4.3}$$

where z_α is the 100α centile of the baseline distribution. Note that for symmetric distributions of Z with parameter $\mu = 0$, then the median $y_{0.5}$ of Y is exactly ψ, since $z_{0.5} = 0$.

The rest of this paper is organized as follows. In Section 4.2 we provide some special and limiting cases of the BSP distribution. In Section 4.3 we consider the maximum likelihood method to estimate the model parameters. In Section 4.4 we discuss the GAMLSS (generalized additive models for location, scale and shape) framework, which we use to develop the BSP regression model and its inference, creating a new generic package in R software. Section 4.5 presents an application modelling the distribution of GAG concentration against age to illustrate the flexibility of the BSP distribution, comparing it with some known models in the literature. Finally, Section 4.6 ends the paper with some concluding remarks.

4.2 Special cases of the BSP distribution

The BSP family of distributions allows great flexibility in its tails since any distribution in the real line can be transformed using (4.1) and then the resulting pdf inherits some properties of the baseline model, such as light or heavy tails and sharp or flat peaks. In this section, we define some of the many novel distributions that belong to the BSP family, highlighting some of their special or limiting cases.

4.2.1 BSPNO distribution

Let Z follow a normal distribution with parameters mean μ and standard deviation σ, i.e. $Z \sim N(\mu, \sigma^2)$ in (4.1). The resulting random variable Y, denoted here by $Y \sim BSPNO\,(\psi, \xi, \mu, \sigma)$, has the Birnbaum-Saunders power normal (BSPNO) distribution with pdf given by

$$f_Y(y|\psi, \xi, \mu, \sigma) = \frac{1}{\xi\sqrt{2\pi\sigma^2 y}}\left[\left(\frac{y}{\psi}\right)^{\frac{1}{\xi}} + \left(\frac{y}{\psi}\right)^{-\frac{1}{\xi}}\right]\exp\left\{-\frac{1}{2}\left(\frac{z-\mu}{\sigma}\right)^2\right\},$$

where z is given by (4.2). For $\xi = 2$, we have the non-central Birnbaum-Saunders distribution (GUIRAUD et al., 2009). If $\mu = 0$ we have the three-parameter Birnbaum-Saunders distribution (DÍAZ-GARCÍA; DOMÍNGUEZ-MOLINA, 2006). Finally, if $\xi = 2$ and $\mu = 0$ we have the standard BS distribution (BIRNBAUM; SAUNDERS, 1969).

4.2.2 BSPT distribution

If in equation (4.1) Z has a scaled and shifted t distribution with its location, scale and degrees of freedom parameters μ, σ and ν, respectively, then Y follows a Birnbaum-Saunders power t (BSPT) distribution with pdf given by

$$f_Y(y|\psi, \xi, \mu, \sigma, \nu) = \frac{\Gamma\left(\frac{\nu+1}{2}\right)}{\xi\sigma\sqrt{\nu\pi}\,\Gamma\left(\frac{\nu}{2}\right) y}\left[\left(\frac{y}{\psi}\right)^{\frac{1}{\xi}} + \left(\frac{y}{\psi}\right)^{-\frac{1}{\xi}}\right]\left(1 + \frac{(z-\mu)^2}{\nu\sigma^2}\right)^{-\left(\frac{\nu+1}{2}\right)},$$

where z is given by (4.2), denoted here as $Y \sim BSPT(\psi, \xi, \mu, \sigma, \nu)$. If $\xi = 2$ the model is reduced to the non-central Birnbaum-Saunders t distribution (LEIVA et al., 2012) and if also

$\mu = 0$ we have the standard Birnbaum-Saunders t (BST) distribution (DÍAZ-GARCÍA; LEIVA et al., 2005).

4.2.3 BSPGT distribution

The Birnbaum-Saunders power generalized t (BSPGT) distribution is obtained from (4.1) by letting Z have a generalized t distribution (MCDONALD; NEWEY, 1988), and its resulting pdf is given by

$$
\begin{aligned}
f_Y(y|\psi,\xi,\mu,\sigma,\nu,\tau) &= \frac{\tau}{2\xi\sigma\nu^{\frac{1}{\tau}}B\left(\frac{1}{\tau},\nu\right)y}\left[\left(\frac{y}{\psi}\right)^{\frac{1}{\xi}}+\left(\frac{y}{\psi}\right)^{-\frac{1}{\xi}}\right] \\
&\times \left(1+\frac{|z-\mu|^{\tau}}{\nu\sigma^{\tau}}\right)^{-\left(\nu+\frac{1}{\tau}\right)},
\end{aligned}
$$

where z is given by (4.2), denoted here by $Y \sim BSPGT(\psi,\xi,\mu,\sigma,\nu,\tau)$. When $\tau \leq 1$ the resulting density presents a spike. For $\mu = 0$ and $\xi = 2$, we have the BSGT family of distributions (GENÇ, 2013) which has several special or limiting cases, e.g. $\nu \to \infty$ gives the Birnbaum-Saunders power exponential distribution, while $\tau = 2$ gives a reparametrization of the BST distribution.

4.2.4 Plots of the BSP probability density functions

Figure 4.1 (a) and (b) display the BSPNO distribution with $\sigma = 0.5$ and $\sigma = 1.5$ respectively, while Figure 4.1 (c) and (d) display the BSPT distribution with $\sigma = 0.5$ and $\sigma = 1.5$, respectively. Figure 4.2 displays the BSPGT distribution. Panels (a) and (b) display $\sigma = 0.5$ and $\sigma = 1.5$, respectively, and $\tau = 1$, while panels (c) and (d) display $\sigma = 0.5$ and $\sigma = 1.5$, respectively, and $\tau = 5$. Within each of the panels in Figures 4.1 and 4.2, the pdf for all combinations of $\xi = 1, 2, 3$ and $\mu = -0.5, 0.5$ are plotted. As we can see, the non-centrality parameter μ affects the location of the distribution while, as described after (4.1), ξ may be interpreted as a skewness parameter. Some different shapes are not presented here due to the limited space, e.g. all BSP distributions present bimodality when $\xi \to 0$ and σ is large.

4.3 Inference

Let Y be a random variable following a BSP distribution with parameter vector $\boldsymbol{\zeta} = (\psi, \xi; \boldsymbol{\theta})^{\top}$. Let $\boldsymbol{y} = (y_1, \ldots, y_n)^{\top}$ be a random sample from a BSP distribution. The log-likelihood function for $\boldsymbol{\zeta}$ is given by

$$
l(\boldsymbol{\zeta};\boldsymbol{y}) = \sum_{i=1}^{n}\log f_Y(y_i|\boldsymbol{\zeta}) = \sum_{i=1}^{n}\log f_Z(z_i|\boldsymbol{\theta}) + \sum_{i=1}^{n}\log\left(\left|\frac{dz_i}{dy_i}\right|\right), \tag{4.4}
$$

where the relationship between z and y is given by (4.2). Note that only the first term in (4.4) involves parameters $\boldsymbol{\theta}$, while both terms involve parameters ψ and ξ through z_i.

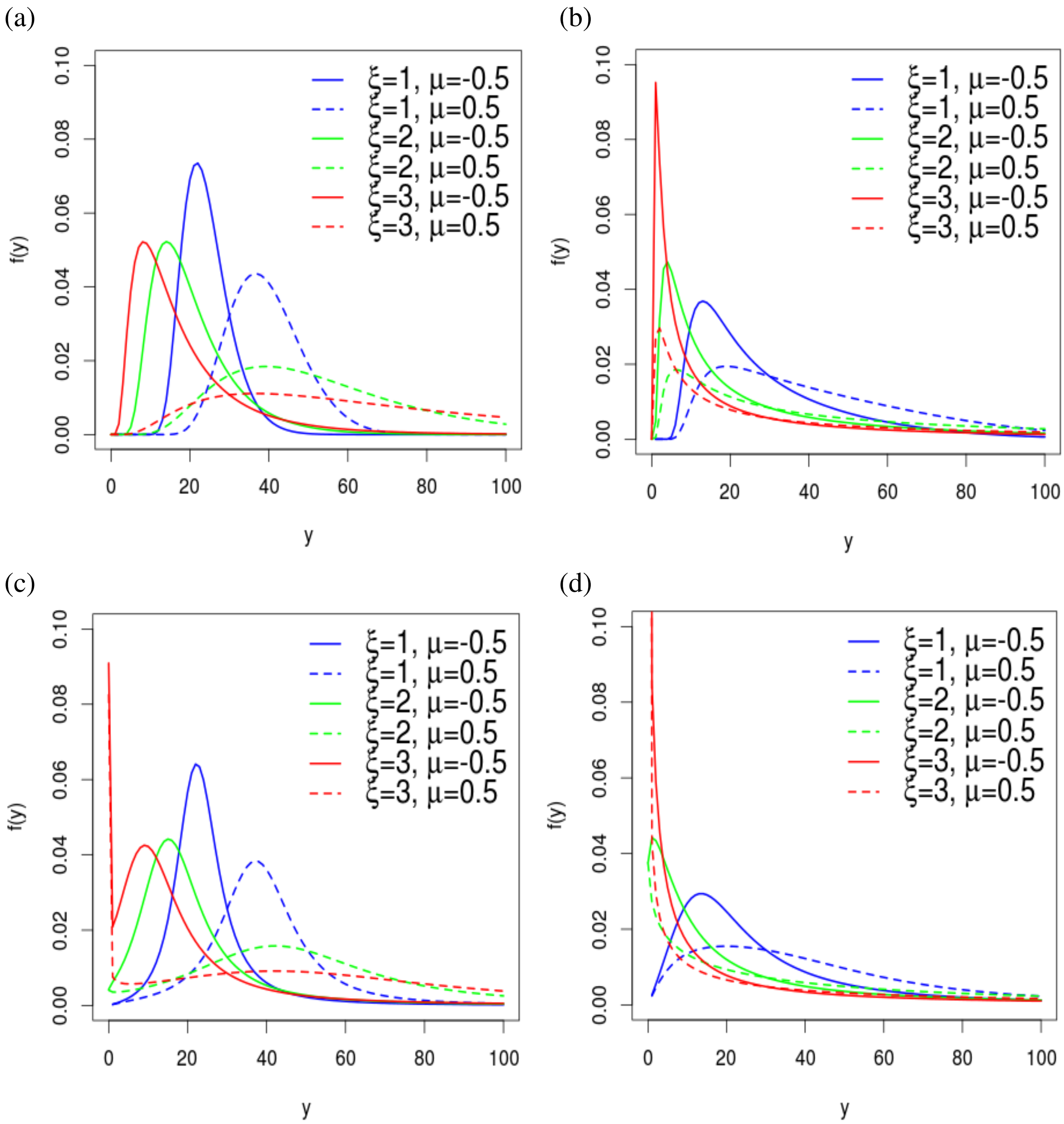


Figure 4.1 – Plots of the (a) $BSPNO(30, \xi, \mu, 0.5)$; (b) $BSPNO(30, \xi, \mu, 1.5)$; (c) $BSPT(30, \xi, \mu, 0.5, 2)$; (d) $BSPT(30, \xi, \mu, 1.5, 2)$

The elements of the score vector of $\boldsymbol{\zeta}$ are obtained from (4.4) and are given by

$$U_\psi(\boldsymbol{\zeta}) = \frac{\partial l(\boldsymbol{\zeta}, \boldsymbol{y})}{\partial \psi} = -\frac{1}{\xi}\left[\sum_{i=1}^{n}\left(\frac{y_i^{\frac{1}{\xi}}}{\psi^{\frac{1}{\xi}+1}} + \frac{\psi^{\frac{1}{\xi}-1}}{y_i^{\frac{1}{\xi}}}\right)\frac{\partial l_{z_i}}{\partial z_i} + \frac{1}{\psi}\sum_{i=1}^{n}\frac{\psi^{\frac{2}{\xi}} - y_i^{\frac{2}{\xi}}}{y_i^{\frac{2}{\xi}} + \psi^{\frac{2}{\xi}}}\right],$$

$$\begin{aligned} U_\xi(\boldsymbol{\zeta}) = \frac{\partial l(\boldsymbol{\zeta}, \boldsymbol{y})}{\partial \xi} &= -\frac{1}{\xi^2}\sum_{i=1}^{n}\log\left(\frac{y_i}{\psi}\right)\left[\left(\frac{y_i}{\psi}\right)^{\frac{1}{\xi}} + \left(\frac{y_i}{\psi}\right)^{-\frac{1}{\xi}}\right]\frac{\partial l_{z_i}}{\partial z_i} - \frac{n}{\xi} \\ &- \frac{1}{\xi^2}\sum_{i=1}^{n}\frac{y_i^{\frac{2}{\xi}}\log\left(\frac{y_i}{\psi}\right) + \psi^{\frac{2}{\xi}}\log\left(\frac{\psi}{y_i}\right)}{y_i^{\frac{2}{\xi}} + \psi^{\frac{2}{\xi}}}, \end{aligned}$$

