



## A model for bimodal rates and proportions

Roberto Vila, Lucas Alfaia, André F.B. Menezes, Mehmet N. Çankaya & Marcelo Bourguignon

To cite this article: Roberto Vila, Lucas Alfaia, André F.B. Menezes, Mehmet N. Çankaya & Marcelo Bourguignon (2022): A model for bimodal rates and proportions, Journal of Applied Statistics, DOI: [10.1080/02664763.2022.2146661](https://doi.org/10.1080/02664763.2022.2146661)


To link to this article: <https://doi.org/10.1080/02664763.2022.2146661>

 View supplementary material [↗](#)

 Published online: 18 Nov 2022.

 Submit your article to this journal [↗](#)





 Article views: 46

 View related articles [↗](#)

 View Crossmark data [↗](#)



## A model for bimodal rates and proportions

Roberto Vila <sup>a</sup>, Lucas Alfaia<sup>a</sup>, André F.B. Menezes <sup>b</sup>, Mehmet N. Çankaya <sup>c,d</sup> and Marcelo Bourguignon <sup>e</sup>

<sup>a</sup>Department of Statistics, Universidade de Brasília, Brasília, Brazil; <sup>b</sup>Department of Statistics, Universidade Estadual de Campinas, Campinas, Brazil; <sup>c</sup>Department of International Trading and Finance, Faculty of Applied Sciences, Uşak University, Uşak, Turkey; <sup>d</sup>Department of Statistics, Faculty of Art and Sciences, Uşak University, Uşak, Turkey; <sup>e</sup>Department of Statistics, Universidade Federal do Rio Grande do Norte, Natal, Brazil

### ABSTRACT

The beta model is the most important distribution for fitting data with the unit interval. However, the beta distribution is not suitable to model bimodal unit interval data. In this paper, we propose a bimodal beta distribution constructed by using an approach based on the alpha-skew-normal model. We discuss several properties of this distribution, such as bimodality, real moments, entropies and identifiability. Furthermore, we propose a new regression model based on the proposed model and discuss residuals. Estimation is performed by maximum likelihood. A Monte Carlo experiment is conducted to evaluate the performances of these estimators in finite samples with a discussion of the results. An application is provided to show the modelling competence of the proposed distribution when the data sets show bimodality.

### ARTICLE HISTORY

Received 16 March 2022  
Accepted 5 November 2022




### KEYWORDS

Bimodal model; bimodality; bounded data; beta distribution; maximum likelihood; regression model

## 1. Introduction

The need for modelling and analysing the bimodal bounded data, especially for data on the unit interval, occurs in many fields of real life, such as bioinformatics [12], image classification [16], transaction at a car dealership [26] and so on. In such situations, in order to apply probabilistic modelling for these phenomena, under a parametric paradigm, probability distributions limited to  $(0, 1)$  are indispensable. The unimodal beta model is the most widely used model in the literature to describe data in the unit interval, especially because of its flexibility and fruitful properties [13]. However, despite its broad sense applicability in many fields, the beta distribution is not suitable to model bimodal data on the unit interval.

In general, one uses mixtures of distributions to describe the bimodal data. For example, the studies [26] and [25] consider finite mixtures of beta regression models to analyze the priming effects in judgements of imprecise probabilities. However, in general, mixtures of distributions may suffer from identifiability problems in the parameter estimation;

**CONTACT** Roberto Vila  rovig161@gmail.com Department of Statistics,  Universidade de Brasília, Brasília, DF, Brazil Supplemental data for this article can be accessed online at  <http://dx.doi.org/10.1080/02664763.2022.2146661>.

see Refs. [14,15]. Thus, new mixture-free models which have the capacity to accommodate both unimodal and bimodal data are very important. The nature of phenomena can show bimodality due to many reasons, such as economical policies, uncertainty of social movement and its effects on the economy [28,30].