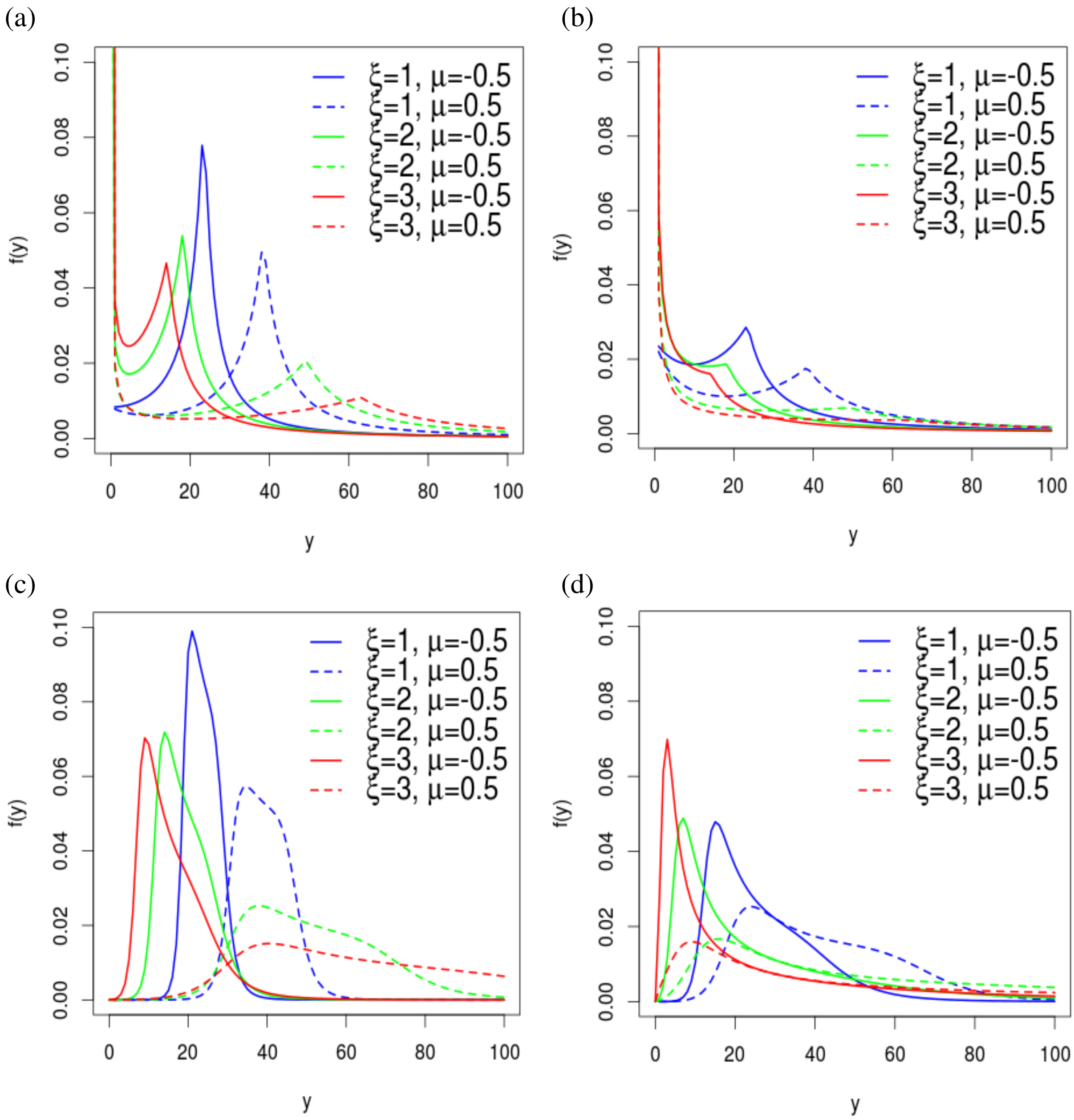


Figure 4.2 – Plots of the (a) $BSPGT(30, \xi, \mu, 0.5, 1, 1)$; (b) $BSPGT(30, \xi, \mu, 1.5, 1, 1)$; (c) $BSPGT(30, \xi, \mu, 0.5, 1, 5)$; (d) $BSPGT(30, \xi, \mu, 1.5, 1, 5)$

and

$$U_{\boldsymbol{\theta}_h}(\boldsymbol{\zeta}) = \frac{\partial l(\boldsymbol{\zeta}, \boldsymbol{y})}{\partial \theta_h} = \sum_{i=1}^{n} \frac{\partial l_{z_i}}{\partial \theta_h},$$

where θ_h, $h = 1, 2, 3, 4$, are the elements of the parameter vector $\boldsymbol{\theta}$ for a four parameter distribution $D(\boldsymbol{\theta})$ and $l_{z_i} = \log f_Z(z_i|\boldsymbol{\theta})$.

The maximum likelihood estimate $\hat{\boldsymbol{\zeta}}$ of $\boldsymbol{\zeta}$ is obtained solving the equations $U_\psi(\boldsymbol{\zeta}) = 0$, $U_\xi(\boldsymbol{\zeta}) = 0$ and $U_{\boldsymbol{\theta}_h}(\boldsymbol{\zeta}) = 0$, for $h = 1, 2, 3, 4$. These equations cannot be solved analytically and thus need to be solved numerically, e.g. using the `optim` function implemented in `R` software (R CORE TEAM, 2013).

For hypothesis testing and confidence interval estimation of the model parameters $\boldsymbol{\zeta} = (\psi, \xi; \boldsymbol{\theta})^\top$, we can use the $k \times k$ observed information matrix $J(\boldsymbol{\zeta})$, where k is the number of parameters of Y. Under regularity conditions, including that the true parameter vector $\boldsymbol{\zeta}$ lies in the interior of the parameter space for $\boldsymbol{\zeta}$ and not at the boundary, the asymptotic distribution of $\sqrt{n}(\hat{\boldsymbol{\zeta}} - \boldsymbol{\zeta})$ is $N_k(\mathbf{0}, \boldsymbol{I}(\boldsymbol{\zeta})^{-1})$, where $\boldsymbol{I}(\boldsymbol{\zeta})$ is the expected information matrix. In practice $\boldsymbol{I}(\boldsymbol{\zeta})$ is replaced by $\boldsymbol{J}(\hat{\boldsymbol{\zeta}})$, the observed information matrix evaluated at $\hat{\boldsymbol{\zeta}}$. Hence approximate standard errors can be calculated and used to construct Wald tests and Wald confidence intervals for the distribution parameters. However, generalized likelihood ratio tests and profile likelihood confidence intervals (AITKIN et al., 2009) for parameter values are more reliable.

It is noteworthy that when $Y \sim BSPGT(\psi, \xi, \mu, \sigma, \nu, \tau)$, then for all its special or limiting cases where $0 < \tau \leq 1$ the regularity conditions assumed for asymptotic normality of the maximum likelihood estimators are not valid, since the likelihood function is not always differentiable since it has spikes, and thus it is not reliable to use the tests and confidence intervals for parameters described above.

4.4 GAMLSS framework

In order to provide a regression model for the BSP family of distributions we used the generalized additive models for location, scale and shape, GAMLSS, framework (RIGBY; STASINOPOULOS, 2005). GAMLSS are semi-parametric regression models that involve a distribution for the response variable (parametric part) and may involve parametric linear and/or non-parametric smoothing terms when modelling any or all of the parameters of the distribution as functions of explanatory variables. This approach is being widely used in different fields, such as in long-term survival models (de CASTRO et al., 2010), economics (MATSUMOTO et al., 2012), and natural sciences (ZHANG et al., 2015), among others.

Generically, let $Y \sim D(\boldsymbol{\theta})$, where D represents the response variable distribution that does not necessarily belong to the exponential family and $\boldsymbol{\theta}$ is its vector of parameters of length p. The GAMLSS model can be defined as

$$g_k(\theta_k) = \eta_k = \boldsymbol{X}_k \boldsymbol{\beta}_k + \sum_{j=1}^{J_k} h_{jk}(\boldsymbol{x}_{jk}), \tag{4.5}$$

for $k = 1, \ldots, p$, where $g_k(\cdot)$ denote the known monotonic link functions, $\boldsymbol{\beta}_k^\top = (\beta_{1k}, \ldots, \beta_{J'_k k})$ is the parameter vector associated with the explanatory variables in design matrix $\boldsymbol{X}_k$ and each h_{jk} function is a smooth non-parametric function of an explanatory variable $\boldsymbol{x}_{jk}$, being typically a penalized spline (P-splines) (EILERS; MARX, 1996) or a smoothing spline (e.g. HASTIE; TIBSHIRANI, 1990). We can see from (4.5) that any distribution parameter can be modelled as a function of explanatory variables. If no covariates are selected in the model for the k^{th} parameter then it is modelled as a constant.

Implementation of fitting a BSP regression model is achieved using a new generic package, called `gamlss.BSP`, that allows the BSP distribution for Y to be fitted (with parametric and/or nonparametric functions of explanatory variables) for any corresponding distribution for Z on the real line currently available in the `gamlss.family` package. The following functions are available in the `gamlss.BSP` package:

- `BSP.d`: creates the BSP probability density function;
- `BSP.p`: creates the BSP cumulative density function (cdf);
- `BSP.q`: creates the BSP quantile function, i.e. inverse cdf;
- `BSP.r`: creates the BSP random number generator; and
- `gen.BSP`: automatically generates the new BSP distribution and the four previous functions;
- `gamlss.BSP`: the main function of the package, fits a BSP model;
- `centiles.BSP`: produces centile curves for the fitted BSP model.

For further details about the GAMLSS framework estimation processes, see Appendices A, B and C and Rigby and Stasinopoulos (2005).

4.5 Application

In order to illustrate their usefulness, in this section we fit some BSP regression models, using the GAMLSS framework, to data on the concentration of a chemical GAG in the urine of 314 children between the ages 0 and 17 (VENABLES; RIPLEY, 2002; available in `R` software under the name `GAGurine` in package `MASS`). The response variable GAG ranges from 1.8 to 56.3, with mean= 13.17, median= 10.60, standard deviation=8.99, skewness= 1.60 and kurtosis= 6.40. The explanatory variable is age in years of the child.

Since GAG is a positively skewed response variable, we used five different distributions, besides the BSPNO, BSPT and BSPGT distributions, that are possible suitable candidates for the response variable distribution: Box-Cox Cole and Green (BCCGo) (COLE; GREEN, 1992), Box-Cox power exponential (BCPEo) (RIGBY; STASINOPOULOS, 2004), Box-Cox t (BCTo) (RIGBY; STASINOPOULOS, 2006), gamma (GA) and inverse Gaussian (IG) distributions which are already available in `gamlss.dist` package (STASINOPOULOS; RIGBY, 2007).

In order to model the distribution of the response variable GAG, we used the age of the child as an explanatory variable. Preliminary analysis indicated a single outlier (GAG=1.8 at age 7.07 years) which was removed and also indicated that a transformation of the explanatory variable age was needed. The transformation $x = \log(\text{age} + 0.01)$ was used.

We modelled the dependence of each (link transformed) parameter (of each distribution for GAG) on x (the transformed age) using a penalized smoothing spline, i.e. P-spline (for further details, see EILERS; MARX, 1996). Moreover, appropriate link functions for each of the parameters were chosen in all fitted models: when $-\infty < \theta_k < \infty$ we used the identity link function, whereas when $\theta_k > 0$ the logarithm link function was used.

Table 4.1 displays the number of distribution parameters, the total (effective) degrees of freedom (df) used in the model, the values of the global deviance (GD, equals twice the fitted log likelihood) and Akaike information criterion (AIC) (AKAIKE, 1974) for each model, which were used to compare the fitted models. The BSPT model outperformed all others since it returned the lowest AIC value (1539.29).

Table 4.1 – Statistics from the fitted models

Model	Parameters	df	GD	AIC
BSPT	5	19.38	1500.54	1539.29
BCCGo	3	15.97	1509.23	1541.17
BCTo	4	18.58	1507.28	1544.45
BSPNO	4	18.41	1508.32	1545.15
IG	2	13.00	1525.95	1551.95
BSPGT	6	20.61	1507.72	1552.79
BCPEo	5	25.73	1504.12	1555.58
GA	2	12.26	1544.44	1568.97

The fitted models for ψ, ξ, μ, σ and ν for the BSPT model are displayed in Figure 4.3. The fitted model for the scale parameter ψ indicates that the value of this parameter increases very rapidly in children with age from 0 to 0.03, after which ψ decreases until the age 17 years. The relationship between parameter ξ and variable age shows that in children with age from 0 to 0.1 the parameter ξ grows very rapidly, slightly decreases until age 0.5 years and then starts to increase gradually until 17 years old. The spikes in the early ages on the fitted models for ψ and ξ are probably a result of a few low values of GAG at ages very close to 0 (see Figure 4.5), which may indicate the necessity of observing more data points in this age region. The fitted model for the non-centrality parameter μ indicates that as the age grows, the parameter μ declines from positive to negative. It is noteworthy here that the standard BST distribution (i.e. BSPT with $\xi = 2$ and $\mu = 0$) does not fit as well to this data set (AIC$= 1548.88$) since ξ is always well below 2 and the parameter μ changes from well above 0 to well below 0 as the age increases. The fitted value for the shape parameter σ declines from 1.42 to 1.25 with age. Finally, the fitted value of the tail parameter ν increases with age, from below 1 to above 6. In the smaller ages the fitted distribution is very heavy tailed (including the Birnbaum-Saunders power Cauchy distribution, i.e. $\nu = 1$).

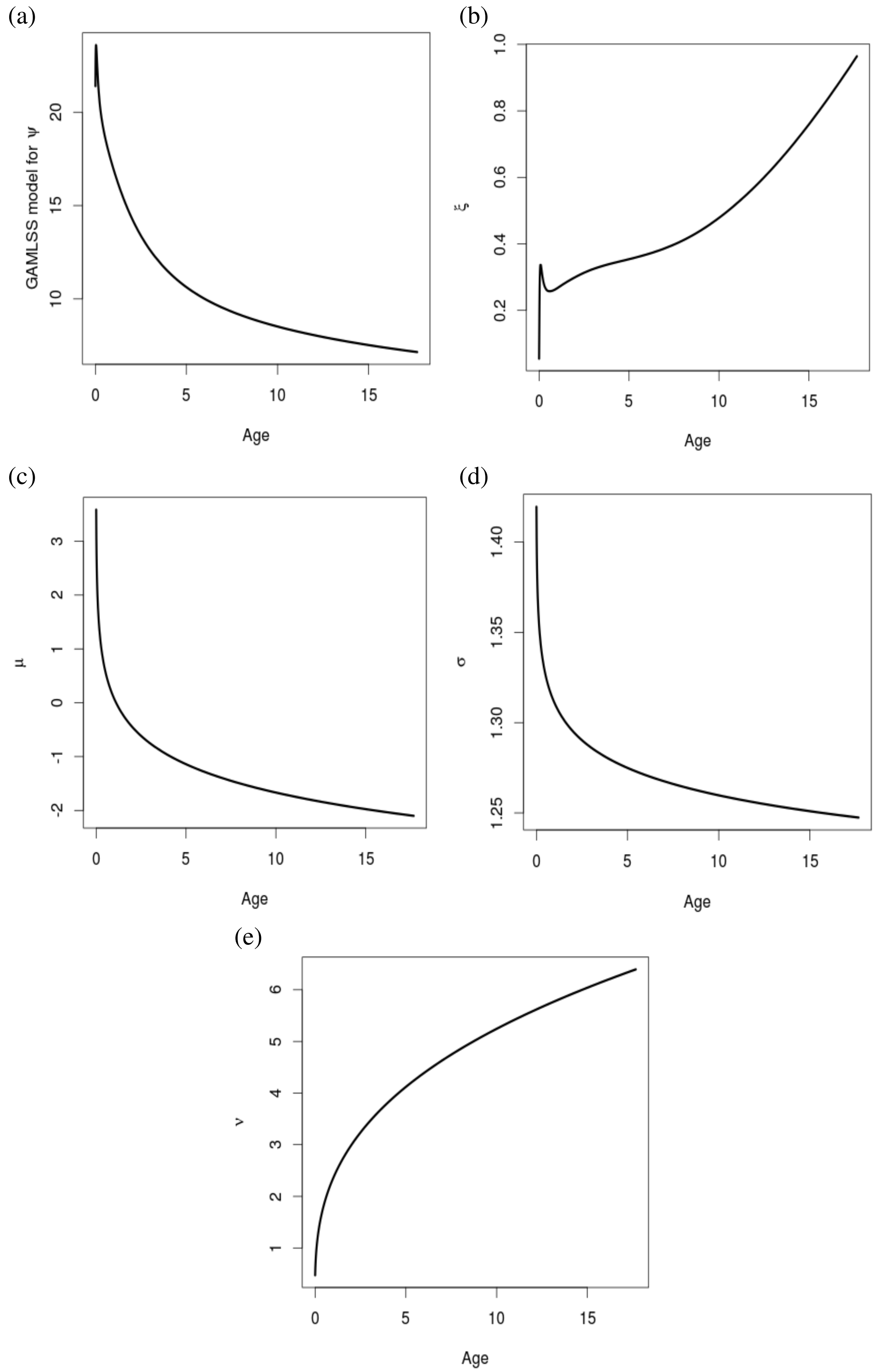


Figure 4.3 – The fitted parameters from the BSPT model against age: (a) ψ; (b) ξ; (c) μ; (d) σ; and (e) ν

Figure 4.4 displays the (normalized quantile) residuals (DUNN; SMYTH, 1996) from the BSPT model. If the model for the response variable is correct, then the residuals have a standard normal distribution. Panel (a) plots the residuals against age, whereas panels (b) and (c) display a kernel density estimate for the residuals and a simulated envelope, respectively. The residuals adequately follow a normal distribution and appear random.