Since the structure of phenomena depends on many factors, it is reasonable to expect that the non-identically distributed data can occur in the data observed by the experimenter. For example, bimodality was introduced by Elal-Olivero [8], Domma *et al.* [6] and Vila and Çankaya [28] is a necessary probabilistic model to perform an efficient fitting on the non-identically distributed data set or the mixed data set. If we have the mixed data set, the mixed form of Beta and Weibull in Vila and Çankaya [28] should be necessary to model the data set efficiently, because the mixing proportions  $\pi_1$  and  $\pi_2 = (1 - \pi_1)$  in the bimodal case cannot be estimated accurately. The analytical expression of the mixed distribution can lead to problem while the optimization of the maximum likelihood estimation method according to parameters from parametric models, such as Beta, Weibull, etc. and the mixing parameters  $\pi_1$  and  $\pi_2$  of Beta from separate populations Beta<sub>1</sub> and Beta<sub>2</sub> is performed. At least, we can come across the numerical error while performing computation. The original working principle of phenomena can depend on the probabilistic model such as the bimodal beta (Bbeta) or the bimodal Weibull (BWeibull) in Vila and Çankaya [28]. On the other hand, the parameters in the mixed form of two Beta distributions, i.e. Beta<sub>1</sub> and Beta<sub>2</sub>, are  $\alpha_1, \alpha_2, \beta_1, \beta_2, \pi_1$  and  $\pi_2$ . However, Bbeta distribution includes four parameters which are  $\alpha, \beta, \rho$  and  $\delta$ . Bbeta has less parameters when compared with the mixed form of Beta. Further, since we have the exact expression for the cumulative distribution function of Bbeta, it is advantageous for us to generate the bimodal artificial data sets, which can be used to check whether or not a data set in the system can be modelled by Bbeta with the estimated parameters. In other words, the results and outputs on the unit interval data can be modelled and tested by using the proposed distribution.

Variations of the beta model can be found in Ferrari and Cribari-Neto [9], Ospina and Ferrari [21], Bayes *et al.* [4], Hahn [11], among others. However, all the models cited above are not suitable for capturing bimodality. Recently, probabilistic models for modelling bimodality on the positive real line were discussed by various authors. Olmos *et al.* [20] introduced recently a bimodal extension of the unit-Birnbaum-Saunders distribution. Vila *et al.* [29] proposed the bimodal gamma distribution. Vila and Çankaya [28] considered a bimodal Weibull distribution. Recently, [17] proposed a family of bimodal distributions generated by distributions with positive support. Despite this, to the best of our knowledge, a specific parametric model to describe bimodality data observed of the unit interval has never been considered in the literature recently. Despite this, to the best of our knowledge, a specific regression bimodal model to unit interval data with a regression structure for the parameters has never been considered in the literature. Martinez-Flrez *et al.* [18] considered a transformation in a random variable that follows a unit-bimodal Birnbaum-Saunders (UBBS for short) distribution only in the case of identically and independently variables.

Based on the above discussion and motivated by the presence of bimodality in proportion responses, we develop a model for double-bounded response variables. In particular, we extended the usual beta distribution using a quadratic transformation technique used to generate bimodal functions [8,28]. The approach, therefore, appears to be a new development for the literature. We discuss several properties of the proposed model, such as

bimodality, real moments, hazard rate, entropies and identifiability. Furthermore, we study the effects of the explanatory variables on the response variable using a regression model.

In what follows, we list some of the main contributions and advantages of the proposed model.

- We introduce a new family of distributions that is flexible version of the usual beta distribution so that it is capable of fitting bimodal as well as unimodal data. We provide general properties of the proposed model;
- We propose an extended version of the quadratic transformation technique used to generate bimodal functions;

The rest of the article proceeds as follows. In Sections 2 and 3, we present the new distribution and derive some of its properties. Then in Section 4, we present the main properties of the bimodal beta, which include entropies, stochastic representation and identifiability. Section 5 presents the bimodal beta regression model. Also, the estimation method for the model parameters and diagnostic measures are discussed. In Section 6, some numerical results of the estimators and the empirical distribution of the residuals are presented with a discussion of the results. A real-life application related to the proportion of votes that Jair Bolsonaro received in the second turn of Brazilian elections in 2018 is analysed in Section 7. Section 8 summarizes the main findings of the paper.