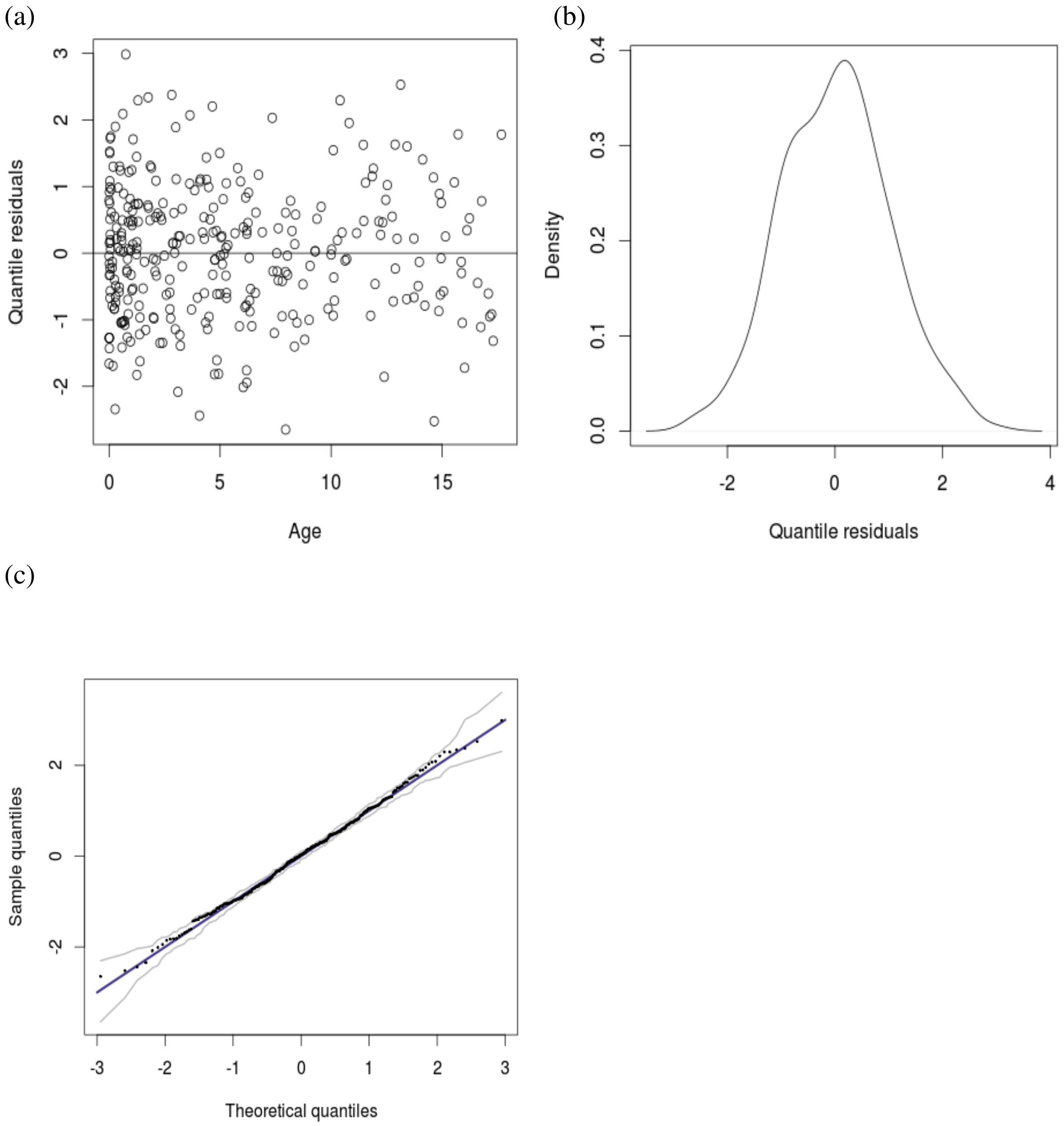


Figure 4.4 – The residuals from the BSPT model: (a) against age; (b) kernel density estimate; and (c) simulated envelope

Figure 4.5 displays seven fitted centile curves, defined by equation (4.3), for concentration of chemical GAG against age for the fitted BSPT model, with centiles 100α=5, 10, 25, 50, 75, 90, 95. For clarity of presentation they are plotted for age ranging from 0 to 1 year in Figure

4.5(a) and for age ranging from 1 to 17 years in Figure 4.5(b).

(a) (b)

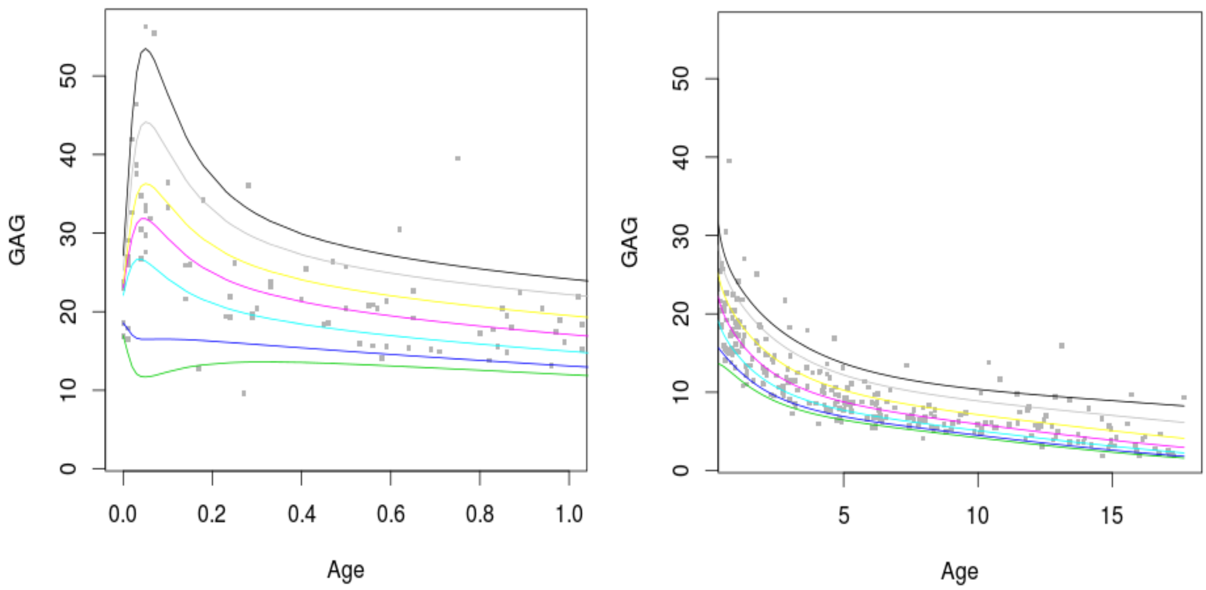


Figure 4.5 – Observed chemical GAG, with seven fitted centile curves (5, 10, 25, 50, 75, 90, 95) from the fitted BSPT model, against age: (a) 0–1 years and (b) 1–17 years

Figure 4.6 shows the worm plots (VAN BUUREN; FREDRIKS, 2001), that are used as a residual-based diagnostic. In the plot the cases are split into nine age intervals with equal number of cases. Worm plots are detrended normal Q-Q residual plots for cases in each of the nine age intervals and their different shapes can indicate whether the assumed distribution for the response variable is reasonable or not for a particular age interval: a vertical shift, a slope, a parabola or a S shape, indicate a misfit in the mean, variance, skewness and excess kurtosis of the residuals, respectively (VAN BUUREN; FREDRIKS, 2001). The nine plots are read in rows from the bottom left plot to the top right plot and correspond to the nine age intervals given above the worm plot from lowest to highest age.

As can be seen in Figure 4.6, there is a slight problem in the 8^{th} age interval (corresponding to ages 9.055–12.895 years) where we note a vertical shift above the horizontal origin line, indicating that the location of the distribution of GAG, in this age interval, is too low. However, since at least 95% of the points in each plot lie between the elliptical 95% pointwise interval band curves, we can say that the model is adequate.

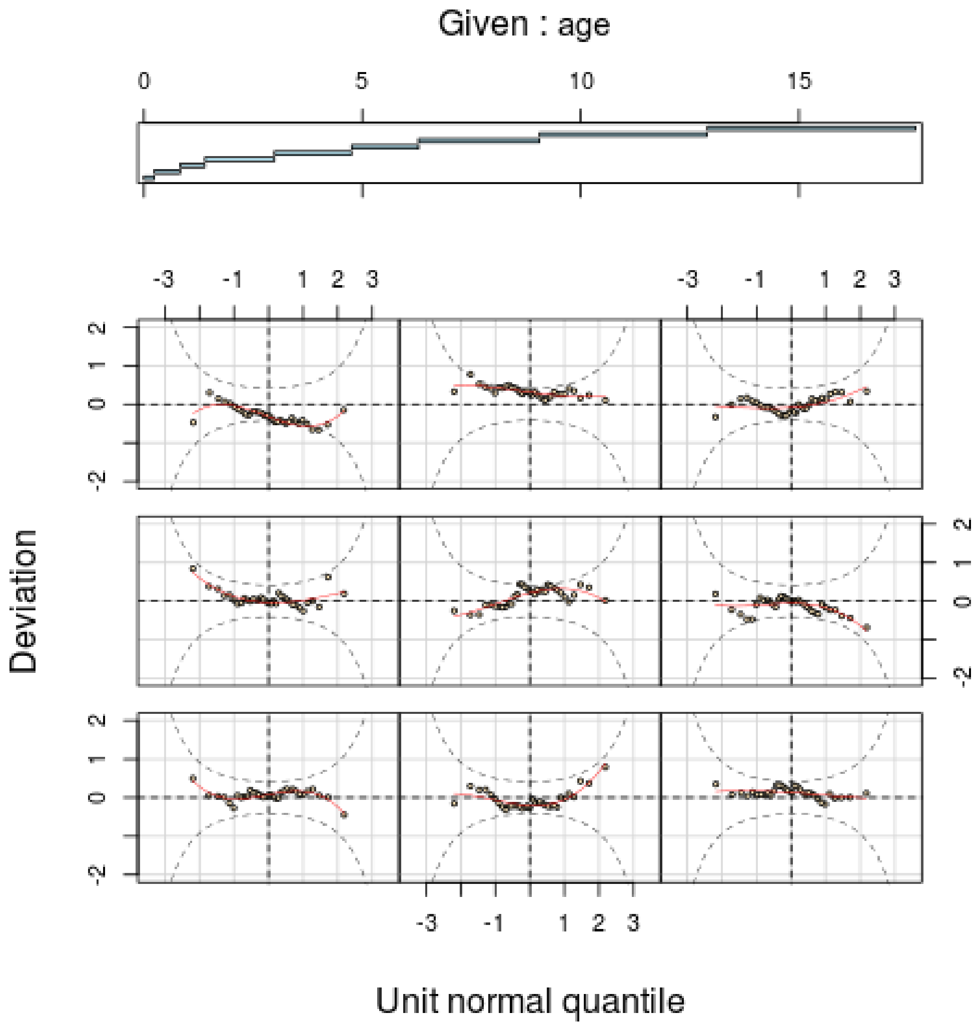


Figure 4.6 – Worm plot for the fitted BSPT model

The last residual-based diagnostic used in this paper is the Z statistics, that are useful to test whether the residuals have a standard normal distribution, since it tests whether the mean, variance, skewness and excess kurtosis of the residuals are 0, 1, 0 and 0, respectively, within each age interval group (further information about the statistical tests used can be found in ROYSTON; WRIGHT, 2000; D'AGOSTINO et al., 1990). Figure 4.7 presents the visual display of the four statistical tests for nine age intervals (which are exactly the same age intervals as presented in Figure 4.6), for the fitted BSPT model. The interpretation of Figure 4.7 is as follows: the first (Z_1), second (Z_2), third (Z_3) and fourth (Z_4) columns represent the test statistics for the mean, variance, skewness and kurtosis of the residuals (for the nine age intervals), respectively. The colors blue and red represent whether the test statistic Z is negative or positive, respectively. The larger the value of $|Z|$, the larger is the circle. Finally, a square within a circle indicates that $|Z| > 1.96$ suggesting a misfit of the model to the response variable within the corresponding inverval of age (HOSSAIN et al., 2016).

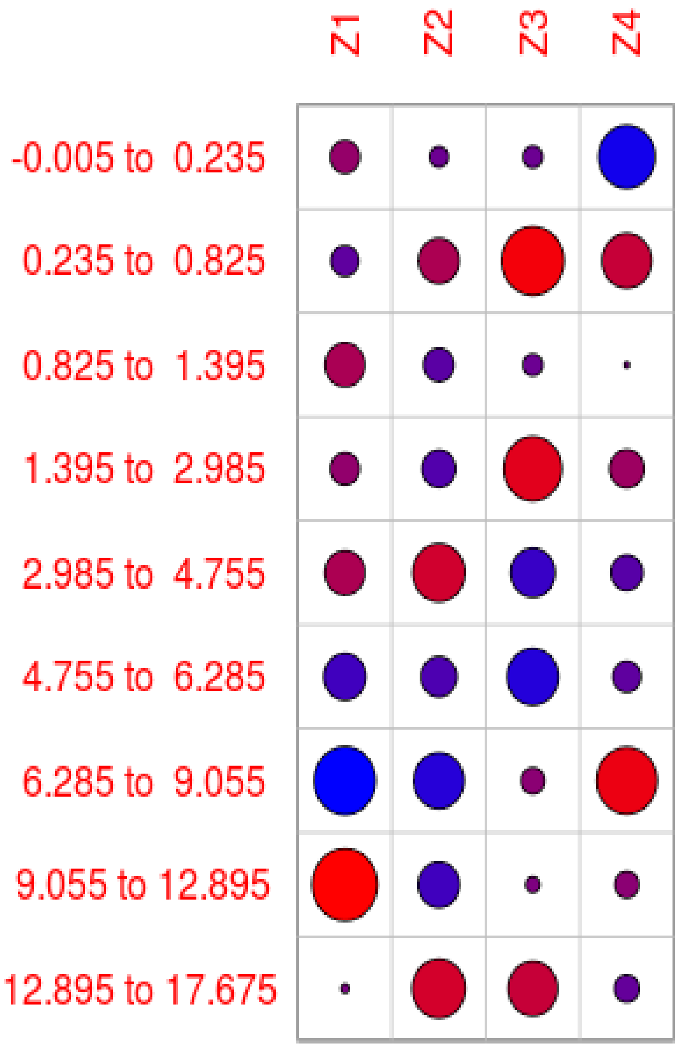


Figure 4.7 – Z statistics for the fitted BSPT model

As we can note from Figure 4.7 that the BSPT model does not present any misfits, since no squares can be seen in the figure, even in the 8^{th} age interval. Based on all residual diagnostics we can say that the BSPT model provides a reasonable fit to the data set in the study.

4.6 Concluding remarks

We presented a new very flexible extension of the BS distribution, that has most of the BS type distributions already available in the literature as special cases and described the method of maximum likelihood estimation for its parameters. In order to present its regression model we incorporated the distribution in the GAMLSS framework, and developed and presented a new generic package for fitting the model, called `gamlss.BSP`. A real data set, relating to the concentration of chemical GAG in urine of children to age, was used to illustrate the importance of the BSP regression model, showing that it produced better results than several other flexible distributions.

References

AITKIN, M., FRANCIS, B., HINDE, J.; DARNELL, R. **Statistical Modelling in R**. Oxford, UK: Oxford University Press, 2009. 588 p.

AKAIKE, H. A new look at the statistical model identification. **IEEE Transactions on Automatic Control**, New York, v. 19, p. 716–723, 1974.

BIRNBAUM, Z.W.; SAUNDERS, S.C. A new family of life distributions. **Journal of Applied Probability**, Sheffield, v. 6, p. 637–652, 1969.

CASTRO, M. de; CANCHO, V.G.; RODRIGUES, J.A. A hands-on approach for fitting long-term survival models under the GAMLSS framework. **Computer Methods and Programs in Biomedicine**, Amsterdam, v. 97, 168–177, 2010.

COLE, T. J.; GREEN, P. J. Smoothing reference centile curves: the LMS method and penalized likelihood. **Statistics in Medicine**, Chichester, v. 11, p. 1305–1319, 1992.

D'AGOSTINO, R.B.; BELANGER, A.; D'AGOSTINO JR., R.B. A suggestion for using powerful and informative tests of normality. **The American Statistician**, Abingdon, v. 44, p. 316–321, 1990.

DÍAZ-GARCÍA, J.A.; DOMÍNGUEZ-MOLINA, J.R. Some generalisations of Birnbaum-Saunders and sinh-normal distributions. **International Mathematical Forum**, Ruse, v. 1, p. 1709–1727, 2006.

DÍAZ-GARCÍA, J.A.; LEIVA, V. A new family of life distributions based on the elliptically contoured distributions. **Journal of Statistical Planning and Inference**, Amsterdam, v. 128, p. 445–457, 2005.

DUNN, P.K.; SMYTH, G.K. Randomized quantile residuals. **Journal of Computational and Graphical Statistics**, Alexandria, VA, v. 5, p. 236–245, 1996.

EILERS, P.H.C.; MARX, B.D. Flexible smoothing with B-splines and penalties (with comments and rejoinder). **Statistical Science**, Hayward, v. 11, p. 89–121, 1996.

GENÇ, A.I. The generalized T Birnbaum-Saunders family. **Statistics: A Journal of Theoretical and Applied Statistics**, Hampshire, v. 47, p. 613–625, 2013.

GUIRAUD, P.; LEIVA, V.; FIERRO, R. A non-central version of the Birnbaum-Saunders distribution for reliability analysis. **IEEE Transactions on Reliability**, New York, v. 58, p. 152–160, 2009.

HASTIE, T.J.; TIBSHIRANI, R.J. **Generalized Additive Models**. London: Chapman & Hall, 1990. 352 p.

HOSSAIN, A. RIGBY, R.; STASINOPOULOS, M.; ENEA, M. Centile estimation for a proportion response variable. **Statistics in Medicine**, Chichester, v. 35, p. 895–904, 2016.

LEIVA, V.; PONCE, M.G.; MARCHANT, C.; BUSTOS, O. Fatigue statistical distributions useful for modeling diameter and mortality of trees. **Revista Colombiana de Estadística**, Bogotá, v. 35, p. 349–370, 2012.

MATSUMOTO, K.; VOUDOURIS, V.; STASINOPOULOS, D.M.; RIGBY, R.A.; MAIO, C. Exploring crude oil production and export capacity of the OPEC Middle East countries. **Energy Policy**, Guildford, v. 48, p. 820–828, 2012.

MCDONALD, J.B.; NEWEY, W.K.. Partially adaptive estimation of regression models via the generalized t distribution. **Econometric Theory**, New York, v. 4, p. 428–457, 1988.

R CORE TEAM. **R: A language and environment for statistical computing**. Vienna, 2013. Available at: <http://cran.r-project.org/>. Access: 01 jun. 2016.

RIGBY, R.A.; STASINOPOULOS, D.M. Smooth centile curves for skew and kurtotic data modelled using the Box-Cox power exponential distribution. **Statistics in Medicine**, Chichester, v. 23, p. 3053–3076, 2004.

RIGBY, R.A.; STASINOPOULOS, D.M. Generalized additive models for location, scale and shape, (with discussion). **Journal of the Royal Statistical Society – Series C (Applied Statistics)**, Oxford, UK, v. 54, p. 507–554, 2005.

RIGBY, R.A.; STASINOPOULOS, D.M. Using the Box-Cox t distribution in GAMLSS to model skewness and kurtosis. **Statistical Modelling**, London, v. 6, p. 209–229, 2006.

ROYSTON, P.; WRIGHT, E.M. Goodness-of-fit statistics for age-specific reference intervals. **Statistics in Medicine**, Chichester, v. 19, p. 2943–2962, 2000.

SANHUEZA, A.; LEIVA, V.; BALAKRISHNAN, N. The generalized Birnbaum-Saunders distribution and its theory, methodology, and application. **Communications in Statistics: Theory and Methods**, New York, v. 37, p. 645–670, 2008.

STASINOPOULOS, D.M.; RIGBY, R.A. Generalized additive models for location scale and shape (GAMLSS) in R. **Journal of Statistical Software**, Los Angeles, v. 23, 1–10, 2007.

VAN BUUREN, S.; FREDRIKS, M. Worm plot: a simple diagnostic device for modelling growth reference curves. **Statistics in Medicine**, Chichester, v. 20, p. 1259–1277, 2001.

VENABLES, W.N.; RIPLEY, B.D. **Modern Applied Statistics with S**. Boston: Springer, 2002. 498 p.

ZHANG, D.D.; YAN, D.H.; WANG, Y.C.; LU, F.; LIU, S.H. GAMLSS-based nonstationary modeling of extreme precipitation in Beijing-Tianjin-Hebei region of China. **Natural Hazards**, Dordrecht, v. 77, p. 1037–1053, 2015.

5 THE BIRNBAUM-SAUNDERS POWER FAMILY OF DISTRIBUTIONS: BIMODALITY APPLICATIONS[1]

Abstract

The huge advance in the computational field, which enables the study and estimation of more complex data sets, stimulates the creation of a wide range of new distributions in order to model several distinct problems. The Birnbaum-Saunders power (BSP) family of distributions is a new very flexible model which adds two new parameters (location and scale) to any distribution on the real line. This family includes some important special and/or limiting cases in the literature, including the generalized Birnbaum-Saunders (GBS) family of distributions. In this paper we introduce the BSP family of distributions as an alternative family to model data sets with bimodal response variables using the generalized additive models for location, scale and shape (GAMLSS). A simulation study with covariates and two applications in real data sets are conducted in order to illustrate the great flexibility of this family.

Keywords: Bimodal data; GAMLSS; Generalized additive models; Penalized splines; R software; Regression

5.1 Introduction

The development of generalized families of distributions is fundamental to all topics related to statistics and it is a powerful tool for theoretical and applied statisticians. One of the main purposes to study new families of distributions is to increase the flexibility to model various types of data sets in which there is a clear need for extended forms of these distributions such as agronomic and environmental sciences, engineering, biological and medical studies, lifetime and reliability analysis, economics, finance and insurance. Consequently, significant progress has been achieved over the past decades in the generalization of some well-known distributions and at the same time they have provided great flexibility and applicability in modelling data in practice.

In this sense, recent developments that focus on new techniques for building new families (or generators) of distributions have been proposed in the statistical literature. Some well-known generators are the Azzalini's skewed family by Azzalini (1985), the exponentiated-G (EG) family by Mudholkar et al. (1995), the Marshall-Olkin generated family (MO-G) by Marshall and Olkin (1997), the beta-G by Eugene et al. (2002), the Kumaraswamy-G (Kw-G for short) by Cordeiro and de Castro (2011), the McDonald-G (Mc-G) by Alexander et al. (2012), the Kummer beta-G (KB-G) by Pescim et al. (2012) and more recently the transformer (T-X) by Alzaatreh et al. (2013).

Those families of distributions have received a great deal of attention in recent years because they allow more flexible densities and introduce skewness and vary tail weight. However, they

[1]Luiz R. Nakamura, Rodrigo R. Pescim (UEL), Robert A. Rigby (LMU), Dimitrios M. Stasinopoulos (LMU), Roseli A. Leandro (ESALQ/USP) and Cristian Villegas (ESALQ/USP). To be submitted.

cannot model a bimodal response variable distribution which frequently occurs in real data analysis. According to Famoye et al. (2004), a bimodal response variable distribution can occur in many areas of science. Withington et al. (2000) showed that plasma vecuronium and vecuronium clearance requirements have bimodal behavior in the study of cardiopulmonary bypass in infants. Freeland et al. (2000), Wolf and Sumner (2001), Zangvil et al. (2001) and Isaacson (2000) reported that bimodal distributions frequently occur in the study of genetic diversity, in agricultural farm size distribution, in atmospheric pressure and in the study of anabolic steroids on animals, respectively.

In Chapter 4, we proposed the Birnbaum-Saunders power (BSP) family of distributions which provides different degrees of skewness and kurtosis and includes most Birnbaum-Saunders type models available in the literature. The BSP family of distributions is defined by the transformation

$$Y = \psi \left[\frac{Z}{2} + \sqrt{\left(\frac{Z}{2}\right)^2 + 1} \right]^{\xi}, \tag{5.1}$$

where $Y > 0$, $\psi > 0$ is a scale parameter and $\xi > 0$ is a skewness parameter and the random variable Z follows any arbitrary baseline distribution on the real line, denoted by $Z \sim D(\boldsymbol{\theta})$, with parameter vector $\boldsymbol{\theta}$. Here, we assume that the distribution of Z has up to four parameters, i.e. $\boldsymbol{\theta} = (\mu, \sigma, \nu, \tau)^\top$, where $-\infty < \mu < \infty$ is the location parameter, $\sigma > 0$ is the scale parameter and ν and τ are parameters related to tail weight. Hence, the BSP family of distributions is very flexible and it can be used in many practical situations. In fact, it can be asymmetric (highly positively skewed and highly negatively skewed) or close to symmetric and can also exhibit bimodality.

The study of the BSP distributions is important since it extends some models previously considered in the literature. If $\xi = 2$ and $Z \sim N(0, \sigma^2)$, it yields the standard Birnbaum-Saunders distribution (BIRNBAUM; SAUNDERS, 1969). If $\xi = 2$ and Z follows any symmetric distribution with non-centrality parameter $\mu = 0$, it reduces to the generalized Birnbaum-Saunders (GBS) distribution (DÍAZ-GARCÍA; LEIVA, 2005).

The probability density function (pdf) corresponding to (5.1) can be expressed as

$$f_Y(y|\psi, \xi, \boldsymbol{\theta}) = f_Z(z|\boldsymbol{\theta}) \left| \frac{dz}{dy} \right|, \; y > 0, \tag{5.2}$$

where

$$\frac{dz}{dy} = \frac{1}{y\,\xi} \left[\left(\frac{y}{\psi}\right)^{\frac{1}{\xi}} + \left(\frac{y}{\psi}\right)^{-\frac{1}{\xi}} \right],$$

and z is related to y by (5.1).

Hereafter, we denote Y the random variable with pdf given by (5.2), by $Y \sim$ BSP$(\psi, \xi, \mu, \sigma, \nu, \tau)$. This pdf has up to six parameters depending on the baseline distribution and also it allows for a high degree of flexibility. In Section 4.2, we defined and studied the Birnbaum-Saunders power normal (BSPNO), the Birnbaum-Saunders power t (BSPT) and the Birnbaum-Saunders power generalized t (BSPGT) distributions by taking $f_Z(z|\boldsymbol{\theta})$ to be the pdf of the normal, scaled and shifted t and generalized t distributions, respectively. In this chapter, we use the bimodality properties of the BSPNO and BSPT distributions to model some bimodal data.

The BSP family of distributions defined by the pdf (5.2) is also an alternative family of models to fit a bimodal response variable which cannot be properly fitted by existing families of distributions. We note that while the parameters μ, ψ, ν and τ control the location, scale and weight of tails and adds flexibility, the parameters ξ and σ are responsible for yielding bimodal behavior in the BSP pdfs, i.e., for $\xi \to 0$ and σ large (usually > 1) BSP distributions present bimodality. Figure 5.1 displays some possible shapes of the BSPNO and BSPT pdfs. All the plots ploted have $\psi = 20$ and $\xi = 0.1$. These plots show the great flexibility achieved with the new bimodal distributions.