## 2. The bimodal beta distribution

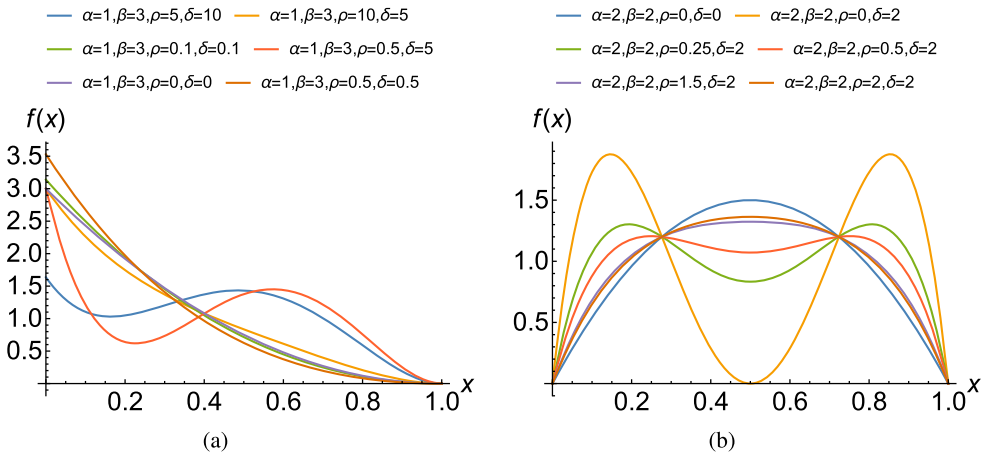
In this section, the bimodal beta (Bbeta) distribution is introduced and its density is derived. Moreover, some results on the bimodality properties are obtained. We say that a random variable (r.v.)  $X$  has a Bbeta distribution with parameter vector  $\theta_\delta = (\alpha, \beta, \rho, \delta)$ ,  $\alpha > 0, \beta > 0, \rho \geq 0$  and  $\delta \in \mathbb{R}$ , denoted by  $X \sim \text{Bbeta}(\theta_\delta)$ , if its probability density function (PDF) is given by

$$f(x; \theta_\delta) = \begin{cases} \frac{\rho + (1 - \delta x)^2}{Z(\theta_\delta)B(\alpha, \beta)} x^{\alpha-1} (1 - x)^{\beta-1}, & 0 < x < 1, \\ 0, & \text{otherwise,} \end{cases} \tag{1}$$

where

$$Z(\theta_\delta) = 1 + \rho - 2\delta \frac{\alpha}{\alpha + \beta} + \delta^2 \frac{\alpha(\alpha + 1)}{(\alpha + \beta)(\alpha + \beta + 1)} \tag{2}$$

denotes the normalization constant and  $B(\alpha, \beta) = \int_0^1 t^{\alpha-1} (1 - t)^{\beta-1} dt$  is the beta function. When  $\delta = 2, \alpha = \beta = 1$  and  $\rho = 0$ , we have the U-quadratic distribution on  $(0, 1)$ . When  $\delta = 0$ , we obtain the classical beta distribution with parameter vector  $\theta_0 = (\alpha, \beta, \rho, 0) := (\alpha, \beta)$ . The parameters  $\alpha, \beta$  (which appear as exponents of the r.v.) and  $\rho$  control the shape of the distribution. The uni- or bimodality is controlled by the parameter  $\delta$ . Note that for  $\alpha, \beta$  and  $\delta \neq 0$  fixed, the parameter  $\rho$  also controls the unimodality or bimodality of the distribution; see Subsection 2.1. From Figure 1, we note some different shapes of the Bbeta PDF for different combinations of parameters. Figure 1(a,b) represents L shape and its bimodal form and bell shaped case of beta distribution, respectively.



**Figure 1.** The PDF of the bimodal beta distribution for different values of parameters. In the figure on the left, the PDF presents strict decreasing monotonicity or decreasing-increasing-decreasing shapes. In the figure on the right, the PDF shows symmetry and uni- or bimodality.

Unlike Figure 1(b), Figure 2(a,b) shows that, a peak can be major peak and the other one can be minor peak.

The asymptotic behaviour of the PDF (1) is as follows:

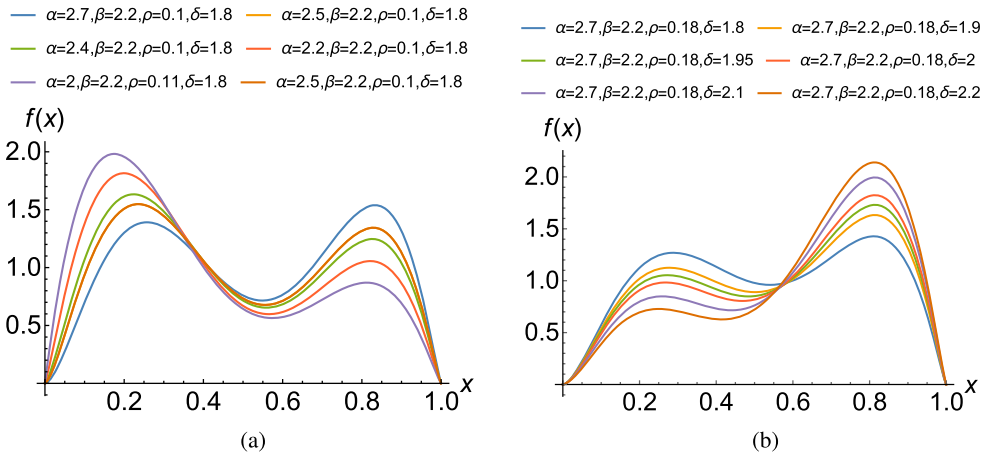
$$f(0^+; \theta_\delta) = \lim_{\substack{x \rightarrow 0, \\ x > 0}} f(x; \theta_\delta) = \begin{cases} \frac{\beta(\rho + 1)}{1 + \rho - \frac{2\delta}{1+\beta} + \frac{2\delta^2}{(1+\beta)(2+\beta)}}, & \alpha = 1, \\ 0, & \alpha > 1, \\ +\infty, & \alpha < 1, \end{cases} \quad (3)$$