(a) (b)

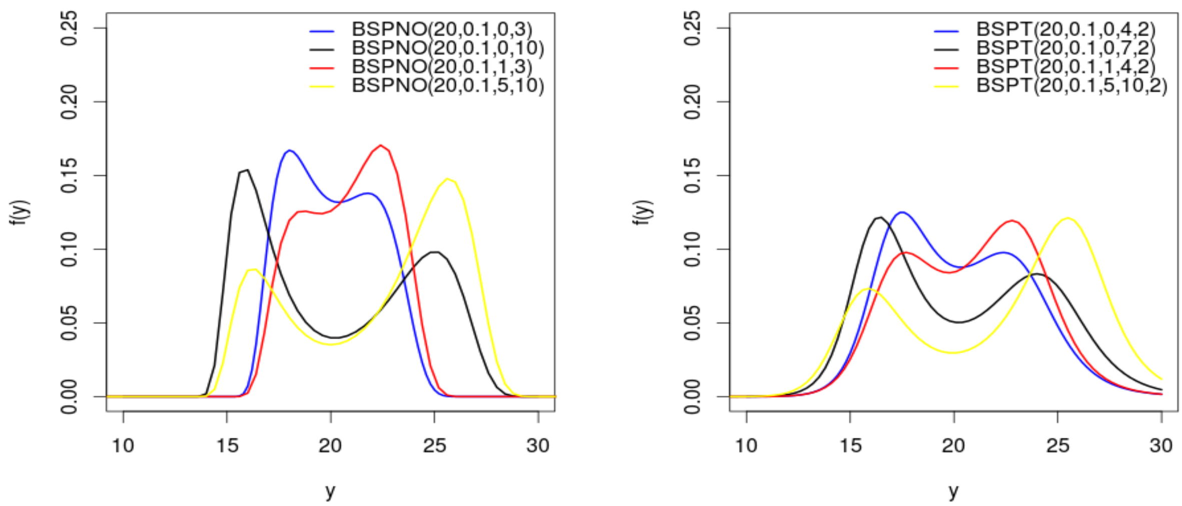


Figure 5.1 – Plots of the (a) BSPNO and (b) BSPT density functions

In this chapter, we introduce the BSP family of distributions as an alternative to fit bimodal response variables with the hope that it will attract wider applications in several areas of research. The rest of paper is outlined as follows. Section 5.2 presents a brief review about the GAMLSS framework. In Section 5.3, we run a simulation study modelling all parameters of the BSP family of distributions using covariates. Two real data set applications are presented in Section 5.4, one related to geysers and other regarding prawns. Section 5.5 ends the paper with some concluding remarks.

5.2 GAMLSS framework

In Section 4.4, we proposed a regression model for the BSP family of distributions based on the generalized additive models for location, scale and shape, GAMLSS, framework. GAMLSS are very flexible semi-parametric regression models that involve a distribution for the response variable, that does not necessarily belong to the exponential family, where all of its parameters can be modelled using parametric and/or non-parametric smooth functions, such as penalized splines (EILERS; MARX, 1996), of a set of explanatory variables, thus allow modelling of the location, scale and shape parameters.

The GAMLSS models are defined as follows. Let $Y \sim D(\boldsymbol{\theta})$, where D is the response variable distribution with parameter vector $\boldsymbol{\theta} = (\theta_1, \ldots, \theta_k)^\top$. For $k = 1, \ldots, p$, let $g_k(\cdot)$ be known monotonic link functions relating each of the parameters with their respectively predictors η_k. Then the GAMLSS models can be written as

$$g_k(\theta_k) = \eta_k = \boldsymbol{X}_k\boldsymbol{\beta}_k + \sum_{j=1}^{J_k} h_{jk}(\boldsymbol{x}_{jk}),$$

where $\boldsymbol{X}_k$ is a known design matrix, $\boldsymbol{\beta}_k^\top = \left(\beta_{1k}, \ldots, \beta_{J'_k k}\right)$ is a parameter vector of length J'_k and each h_{jk} function is a smooth non-parametric function of an explanatory variable $\boldsymbol{x}_{jk}$, for $j = 1, \ldots, J_k$ and $k = 1, \ldots, p$.

We created a new generic package (as mentioned in Section 4.4) in `R` (R CORE TEAM, 2013), that will be available soon in the Comprehensive R Archive Network (CRAN), named `gamlss.BSP`, that allows the BSP family of distributions for Y to be fitted for any corresponding baseline distribution for $-\infty < Z < \infty$ currently available in the `gamlss.family` package. For further details about the GAMLSS estimation, see Appendices A, B and C and Rigby and Stasinopoulos (2005).

5.3 Simulation study

In order to simulate values from the BSP distribution with pdf (5.2), when $F_Z^{-1}(u|\boldsymbol{\theta})$ exists, we can use the quantile function of Y, i.e. the inverse cumulative distribution function, $Q_Y(u) = F_Y^{-1}(u|\psi, \xi, \boldsymbol{\theta})$, given by

$$Q_Y(u) = \psi\left[\frac{F_Z^{-1}(u|\boldsymbol{\theta})}{2} + \sqrt{\left(\frac{F_Z^{-1}(u|\boldsymbol{\theta})}{2}\right)^2 + 1}\,\right]^{\xi}, \tag{5.3}$$

where u is an uniform random variable on the interval from zero to one.

We conduct three Monte Carlo simulation studies to assess the finite sample behavior of the maximum likelihood estimators (MLEs) of the parameters of the BSP distribution for two different sample sizes ($n = 50$ and $n = 100$) and two baseline distributions for Z: i) normal and ii) scaled and shifted student's t distributions. For all scenarios we obtained 1,000 Monte

Carlo replications and, for each replication, we calculate the MLEs of the parameters. After all replications we determined the average estimates (AEs), biases and mean squared errors (MSEs). The simulations are carried out using the package `gamlss.BSP` in `R`.

5.3.1 BSPNO simulation

In order to generate values from the BSPNO distribution, we use the function `rBSPNO` available in the `gamlss.BSP` package. The true parameter values used in the data-generating process are $\psi = 20$, $\xi = 0.1$, $\mu = \beta_{01} + \beta_{11}x_1 = 2 + 3x_1$ and $\sigma = \exp\{\beta_{02} + \beta_{12}x_1\} = \exp\{\log(5) + \log(2)x_1\}$, where x_1 was generated from a binomial$(n, 0.5)$, $n = 50, 100$, distribution. The simulation results are reported in Table 5.1.

Table 5.1 – The average estimates (AE), biases and mean squared errors (MSE) based on 1,000 simulations for the BSPNO model

$n = 50$				$n = 100$			
Parameter	AE	Bias	MSE	Parameter	AE	Bias	MSE
β_{01}	2.292	0.292	3.926	β_{01}	2.120	0.120	1.358
β_{11}	3.909	0.909	22.050	β_{11}	3.331	0.331	7.965
β_{02}	1.743	0.134	0.143	β_{02}	1.656	0.046	0.067
β_{12}	0.741	0.047	0.086	β_{12}	0.713	0.020	0.037
ψ	20.082	0.082	0.233	ψ	20.039	0.039	0.106
ξ	0.093	0.007	0.000	ξ	0.098	0.002	0.000

5.3.2 BSPT simulation

Now, we generate observations from the BSPT distribution using the `rBSPTF` function available in the `gamlss.BSP` package. The true parameter values for this simulation study are $\psi = 20$, $\xi = 0.2$, $\nu = 2$, $\mu = \beta_{01} + \beta_{11}x_1 = 2 + 3x_1$ and $\sigma = \exp\{\beta_{01} + \beta_{11}x_1\} = \exp\{\log(5) + \log(2)x_1\}$, where the explanatory variable x_1 was generated from a binomial$(n, 0.5)$, $n = 50, 100$, distribution as in the previous case. Table 5.2 displays the simulation results for the BSPT distribution.

Table 5.2 – The average estimates (AE), biases and mean squared errors (MSE) based on 1,000 simulations for the BSPT model

$n = 50$				$n = 100$			
Parameter	AE	Bias	MSE	Parameter	AE	Bias	MSE
β_{01}	1.858	0.142	2.444	β_{01}	1.941	0.059	1.102
β_{11}	3.056	0.056	15.516	β_{11}	2.841	0.159	4.810
β_{02}	1.519	0.091	0.145	β_{02}	1.549	0.061	0.058
β_{12}	0.677	0.017	0.141	β_{12}	0.686	0.007	0.058
ν	2.194	0.194	1.094	ν	2.364	0.364	0.914
ψ	20.387	0.387	1.446	ψ	20.116	0.116	0.547
ξ	0.200	0.000	0.001	ξ	0.200	0.000	0.001

We can note the apparent consistency of the estimates of the proposed models in agreement with theoretical properties for MLEs. The results from Tables 5.1 and 5.2 indicate that the MSEs of the parameter estimates decay toward zero as the sample size increases, as expected under standard regularity conditions. Further, as the sample size n increases, the AE tend to be closer to the true parameter value. Figures 5.2 (a) and (b) display the true density and the density of the average values of the parameters, for $n = 100$, from the BSPNO and BSPT simulation studies, respectively.

(a) (b)

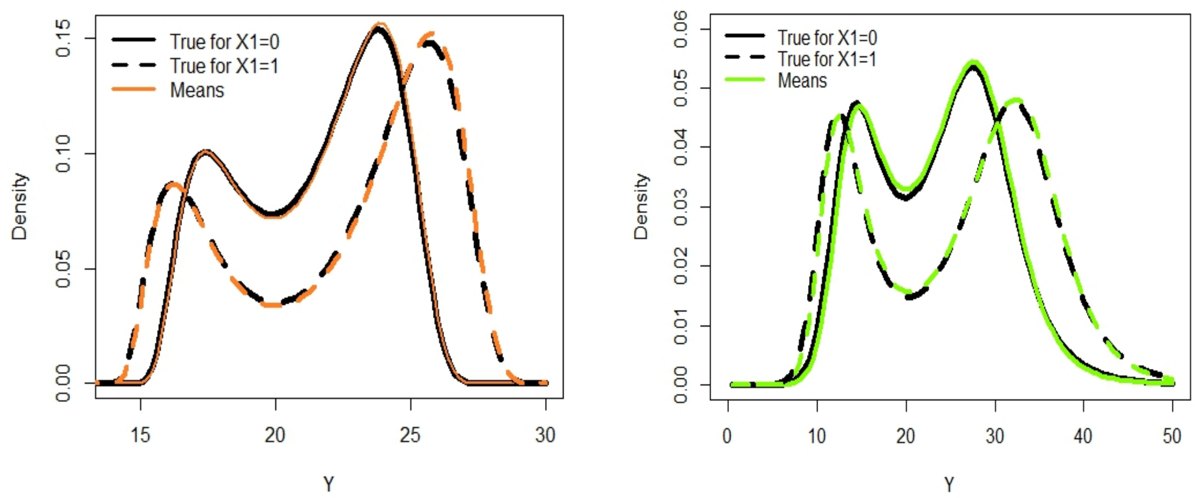


Figure 5.2 – Estimated densities at the true parameter values and at the AEs obtained in the simulation study, considering $n = 100$ of the (a) BSPNO and (b) BSPT distributions

It is noteworthy that despite of the number of parameters from the BSP distributions and their bimodal shapes of the response variable, we did not encounter any convergence problems using the `gamlss.BSP` package in `R`.

5.4 Applications

In this section we provide two applications to real bimodal data to illustrate the great flexibility of the BSP regression model based on the GAMLSS framework. The computations are performed using the `gamlss.BSP` and `gamlss.mx` (STASINOPOULOS; RIGBY, 2007) packages. In both applications we compare the BSPNO and BSPT models with a mixture-normal model.

5.4.1 Eruption data

The first data set modelled in this paper refers to two variables measured on the Old Faithful geyser in Yellowstone National Park, Wyoming, USA, available on `R` software under the data set name `faithful`: waiting time between eruptions (Y) and duration of the eruption (X). The response variable Y, displayed in Figure 5.3, ranges from 43.0 to 96.0 and X ranges from 1.6 to 5.1. There was a total of 272 observations in this data set.

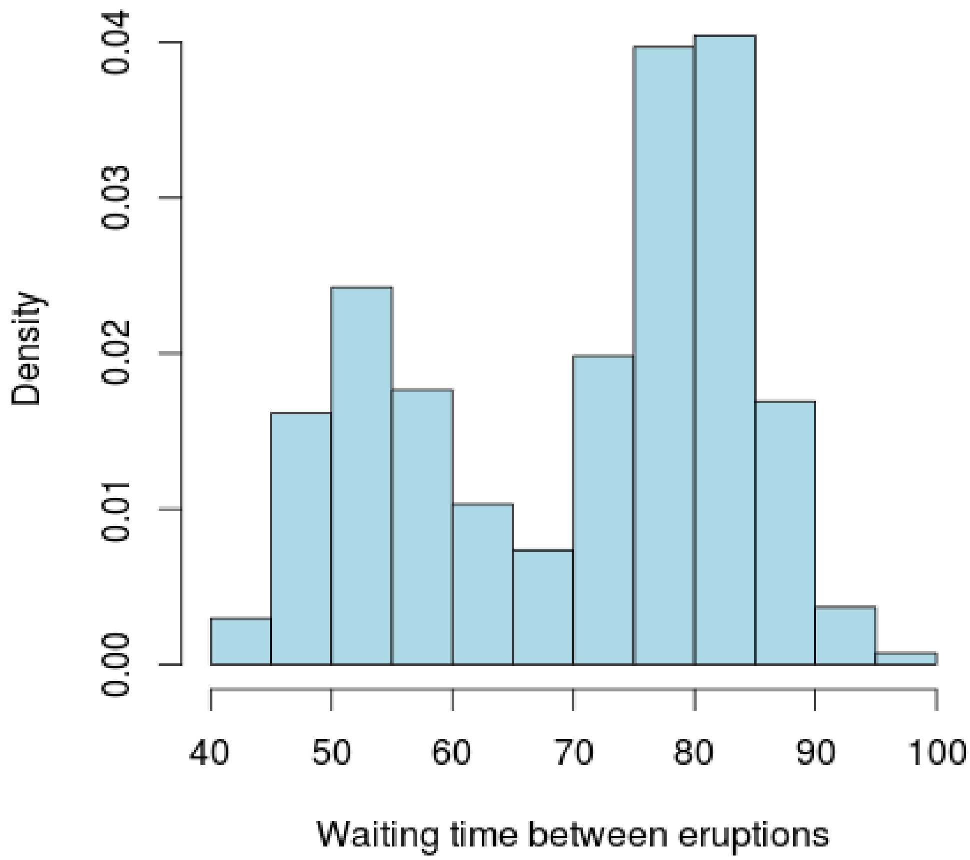


Figure 5.3 – Histogram of the waiting time between eruption on the Old Faithful geyser

Since the variable waiting time between eruptions is bimodal, we used two special models from the BSP family of distributions: BSPNO and BSPT, and also we compare them with a mixture-normal model. Table 5.3 displays the number of distribution parameters, the total (effective) degrees of freedom (df) used in the respective model, the values of global deviance (GD), Akaike information criterion (AIC) (AKAIKE, 1974) and Schwarz Bayesian criterion (SBC) (SCHWARZ, 1978). We can see that the BSPNO model is chosen as the best fitted model to the eruption data, since it returned the smallest GD, AIC and SBC values (1,688.62, 1,714.32 and 1,760.66, respectively).