and

$$f(1^-; \theta_\delta) = \lim_{\substack{x \rightarrow 1, \\ x < 1}} f(x; \theta_\delta) = \begin{cases} \frac{\alpha[\rho + (1 - \delta)^2]}{1 + \rho - \frac{2\delta\alpha}{\alpha+1} + \frac{\delta^2\alpha(\alpha+1)}{(\alpha+1)(\alpha+2)}}, & \beta = 1, \\ 0, & \beta > 1, \\ +\infty, & \beta < 1. \end{cases} \quad (4)$$

This asymptotic behaviour of the Bbeta PDF was expected, since the bimodal beta distribution is defined in terms of the classical beta. It is clear that when  $\delta = 0$ ;  $f(0^+; \theta_\delta) = \beta$  for  $\alpha = 1$  and  $f(1^-; \theta_\delta) = \alpha$  for  $\beta = 1$ .

If  $X \sim \text{Bbeta}(\theta_\delta)$ , the cumulative distribution function (CDF) (see Figure 3), the survival function (SF) and the hazard rate function (HR) of  $X$  are, respectively, given



**Figure 2.** The PDF of the bimodal beta distribution for different values of parameters. Both figures present asymmetry and bimodality.

by

$$F(x; \theta_\delta) = \frac{1}{Z(\theta_\delta)} \left[ (1 + \rho) I_x(\alpha, \beta) - 2\delta \frac{B_x(\alpha + 1, \beta)}{B(\alpha, \beta)} + \delta^2 \frac{B_x(\alpha + 2, \beta)}{B(\alpha, \beta)} \right], \quad (5)$$

$$S(x; \theta_\delta) = \frac{1}{Z(\theta_\delta)} \sum_{i=0}^2 c_i \left[ \frac{B(\alpha + i, \beta)}{B(\alpha, \beta)} - \frac{B_x(\alpha + i, \beta)}{B(\alpha, \beta)} \right] \quad \text{and} \quad (6)$$

$$H(x; \theta_\delta) = \frac{[\rho + (1 - \delta x)^2] x^{\alpha-1} (1 - x)^{\beta-1}}{\sum_{i=0}^2 c_i [B(\alpha + i, \beta) - B_x(\alpha + i, \beta)]}, \quad (7)$$

where  $I_x(\alpha, \beta) = B_x(\alpha, \beta)/B(\alpha, \beta)$  is the incomplete beta function ratio,  $B_x(\alpha, \beta) = \int_0^x t^{\alpha-1} (1 - t)^{\beta-1} dt$  is the incomplete beta function, and  $c_0 = 1 + \rho$ ,  $c_1 = -2\delta$ ,  $c_2 = \delta^2$ . For more details on the derivation of these formulas, see Section 3.

### 2.1. Bimodality properties

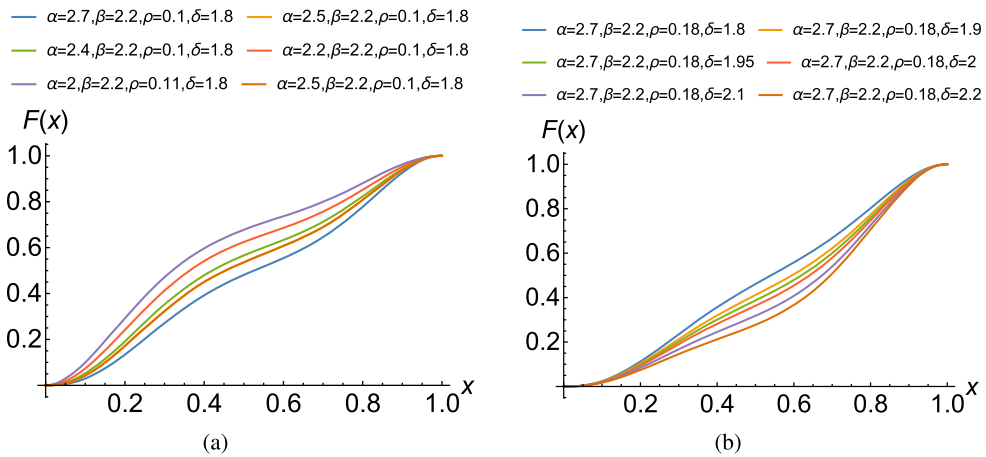
To state the following result that guarantees the uni- or bimodality of the Bbeta distribution, we define the following cubic polynomial:

$$p_3(x) = a_3x^3 + a_2x^2 + a_1x + a_0 = 0, \quad (8)$$

where  $a_3 = -\delta^2(\alpha + \beta)$ ,  $a_2 = \delta[\alpha(\delta + 2) + 2\beta + \delta - 2]$ ,  $a_1 = -[\alpha(2\delta + \rho + 1) + (\beta - 2)(\rho + 1)]$  and  $a_0 = (\alpha - 1)(\rho + 1)$ .

**Theorem 2.1 (Uni- or bimodality):** *Let  $X \sim \text{Bbeta}(\theta_\delta)$  such that  $\alpha > 1, \beta > 1, (\delta - 1)^2 + \rho > 0$  and  $\delta > 0$ .*

- (i) *If  $p_3(x)$  has a single positive zero then the Bbeta distribution is unimodal.*
- (ii) *If  $p_3(x)$  has exactly three zeros in  $(0, 1)$  then the Bbeta distribution is bimodal.*



**Figure 3.** The CDF of the bimodal beta distribution for different values of parameters. Due to bimodality, as seen in the figure, it is natural to expect the CDF graph to present up to three inflection points.

**Proof:** A simple computation shows that

$$f'(x; \theta_\delta) = \frac{x^{\alpha-2}(1-x)^{\beta-2}}{Z(\theta_\delta)B(\alpha, \beta)} p_3(x), \tag{9}$$

where  $p_3(x)$  is given in (8). Under the conditions stated in the theorem, we have  $a_3 < 0$ ,  $a_2 > 0$ ,  $a_1 < 0$  and  $a_0 > 0$ . By definition, the boundary points are never critical points, then we exclude the analysis at these points.

Since  $p_3(0) = a_0 > 0$  and  $p_3(1) = a_3 + a_2 + a_1 + a_0 = (1 - \beta)[(\delta - 1)^2 + \rho] < 0$  because  $\beta > 1$  and  $(\delta - 1)^2 + \rho > 0$ , the Intermediate Value Theorem guarantees that there is at least one root in the interval  $(0, 1)$ . Further, by Descartes rule of signs (see, e.g. Refs. [31] and [10]),  $p_3(x)$  has one or three roots in the interval  $(0, 1)$ .

Assume that  $p_3(x)$  has a single zero. In this case,  $f(x; \theta_\delta)$  has a single critical point, denoted by  $x_0$ . Since, for  $\alpha > 1$  and  $\beta > 1$ ,  $f(0^+; \theta_\delta) = 0$  and  $f(1^-; \theta_\delta) = 0$ , see limits in (3) and (4); it follows that  $f(x; \theta_\delta)$  increases on  $(0, x_0)$  and decreases on  $(x_0, 1)$ . That is,  $x_0$  is a global maximum point of  $f(x; \theta_\delta)$ . This proves Item (i).

On the other hand, if  $p_3(x)$  has exactly three zeros in  $(0, 1)$  then  $f(x; \theta_\delta)$  has three critical points  $x_1, x_2$  and  $x_3$ . Without loss of generality, let us assume that  $x_1 < x_2 < x_3$ . Again, since, for  $\alpha > 1$  and  $\beta > 1$ ,  $f(0^+; \theta_\delta) = 0$  and  $f(1^-; \theta_\delta) = 0$ , it follows that  $f(x; \theta_\delta)$  increases on the intervals  $(0, x_1)$  and  $(x_2, x_3)$ , and decreases on  $(x_1, x_2)$  and  $(x_3, 1)$ . In other words,  $x_1$  and  $x_3$  are two maximum points and  $x_2$  is the unique minimum point. Then the statement in Item (ii) follows.