Table 5.3 – Statistics from the fitted models for the eruption data

Model	Parameters	df	GD	AIC	SBC
BSPNO	4	12.85	1,688.62	1,714.32	1,760.66
Mixture-normal	4	16.72	1,688.86	1,722.30	1,782.57
BSPT	5	20.02	1,688.82	1,728.85	1,801.03

The fitted model from the BSPNO distribution under the GAMLSS framework is given by $Y \sim BSPNO(\psi, \xi, \mu, \sigma)$ where

$$\log(\psi) = 3.677 + h_{11}(X)$$

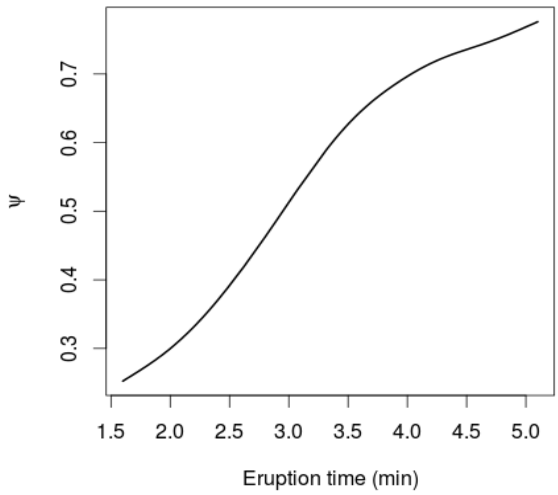
$$\log(\xi) = -1.564 - 0.136X$$

$$\mu = 0.525 - 0.195X$$

$$\log(\sigma) = 0.508 + h_{14}(X), \tag{5.4}$$

where the parameters ψ and σ are explained by non-parametric (penalized splines) functions h_{11} and h_{14}, respectively. Both functions are displayed in Figure 5.4. Panel (a) presents the relationship between the location parameter ψ and the covariate X. Although it has an almost linear relationship, we can see that when the eruption time is less than three minutes, the function grows fast and after this point it starts to grow slower. The fitted model for the shape parameter σ indicates that the value of this parameter increases in eruption times up to about 2.75 minutes, decreases until eruption times about 4.6 minutes, then increases. We can also see from model (5.4) that the longer is the eruption time the lower are the values of parameters ξ and μ. Note that parameter ξ will always be close to zero (one of the assumptions in order to obtain a bimodal shape for a BSP model).

(a) (b)



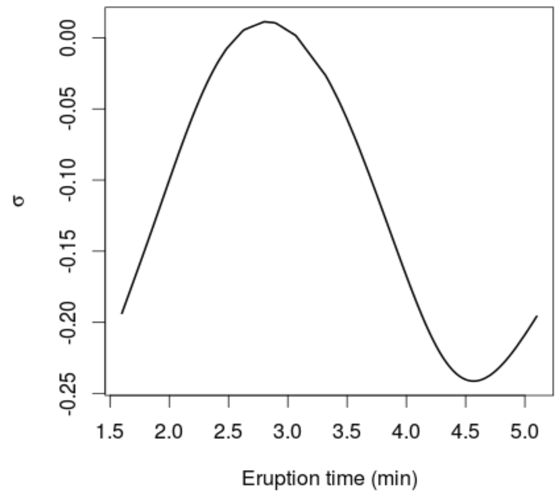
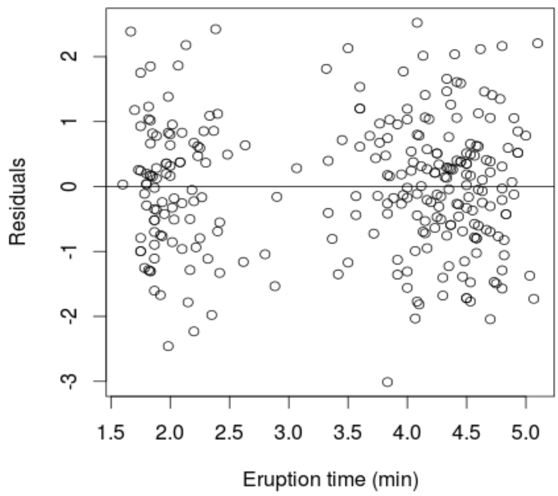


Figure 5.4 – The fitted parameters from the BSPNO model against eruption time: (a) ψ and (b) σ

Figure 5.5 displays the normalized quantile residuals (DUNN; SMYTH, 1996) from the BSPNO fitted model (5.4). The true residuals follow a standard normal distribution if the model is correct. Panels (a) and (b) give a plot of the residuals against the eruption time and a simulated envelope, respectively, and they indicate that a normal distribution for the residuals appears reasonable and that the residuals appear to be random.

(a) (b)



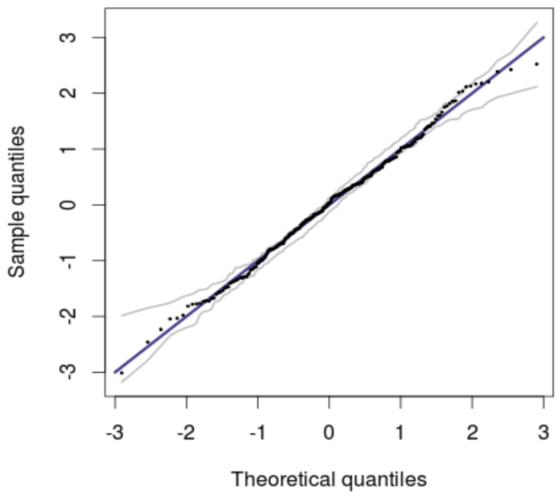


Figure 5.5 – The residuals from the BSPNO model: (a) against the eruption time and (b) simulated envelope

Figure 5.6 presents the worm plot (VAN BUUREN; FREDRIKS, 2001). If a vertical shape, a slope, a quadratic or a cubic shape is observed, it may indicate misfits in the location, scale, skewness and excess kurtosis of the residuals, respectively, and hence, there is a problem with the fitted model. Moreover, 95% of the residual points should lie between the elliptical 95% pointwise interval band curves. The plots are read in rows from the bottom left plot to the top right plot and correspond, in this case, to the nine intervals in the eruption time variable given above the worm plot, from the shortest to the longest eruption time. We can verify from the worm plots that the model is adequate throughout all the eruption time intervals.

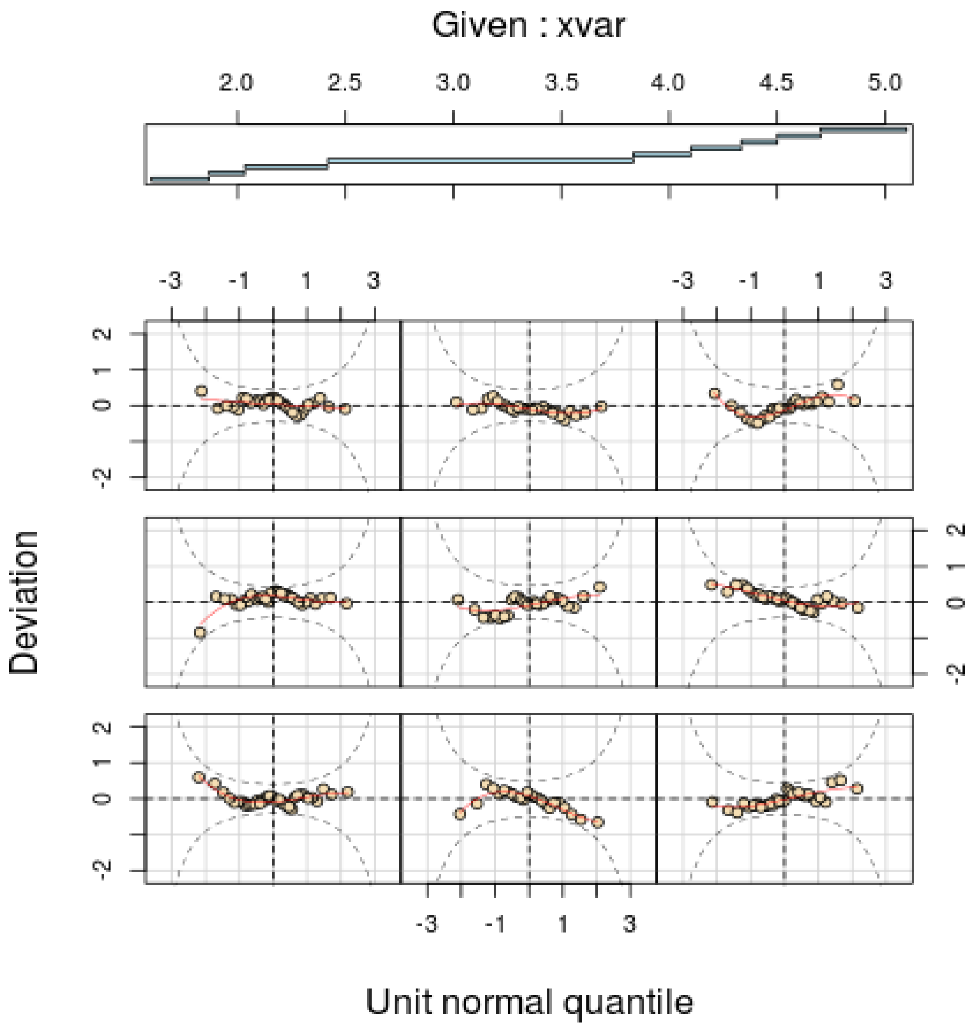


Figure 5.6 – Worm plot for the fitted BSPNO model

5.4.2 Prawn data

The second data set analyzed in this chapter regards prawn (*farfantepenaeus brasiliensis*) biometric measurements, collected in three different regions of Rio Grande do Norte State coast, Brazil. This data set, comprehending 120 observations, was collected by Pinheiro (2008). Here we have the response variable prawn weight (Figure 5.7), which ranges from 1.98 and 35.78 g, and we try to explain it using three different covariates: X_1 is the region where prawns were collected and has three different levels (`1`: Baia Formosa, `2`: Diogo Lopes and `3`: Touros), X_2 is the prawn gender (`0`: female and `1`: male) and X_3 is the prawn length that ranges from 55.8 and 138.2 mm.

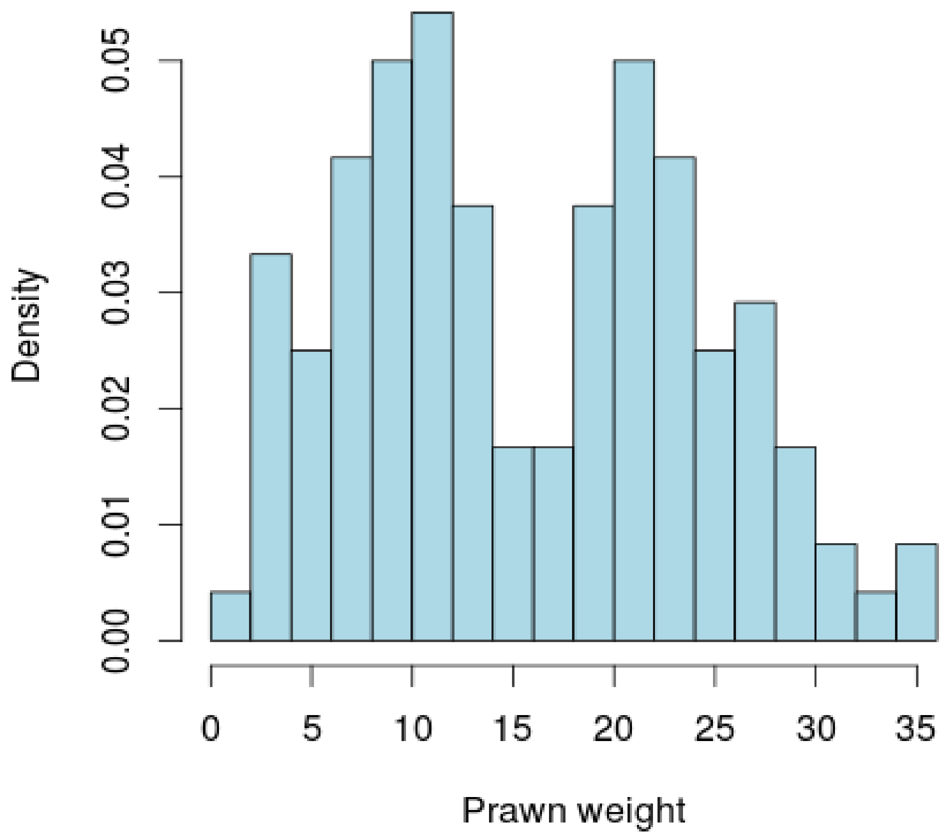


Figure 5.7 – Histogram of the prawn weight

Since the response variable prawn weight has a bimodal shape we have as possible suitable distributions the BSP family of distributions. In this data set, as in the previous one, we used the BSPNO, BSPT and mixture-normal models. Table 5.4 displays results of the best fitted models for each distribution, presenting the number of distribution parameters, the effective degrees of freedom (df) used in fitting the model, global deviance (GD), AIC and SBC.

Table 5.4 – Statistics from the fitted models for the prawn data

Model	Parameters	df	GD	AIC	SBC
BSPT (with $\mu = 0$)	4	15.63	290.52	321.78	365.33
BSPNO	4	11.93	309.52	333.38	366.63
Mixture-normal	4	13.00	448.29	474.29	510.52

We can see from Table 5.4 that BSPT (with $\mu = 0$) is the best fitted model according to the AIC and SBC measures (321.78 and 365.33, respectively). The BSPT model as the best model here is given by

$$\log(\psi) = -0.475 - 0.048(\text{if } X_2 = 1) + h_{11}(X_3)$$

$$\log(\xi) = -1.199 + 2.365(\text{if } X_1 = 2) + 2.133(\text{if } X_1 = 3) + h_{12}(X_3)$$

$$\mu = 0.000$$

$$\log(\sigma) = 1.346 - 3.146(\text{if } X_1 = 2) - 2.958(\text{if } X_1 = 3)$$

$$\log(\nu) = 1.285. \tag{5.5}$$

From (5.5), we can see that $\mu = 0$, which implies that the parameter ψ is exactly the median of the BSP distribution (see Section 4.1). Hence, from the model for the median ψ, we observe that male prawns tend to be smaller than female ones as expected (BAUER, 2004). Moreover ψ is also modelled by a non-parametric function, denoted by h_{11}, over the covariate length (Figure 5.8(a)). Although this function is roughly linear, the penalized spline was important to obtain reliable residuals, which are presented in Figure 5.9. As expected, the larger is the animal the heavier it will be. Prawns collected from locations number two (Diogo Lopes) and three (Touros) produce a greater value of the parameter ξ. Parameter ξ is also modelled by a non-parametric function (h_{12}). This relationship is shown in Figure 5.8(b) and indicates that the value of parameter ξ decreases until about 115 mm and then stabilizes as a constant up to 138.2 mm. Variable location was the only one selected to explain the parameter σ. As we can see, prawns collected from region number one (Baia Formosa) produces high values of this parameter. Finally, parameter ν is modelled as a constant (1.285). As we can see, the BSPT distribution was needed to model this data set since a small value of ν produces a heavy-tailed

(a) (b)

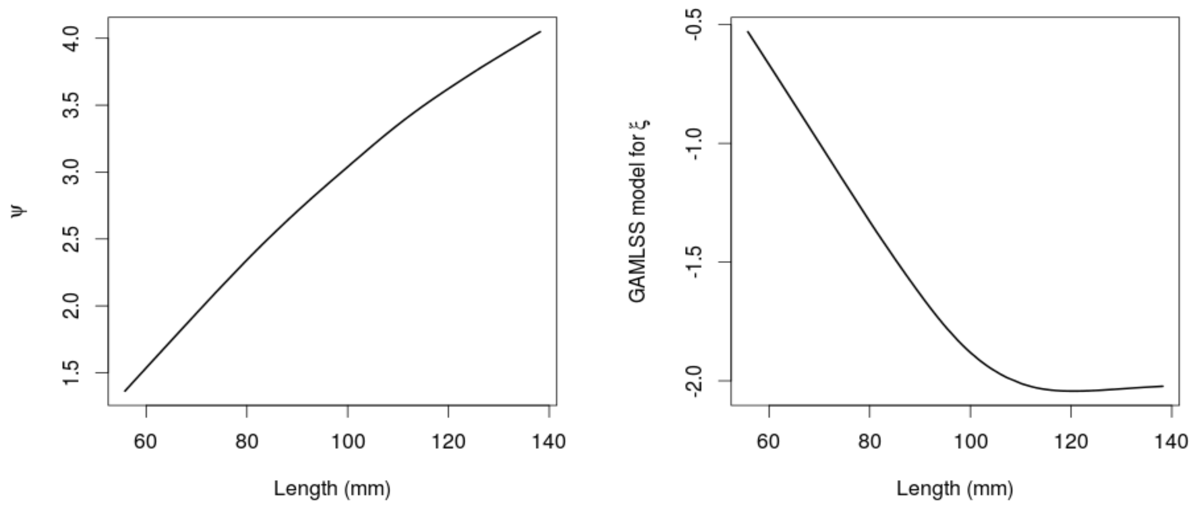


Figure 5.8 – Relationship between length and parameters ψ and ξ

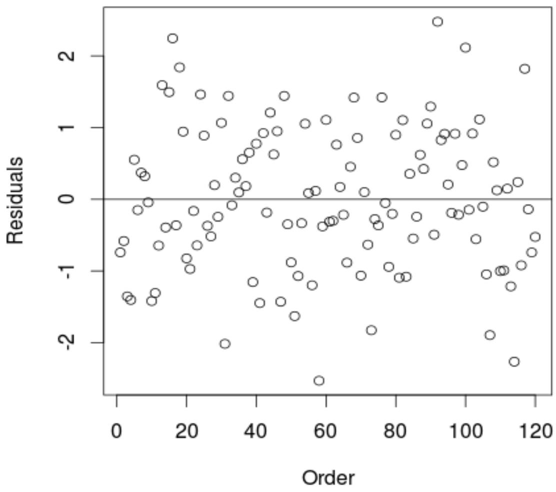
distribution (close to the Birnbaum-Saunders power Cauchy distribution, which is obtained when $\nu = 1$).

Figure 5.9 displays the normalized quantile residuals for the prawn data. Panel (a) shows us that the residuals adequately follow a normal distribution. This information is corroborated by the simulation envelope presented in Panel (b), since the majority of the points are within the simulated confidence bands. Panel (c) presents the worm plot for the BSPT fitted model and there are no evidences of inadequacies in it, since all the residuals fall in the "acceptance" region inside the two elliptic curves and no specific shape is detected in the points.

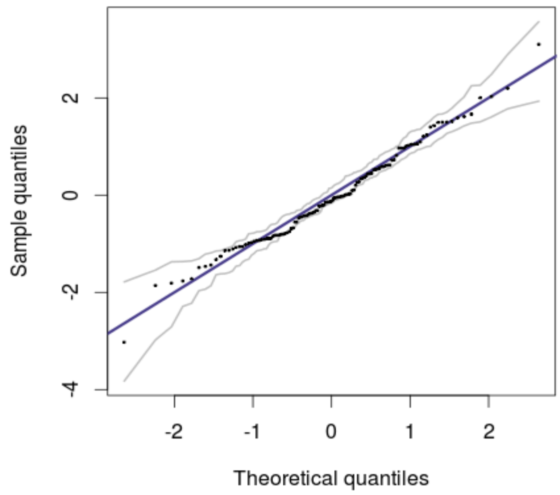
5.5 Concluding remarks

In this chapter, we used the new very flexible Birnbaum-Saunders power (BSP) family of distributions as a new alternative for fitting bimodal response variables. In order to fit a BSP regression model, we used the GAMLSS framework, which proved to be a very powerful methodology. A simulation study using the BSPNO and BSPT distributions (with a covariate) was performed, showing that the estimates of all parameters are satisfactory. Further, we did not face any convergence problems using the `gamlss.BSP` package in software `R`. Finally, two applications to real data sets show the great flexibility of the BSP family of distributions when applied to problems involving a bimodal response variable and we hope that it can attract wider applications in statistics.

(a)



(b)



(c)

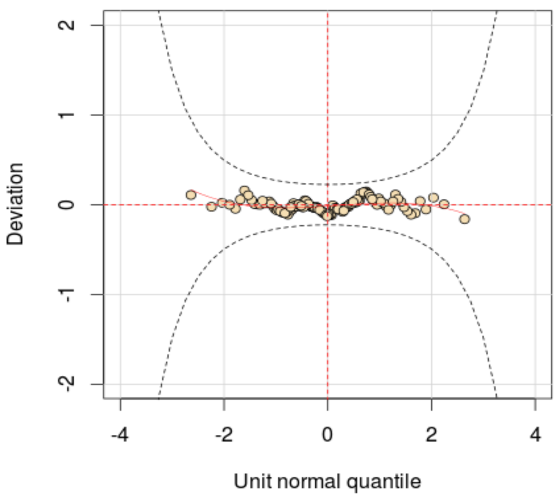


Figure 5.9 – The residuals from the BSPT model: (a) against the index; (b) simulated envelope; and (c) worm plot

References

AKAIKE, H. A new look at the statistical model identification. **IEEE Transactions on Automatic Control**, New York, v. 19, p. 716–723, 1974.

ALEXANDER, C.; CORDEIRO, G. M.; ORTEGA, E. M. M.; SARABIA, J. M. Generalized beta-generated distributions. **Computational Statistics & Data Analysis**, Amsterdam, v. 56, p. 1880–1897, 2012.

ALZAATREH, A.; LEE, C.; FAMOYE, F. A new method for generating families of continuous distributions. **Metron**, Padua, v. 71, n. 1, p. 63–79, 2013.

AZZALINI, A. A class of distributions which includes the normal ones. **Scandinavian Journal of Statistics**, Stockholm, v. 12, n. 2, p. 171–178, 1985.

BAUER, R.T. **Remarkable Shrimps: Adaptations and Natural History of the Carideans**. Oklahoma: University of Oklahoma Press, 2004. 316 p.

BIRNBAUM, Z. W.; SAUNDERS, S. C. A new family of life distribution. **Journal of Applied Probability**, Sheffield, v. 6, p. 319–327, 1969.

CORDEIRO, G. M.; CASTRO, M. de. A new family of generalized distributions. **Journal of Statistical Computation and Simulation**, Philadelphia, v. 81, n. 7, p. 883–898, 2011.

DUNN, P.K.; SMYTH, G.K. Randomized quantile residuals. **Journal of Computational and Graphical Statistics**, Alexandria, VA, v. 5, p. 236–245, 1996.

EILERS, P.H.C.; MARX, B.D. Flexible smoothing with B-splines and penalties (with comments and rejoinder). **Statistical Science**, Hayward, v. 11, p. 89–121, 1996.

EUGENE, N.; LEE, C.; FAMOYE, F. Beta-normal distribution and its applications. **Communications in Statistics: Theory and Methods**, New York, v. 31, n. 4, p. 497–512, 2002 .

FAMOYE, F.; LEE, C.; EUGENE, N. Beta-normal distribution: bimodality properties and application. **Journal of Modern Applied Statistical Methods**, Oak Park, MI, v. 3, n. 1, p. 85–103, 2004.

FREELAND, J. R.; ROMUALDI, C.; OKAMURA, B. Gene flow and genetic diversity; a comparison of freshwater bryozoan populations in Europe and North America. **Heredity**, London, v. 85, p. 498–508, 2000.

ISAACSON, R. L. Non linear effects in the retention of an avoidance task induced by anabolic steroids. **European Journal of Pharmacology**, Amsterdam, v. 405, p. 177–186, 2000.

MARSHALL, A. W.; OLKIN, I. A new method for adding a parameter to a family of distributions with application to the exponential and Weibull families. **Biometrika**, Oxford, UK, v. 84, n. 3, p. 641–652, 1997.

MUDHOLKAR, G. S.; SRIVASTAVA, D. K.; MARSHALL, F. The exponentiated Weibull family: a reanalysis of the bus-motor-failure data. **Technometrics**, Richmond, VA, v. 37, n. 4, p. 436–445, 1995.

PESCIM, R. R.; CORDEIRO, G .M.; DEMÉTRIO, C. G. B.; ORTEGA, E. M. M.; NADARAJAH, S. The new class of Kummer beta generalized distributions. **Statistics and Operations Research Transactions**, Barcelona, v. 36, p. 153–180, 2012.

PINHEIRO, A.P. **Caracterização genética e biométrica das populações de camarão rosa *Farfantepenaeus brasiliensis* de três localidades da costa do Rio Grande do Norte**. 2008. 82 f. Ph.D. Thesis (Ecology and natural resources) – Universidade Federal de São Carlos, São Carlos, 2008.

R CORE TEAM. **R: A language and environment for statistical computing**. Vienna, 2013. Available at: <http://cran.r-project.org/>. Access: 01 jun. 2016.

RIGBY, R.A.; STASINOPOULOS, D.M. Generalized additive models for location, scale and shape, (with discussion). **Journal of the Royal Statistical Society – Series C (Applied Statistics)**, Oxford, UK, v. 54, p. 507–554, 2005.

SCHWARZ, G. Estimating the dimension of a model. **The Annals of Statistics**, Hayward, v. 6, p. 461–464, 1978.

STASINOPOULOS, D.M.; RIGBY, R.A. Generalized additive models for location scale and shape (GAMLSS) in R. **Journal of Statistical Software**, Los Angeles, v. 23, n. 7, p. 1–10, 2007.

VAN BUUREN, S.; FREDRIKS, M. Worm plot: a simple diagnostic device for modelling growth reference curves. **Statistics in Medicine**, Chichester, v. 20, p. 1259–1277, 2001.

WITHINGTON, D.; MENARD, G.; HARRIS, J. KULKARNI, P.; DONATI, F.; VARIN, F. Vecuronium pharmacokinetics and pharmacodynamics during hypothermic cardiopulmonary bypass in infants and children. **Canadian Journal of Anesthesia**, Toronto, v. 47, n. 12, p. 1188–1195, 2000.

WOLF, C. A.; SUMNER, D. A. Are farm size distributions bimodal? Evidence from kernel density estimates of dairy farm size distributions. **American Journal of Agricultural Economics**, St. Paul, MN, v. 83, p. 77–88, 2001.

ZANGVIL, A.; PORTIS, D. H.; LAMB, P. J. Investigation of the large-scale atmospheric moisture field over the Midwestern United States in relation to summer precipitation. Part I: Relationships between moisture budget components on different timescales. **Journal of Climate**, Boston, v. 14, p. 582–597, 2001.

6 CONCLUSION

In this thesis, we provided some advances on the Birnbaum-Saunders (BS) distribution. Firstly, we adopted a Bayesian approach in order to estimate the two parameters of the Birnbaum-Saunders special-case (BS-SC) distribution. Although this distribution has been previously developed in the literature, its properties were never studied. We showed that the BS-SC distribution has heavier tails and allocates more information around the mean than the BS distribution and hence, it could be an interesting competitive model to the BS and the BS-t distributions when there are extreme values in tails of the distribution. In the Bayesian approach the BS-SC model is simpler considering that the BS-t distribution has a degree of freedom parameter (ν) which is somewhat not very easy to estimate. The Bayesian approach was necessary since only the first moment of the BS-SC distribution can be obtained. We showed that we can generate data from the BS-SC distribution using a Metropolis-Hastings algorithm. An application was conducted to demonstrate that this distribution can produce a better fit than the BS and BS-t distributions according to the deviance information criterion (DIC).

Another interesting and more flexible extension of the BS distribution studied in this thesis was the Birnbaum-Saunders generalized t (BSGT) distribution, which admits different degrees of skewness and kurtosis and includes some important special or limiting cases available in the literature, such as the BS and BS-t distributions. We developed a regression model for this distribution based on the generalized additive models for location, scale and shape (GAMLSS) framework, allowing any parameter of the BSGT distribution to be modelled as parametric and/or nonparametric functions of explanatory variables. A simulation study was conducted to investigate the behavior of the estimators and this new regression model was applied to a real motor vehicle insurance data set. In the application, the BSGT regression model produced better results than some well-known models in the GAMLSS literature according to the Akaike information and Schwarz Bayesian criteria (AIC and SBC, respectively).