Thus we have completed the proof of the theorem. ■

**Remark 2.1:** By considering  $\alpha, \beta, \delta$  and  $\rho$  as in the Table 1, it is clear that the conditions of Theorem 2.1 are satisfied. Then, depending on the number of roots of  $p_3(x)$ , Theorem 2.1 guarantees the uni- or bimodality (U- or B) of PDF  $f(x; \theta_\delta)$ . These results are compatible with Figure 1(b).

**Table 1.** Roots of the polynomial  $p_3(x)$  and shapes of the PDF bimodal beta using the values of the parameters of Figure 1(b).

$\alpha$	$\beta$	$\delta$	$\rho$	$a_3$	$a_2$	$a_1$	$a_0$	Real roots of $p_3(x)$ in $(0, 1)$	Shape
2	2	2	0.25	-16	24	-10.5	1.25	$x = 0.19, x = 0.5, x = 0.81$	B
2	2	2	1.5	-16	24	-13	2.5	$x = 0.5$	U
2	2	2	0	-16	24	-10	1	$x = 0.15, x = 0.5, x = 0.85$	B
2	2	2	0.5	-16	24	-11	1.5	$x = 0.25, x = 0.5, x = 0.75$	B
2	2	2	2	-16	24	-14	3	$x = 0.5$	U

Again, using the values of Figure 2(a, b), it can be verified that the conditions of Theorem 2.1 are satisfied, which allows concluding the bimodality of the PDF  $f(x; \theta_\delta)$ . This contrasts the shape of PDF shown in Figure 2.

**Theorem 2.2 (Bimodality; case  $\rho = 0$ ):** *If  $X \sim \text{Bbeta}(\theta_\delta)$ ,  $\alpha > 1, \beta > 1, \rho = 0, \delta > 1$ , and*

$$[\delta(\alpha + 1) + \alpha + \beta - 2]^2 > 4\delta(\alpha + \beta)(\alpha - 1), \tag{10}$$

*then the Bbeta distribution is bimodal with maximum points*

$$x_{\max, \pm} = \frac{1}{\delta} + \frac{\delta(\alpha + 1) - (\alpha + \beta + 2)}{2\delta(\alpha + \beta)} \pm \frac{\sqrt{[\delta(\alpha + 1) + \alpha + \beta - 2]^2 - 4\delta(\alpha + \beta)(\alpha - 1)}}{2\delta(\alpha + \beta)}$$

*and minimum point  $x = 1/\delta$ , where  $0 < x_{\max, -} < x = 1/\delta < x_{\max, +} < 1$ .*

**Proof:** Taking  $\rho = 0$  in (9), we have

$$f'(x; \theta_\delta) = \frac{x^{\alpha-2}(1-x)^{\beta-2}}{Z(\theta_\delta)B(\alpha, \beta)} (1-\delta x) \left\{ (\alpha + \beta)\delta x^2 - [\delta(\alpha + 1) + \alpha + \beta - 2]x + (\alpha - 1) \right\}.$$

A direct calculus shows that  $f'(x; \theta_\delta) = 0$  if and only if (excluding the boundary points)  $x = 1/\delta$  and

$$x_{\pm} = \frac{\delta(\alpha + 1) + \alpha + \beta - 2 \pm \sqrt{[\delta(\alpha + 1) + \alpha + \beta - 2]^2 - 4\delta(\alpha + \beta)(\alpha - 1)}}{2\delta(\alpha + \beta)} = x_{\max, \pm}.$$

Hence, under condition (10), it follows that the equation  $f'(x; \theta_\delta) = 0$  has three roots  $x = 1/\delta, x_-$  and  $x_+$  within the interval  $(0, 1)$ , where  $x_- < x = 1/\delta < x_+$ .

Since, for  $\alpha > 1$  and  $\beta > 1, f(0^+; \theta_\delta) = 0$  and  $f(1^-; \theta_\delta) = 0$ , see limits in (3) and (4); the bimodality of the Bbeta distribution is guaranteed, where  $x_- = x_{\max, -}$  and  $x_+ = x_{\max, +}$  are two maximum points and  $x = 1/\delta$  is the unique minimum point. ■

**Remark 2.2:** Let  $\alpha = \beta = \delta = 2$  and  $\rho = 0$ . It is clear that the condition (10) is satisfied. Then, by Theorem 2.2, the Bbeta distribution is bimodal with maximum points  $x_{\max, -} = (2 - \sqrt{2})/4 \approx 0.15$  and  $x_{\max, +} = (2 + \sqrt{2})/4 \approx 0.85$ , and minimum point  $x = 1/2$ ; which is compatible with Figure 1(b).



**Table 2.** Modes, maximum values and shapes of the PDF bimodal beta using the parameter values in Figure 1 (b) and Figure 2(a)–(b).

$\alpha$	$\beta$	$\delta$	$\rho$	$x_{\max,-}^*$	$x_{\max,+}^*$	$f(x_{\max,-}^*; \theta_\delta)$	$f(x_{\max,+}^*; \theta_\delta)$	Shape
2	2	2	0.25	0.19	0.81	1.30	1.30	B
2	2	2	0	0.15	0.85	1.87	1.87	B
2	2	2	0.5	0.25	0.75	1.21	1.21	B

**Proposition 2.3:** *The Bbeta PDF  $f(x; \theta_\delta)$  is symmetric at the point  $x = 1/2$  whenever  $\alpha = \beta$  and  $\delta = 2$ .*

**Proof:** A simple algebraic manipulation shows that, if  $\alpha = \beta$  and  $\delta = 2$  then  $f(0.5 - x; \theta_\delta) = f(0.5 + x; \theta_\delta)$ ,  $\forall 0 < x < 1$ . Then the proof follows. ■

**Theorem 2.3:** *If  $X \sim \text{Bbeta}(\theta_\delta)$ ,  $\alpha = \beta > 1$ ,  $\rho < 1/(\alpha - 1)$  and  $\delta = 2$ , then the Bbeta distribution is bimodal with maximum points*

$$x_{\max,\pm}^* = \frac{1}{2} \pm \frac{\sqrt{\rho\alpha(1-\alpha) + \alpha}}{2\alpha}$$

and minimum point  $x = 1/2$ , where  $0 < x_{\max,-}^* < x = 1/2 < x_{\max,+}^* < 1$ . Moreover, the maximum values coincide, that is,  $f(x_{\max,-}^*; \theta_\delta) = f(x_{\max,+}^*; \theta_\delta)$ .

**Proof:** As a by-product of proof of the Theorem 2.1, we have  $f'(x; \theta_\delta) = 0$  if and only if  $x$  is a zero of polynomial  $p_3(x)$  defined in (8). Setting  $\alpha = \beta > 1$  and  $\delta = 2$  in polynomial  $p_3(x)$ , we get  $f'(x; \theta_\delta) = 0$  if and only if

$$p_3^*(x) = -8\alpha x^3 + 12\alpha x^2 - [\alpha(5 + \rho) + (\alpha - 2)(\rho + 1)]x + (\alpha - 1)(\rho + 1) = 0.$$

Note that the above polynomial can be written as  $p_3^*(x) = 2(x - (1/2))[4\alpha x^2 - 4\alpha x + (\alpha - 1)(\rho + 1)]$ . Then, it is clear that  $x = 1/2$  and  $x_{\max,\pm}^*$  are critical points of  $f(x; \theta_\delta)$ , where  $0 < x_{\max,-}^* < x = 1/2 < x_{\max,+}^* < 1$ . Note that the restriction  $\rho < 1/(\alpha - 1)$  guarantees that the discriminant of the quadratic polynomial implicit in  $p_3^*(x)$  is positive.

By using that  $\alpha > 1$  and  $\beta > 1$ , and by following the same steps as in the final paragraph of proof of the Theorem 2.2, we guarantee bimodality of the Bbeta distribution.

Finally, the identity  $f(x_{\max,-}^*; \theta_\delta) = f(x_{\max,+}^*; \theta_\delta)$  follows from Proposition 2.3. ■

**Remark 2.4:** By considering  $\alpha$ ,  $\beta$ ,  $\delta$  and  $\rho$  as in the Table 2, it is clear that the restriction  $\rho < 1/(\alpha - 1)$  is satisfied. Then, Theorem 2.3 guarantees the bimodality (B) of PDF  $f(x; \theta_\delta)$  with minimum point  $x = 1/2$  and points (and values) of maximum specified in this table. These results are compatible with Figures 1(b) and 2(a)–(b).

### 3. Moments

In this section, some closed expressions for truncated moments and real moments of the Bbeta distribution are obtained. Other properties as raw moments, mean residual life function and moment generating function were also analysed in Section I of the Supplementary Material.