We developed a new extension of the BS distribution, called the Birnbaum-Saunders power (BSP) family of distributions which has several special cases, including the BSGT distribution. For any baseline distribution with support on the real line, we can add two extra parameters using simple formulae to create a BSP distribution. This family can produce unimodal or bimodal shapes, depending on the values of its parameters. Implementation of fitting a BSP regression model was achieved using a new generic package, that we called `gamlss.BSP` in software `R`, that allows the BSP distribution for Y to be fitted with parametric and/or nonparametric functions of explanatory variables, for any corresponding distribution for Z on the real line currently available in the `gamlss.family` package. A simulation study using two of its special cases – Birnbaum-Saunders power normal (BSPNO) and Birnbaum-Saunders power t (BSPT) distributions – with a response variable with bimodal shape and a covariate, performed by the new package, was presented and regardless this behavior and the number of parameters (four and five) we did not face any convergence problems. The potentiality of the BSP family of distributions was illustrated in three different applications to real data sets throughout this thesis, showing its great flexibility.

Finally, we hope this thesis encourages people to use different extensions of the BS distributions and also to use the GAMLSS framework to produce very flexible semi-parametric regression models to explain the behavior of very complex data sets.

APPENDIX

APPENDIX A - Estimation in GAMLSS

The generalized additive models for location, scale and shape (GAMLSS) defined in (3.3) and (4.5) can be written in the form $h(\boldsymbol{x}) = \boldsymbol{Z}\boldsymbol{\gamma}$, i.e.

$$g_k(\boldsymbol{\theta}_k) = \boldsymbol{X}_k\boldsymbol{\beta}_k + \sum_{j=1}^{J_k} \boldsymbol{Z}_{kj}\boldsymbol{\gamma}_{kj}, \tag{1}$$

where $\boldsymbol{\theta}_k^\top = (\theta_{i1}, \ldots, \theta_{ip})$ is the vector of parameters of length p, $g_k(\cdot)$ denote known monotonic link functions, $\boldsymbol{\beta}_k^\top = (\beta_{1k}, \ldots, \beta_{J'k})$ is the parameter vector associated with the explanatory variables in design matrix $\boldsymbol{X}_k$, $\boldsymbol{Z}$ is the basis matrix which depends on the explanatory variable $\boldsymbol{x}$, $\boldsymbol{\gamma}$ is a parameter vector to be estimated, subject to a quadratic penalty of the form $\lambda\boldsymbol{\gamma}^\top\boldsymbol{G}\boldsymbol{\gamma}$, for a known matrix $\boldsymbol{G} = \boldsymbol{D}^\top\boldsymbol{D}$, λ is the parameter responsible by the smoothing needed for the fit and $\boldsymbol{D}$ is a difference matrix of order k.

The GAMLSS model can be fitted by the maximum penalized likelihood method with respect to $\boldsymbol{\beta} = \left(\boldsymbol{\beta}_1^\top, \ldots, \boldsymbol{\beta}_k^\top\right)^\top$ and $\boldsymbol{\gamma} = (\boldsymbol{\gamma}_{11}, \ldots, \boldsymbol{\gamma}_{1J_1}, \boldsymbol{\gamma}_{21}, \ldots, \boldsymbol{\gamma}_{kJ_k})^\top$, for a fixed $\boldsymbol{\lambda} = (\lambda_{11}, \ldots, \lambda_{1J_1}, \lambda_{21}, \ldots, \lambda_{kJ_k})^\top$. The penalized log-likelihood function for model (1) can be written as

$$l_p = l - \frac{1}{2}\sum_{k=1}^{p}\sum_{j=1}^{J_k} \lambda_{kj}\boldsymbol{\gamma}_{kj}^\top\boldsymbol{G}_{kj}\boldsymbol{\gamma}_{kj},$$

where $l = \sum_{i=1}^{n} \log f(y_i|\mu_i, \sigma_i, \nu_i, \tau_i)$. Note that if no smooth functions are adressed in the model, a simple maximum likelihood estimation should be performed.

In order to estimate $\boldsymbol{\beta}$ and $\boldsymbol{\gamma}$ for fixed $\boldsymbol{\lambda}$, Rigby and Stasinopoulos (2005) provide two different algorithms: CG (stands for Cole and Green) and RS (stands for Rigby and Stasinopoulos). The main difference between them is that RS algorithm maximizes the penalized likelihood over each of the parameters in turn and CG jointly updates all parameters. Both algorithms are described in Appendices B and C, respectively. There is also a third option, which is a combination of the two algorithms, where the RS is performed in early iterations of a given problem and later switches to the CG algorithm. Full details can be seen in Rigby and Stasinopoulos (2005).

APPENDIX B - The RS algorithm

The RS algorithm is divided in three different phases: the outer iteration, the inner iteration and the modified backfitting algorithm. For simplicity, we will assume that the distribution of the response variable we are working with has four parameters, hence given the initial values ($\boldsymbol{\mu}_0 = \boldsymbol{\theta}_1^0, \boldsymbol{\sigma}_0 = \boldsymbol{\theta}_2^0, \boldsymbol{\nu}_0 = \boldsymbol{\theta}_3^0,\ \ \boldsymbol{\tau}_0 = \boldsymbol{\theta}_4^0$) for ($\boldsymbol{\mu} = \boldsymbol{\theta}_1, \boldsymbol{\sigma} = \boldsymbol{\theta}_2, \boldsymbol{\nu} = \boldsymbol{\theta}_3, \boldsymbol{\tau} = \boldsymbol{\theta}_4$), the first step, outer iteration has the following steps:

1) Given the estimates $\hat{\boldsymbol{\sigma}}$, $\hat{\boldsymbol{\nu}}$ and $\hat{\boldsymbol{\tau}}$ from the last cycle, maximize the penalized log-likelihood over $\boldsymbol{\mu}$, i.e. fit a model for $\boldsymbol{\mu}$;

2) Given the estimates $\hat{\boldsymbol{\mu}}$, $\hat{\boldsymbol{\nu}}$ and $\hat{\boldsymbol{\tau}}$ from the last cycle, fit a model for $\boldsymbol{\sigma}$;

3) Given the estimates $\hat{\boldsymbol{\mu}}$, $\hat{\boldsymbol{\sigma}}$ and $\hat{\boldsymbol{\tau}}$ from the last cycle, fit a model for $\boldsymbol{\nu}$;

4) In the last step we fit a model for $\boldsymbol{\tau}$ given the estimates $\hat{\boldsymbol{\mu}}$, $\hat{\boldsymbol{\sigma}}$ and $\hat{\boldsymbol{\tau}}$ from the last cycle.

When all steps are completed, the global deviance (G$D = -2\hat{l}$, where $\hat{l}$ is the current log-likelihood), is calculated and a checked for convergence. If the convergence is achieved then the algorithm stops, otherwise all the process is repeated.

In each of the itens described before, i.e. for each fitting of a distribution parameter $\boldsymbol{\theta}_k$, $k = 1, \ldots, p$, the inner iteration is performed. Basically, a working variable is generated to fit each $\boldsymbol{\theta}_k$ as follows

$$\boldsymbol{z}_k = \boldsymbol{\eta}_k + \boldsymbol{w}_k^{-1} \circ \boldsymbol{u}_k, \qquad (2)$$

where $\boldsymbol{\eta}_k = g_k(\boldsymbol{\theta}_k)$ is the predictor vector of θ_k, $\boldsymbol{u}_k$ is the score function, i.e.

$$\boldsymbol{u}_k = \frac{\partial l}{\partial \boldsymbol{\eta}_k} = \left(\frac{\partial l}{\partial \boldsymbol{\theta}_k}\right) \circ \left(\frac{d\boldsymbol{\theta}_k}{d\boldsymbol{\eta}_k}\right),$$

$\boldsymbol{w}_k = (w_{k1}, \ldots, w_{kn})^\top$ is the vector of iterative weights, which, in our case, is given by

$$\boldsymbol{w}_k = \boldsymbol{u}_k \circ \boldsymbol{u}_k,$$

and $\boldsymbol{w}_k^{-1} \circ \boldsymbol{u}_k = \left(w_{k1}^{-1} u_{k1}, \ldots, w_{kn}^{-1} u_{kn}\right)^{-1}$ is the Hadamard product.

Given the current estimates for each parameter, $\boldsymbol{w}_k$ and $\boldsymbol{z}_k$ are recalculated and used in a weighted fit against all the explanatory variables considered for parameter k using modified backfitting process. This process is performed until there is no change in the GD.

For simplicity, let us consider only two smoothers, γ_{k1} and γ_{k2} with basis matrices $\boldsymbol{Z}_{k1}$ and $\boldsymbol{Z}_{k2}$, respectively. Given the current values for the working variable $\boldsymbol{z}_k$, the working weights $\boldsymbol{w}_k$, the smoothers for parameter and their respectively basis matrices, the backfitting process works as follows:

1) Calculate the partial residuals for $\boldsymbol{\beta}_k$, i.e. $\boldsymbol{\varepsilon} = \boldsymbol{z}_k - \boldsymbol{Z}_{k1}\hat{\gamma}_{k1} - \boldsymbol{Z}_{k2}\hat{\gamma}_{k2}$;

2) Fit a weighted least squares algorithm on ε againt $\boldsymbol{X}_k$ to get a new estimate for $\hat{\boldsymbol{\beta}}_k$;

3) Calculate the partial residuals for $\hat{\gamma}_{k1}$, i.e. $\boldsymbol{\varepsilon} = \boldsymbol{z}_k - \boldsymbol{X}_k\hat{\boldsymbol{\beta}}_k - \boldsymbol{Z}_{k2}\hat{\gamma}_{k2}$;

4) Fit a penalized weighted least squares algorithm on ε against $\boldsymbol{Z}_{k1}$ using weights $\boldsymbol{w}_k$ to obtain a new $\hat{\gamma}_{k1}$;

5) Calculate the partial residuals for $\hat{\gamma}_{k2}$, i.e. $\boldsymbol{\varepsilon} = \boldsymbol{z}_k - \boldsymbol{X}_k\hat{\boldsymbol{\beta}}_k - \boldsymbol{Z}_{k1}\hat{\gamma}_{k1}$;

6) Fit a penalized weighted least squares algorithm on ε against $\boldsymbol{Z}_{k2}$ using weights $\boldsymbol{w}_k$ to obtain a new $\hat{\gamma}_{k2}$;

The backfitting process is finished when $\hat{\boldsymbol{\beta}}_k$, $\hat{\gamma}_{k1}$ and $\hat{\gamma}_{k2}$ converge by some criterion.

APPENDIX C - The CG algorithm

The CG algorithm is divided in two different phases: the outer and the inner iteration. In the outer iteration, we need to create a working variable given by

$$\boldsymbol{z}_k = \boldsymbol{\eta}_k + \boldsymbol{w}_{ks}^{-1} \circ \boldsymbol{u}_k,$$

where $\boldsymbol{z}_k$ is defined in equation (2). For simplicity, we will consider that the distribution of the response variable we are working with has four parameters. Considering $k = 1, 2, 3, 4$ and $s = 1, 2, 3, 4$, where $k \leq s$, in our case, we can define the working weights $\boldsymbol{w}_{ks}$ as

$$\boldsymbol{w}_{ks} = \left(\frac{\partial l}{\partial \boldsymbol{\theta}_s}\right) \circ \left(\frac{\partial \boldsymbol{\theta}_s}{\partial \boldsymbol{\eta}_s}\right) \circ \boldsymbol{u}_k.$$

During the inner iteration of the CG algorithm, we shall consider a new working variable, defined by

$$\boldsymbol{z}_k^* = \boldsymbol{z}_k + \boldsymbol{z}_k',$$

where z_k' is a combination of the cross derivatives of the log-likelihood with respect to pairs of parameters of the distribution. Considering the four parameter distribution we will have

$$\begin{aligned}
\boldsymbol{z}_1' &= -\boldsymbol{w}_{11}^{-1} \circ \left[\boldsymbol{w}_{12} \circ \left(\boldsymbol{\eta}_2 - \boldsymbol{\eta}_2^0\right) + \boldsymbol{w}_{13} \circ \left(\boldsymbol{\eta}_3 - \boldsymbol{\eta}_3^0\right) + \boldsymbol{w}_{14} \circ \left(\boldsymbol{\eta}_4 - \boldsymbol{\eta}_4^0\right)\right] \\
\boldsymbol{z}_2' &= -\boldsymbol{w}_{22}^{-1} \circ \left[\boldsymbol{w}_{12} \circ \left(\boldsymbol{\eta}_2 - \boldsymbol{\eta}_2^0\right) + \boldsymbol{w}_{23} \circ \left(\boldsymbol{\eta}_3 - \boldsymbol{\eta}_3^0\right) + \boldsymbol{w}_{24} \circ \left(\boldsymbol{\eta}_4 - \boldsymbol{\eta}_4^0\right)\right] \\
\boldsymbol{z}_3' &= -\boldsymbol{w}_{33}^{-1} \circ \left[\boldsymbol{w}_{13} \circ \left(\boldsymbol{\eta}_2 - \boldsymbol{\eta}_2^0\right) + \boldsymbol{w}_{23} \circ \left(\boldsymbol{\eta}_3 - \boldsymbol{\eta}_3^0\right) + \boldsymbol{w}_{34} \circ \left(\boldsymbol{\eta}_4 - \boldsymbol{\eta}_4^0\right)\right] \\
\boldsymbol{z}_4' &= -\boldsymbol{w}_{44}^{-1} \circ \left[\boldsymbol{w}_{14} \circ \left(\boldsymbol{\eta}_2 - \boldsymbol{\eta}_2^0\right) + \boldsymbol{w}_{24} \circ \left(\boldsymbol{\eta}_3 - \boldsymbol{\eta}_3^0\right) + \boldsymbol{w}_{34} \circ \left(\boldsymbol{\eta}_4 - \boldsymbol{\eta}_4^0\right)\right].
\end{aligned}$$

Using the new adjusted working variables we fit a model for each parameter using the modified backfitting algorithm described in Appendix B. The inner process continues until the global deviance (GD) reaches a value that does not change. Then a new outer iteration is performed recalculating the quantities $\boldsymbol{z}_k$, $\boldsymbol{w}_{ks}$ and $\boldsymbol{\eta}_k^0$. The CG algorithm continues until the deviance evaluated in the outer iteration converges according to some criterion.