**Theorem 3.1:** If  $X \sim \text{Bbeta}(\theta_\delta)$  then, for  $0 \leq a < b \leq 1$  and  $r > -\alpha$ ,

$$\mathbb{E}(X^r \mathbb{1}_{\{a \leq X \leq b\}}) = \frac{1}{Z(\theta_\delta)} \sum_{i=0}^2 c_i \left[ \frac{B_b(\alpha + r + i, \beta)}{B(\alpha, \beta)} - \frac{B_a(\alpha + r + i, \beta)}{B(\alpha, \beta)} \right],$$

where  $c_0 = 1 + \rho$ ,  $c_1 = -2\delta$ ,  $c_2 = \delta^2$ , and  $B_x(\alpha, \beta)$  is the incomplete beta function.

**Proof:** By using definition of expectation and definition of Bbeta density, we have

$$\mathbb{E}(X^r \mathbb{1}_{\{a \leq X \leq b\}}) = \frac{1}{Z(\theta_\delta)} \sum_{i=0}^2 c_i \mathbb{E}(Y^{r+i} \mathbb{1}_{\{a \leq Y \leq b\}}), \quad Y \sim \text{Bbeta}(\theta_0).$$

Since  $\mathbb{E}(Y^{r+i} \mathbb{1}_{\{a \leq Y \leq b\}}) = [B_b(\alpha + r + i, \beta) - B_a(\alpha + r + i, \beta)]/B(\alpha, \beta)$ , the proof of the theorem follows. ■

Taking  $r = 0$ ,  $b = x$  and  $a = 0$  in Theorem 3.1, we get the formula (5) for the CDF. Letting  $r = 0$ ,  $b = 1$  and  $a = x$  in Theorem 3.1, we get the formula (6) for the SF. By combining the formula (6) of CDF and definition of the Bbeta distribution, we obtain the formula (7) for the HR.

Taking  $r = 1$ ,  $a = x$  and  $b = 1$  in Theorem 3.1, we get a closed formula for the mean residual life function, see Corollary 1.1 of the Supplementary Material.

**Corollary 3.1 (Real moments):** If  $X \sim \text{Bbeta}(\theta_\delta)$  and  $r > -\alpha$ , then

$$\mathbb{E}(X^r) = \frac{1}{Z(\theta_\delta)} \left[ (1 + \rho) \frac{B(\alpha + r, \beta)}{B(\alpha, \beta)} - 2\delta \frac{B(\alpha + r + 1, \beta)}{B(\alpha, \beta)} + \delta^2 \frac{B(\alpha + r + 2, \beta)}{B(\alpha, \beta)} \right].$$

**Proof:** By taking  $b = 1$  and  $a = 0$  in Theorem 3.1 we have the following:

$$\mathbb{E}(X^r) = \frac{1}{Z(\theta_\delta)} \sum_{i=0}^2 c_i \frac{B(\alpha + r + i, \beta)}{B(\alpha, \beta)},$$

where  $c_0 = 1 + \rho$ ,  $c_1 = -2\delta$  and  $c_2 = \delta^2$ . ■

As a consequence of the above corollary, the closed expressions for the standardized moments, variance, skewness and kurtosis of the bimodal beta r.v.  $X$  are easily obtained.

#### 4. Further properties

In this section, we consider some properties of the Bbeta distribution, such as stochastic representation and identifiability. For reasons of space, entropy measures such as Tsallis [27], quadratic [23] and Shannon [24] ones were studied in Section II of the Supplementary Material.

#### 4.1. Stochastic representation

Let  $W$  be a discrete r.v. with the following probability function  $\mathbb{P}(W = w_k) = \pi_k$ ,  $k = 0, 1, 2$ , where

$$\pi_0 = \frac{1 + \rho}{Z(\boldsymbol{\theta}_\delta)}, \quad \pi_1 = -\frac{2\alpha\delta}{Z(\boldsymbol{\theta}_\delta)(\alpha + \beta)}, \quad \pi_2 = \frac{\alpha(\alpha + 1)\delta^2}{Z(\boldsymbol{\theta}_\delta)(\alpha + \beta)(\alpha + \beta + 1)}, \quad \text{for } \delta < 0,$$

and  $Z(\boldsymbol{\theta}_\delta)$  is as in (2). Notice that  $\pi_0 + \pi_1 + \pi_2 = 1$ .

Let's consider the following three r.v.'s:  $Y_{0;\alpha,\beta} \sim \text{Beta}(\alpha, \beta)$ ,  $Y_{1;\alpha+1,\beta} \sim \text{Beta}(\alpha + 1, \beta)$  and  $Y_{2;\alpha+2,\beta} \sim \text{Beta}(\alpha + 2, \beta)$ . Then we define a new r.v.  $X$  as follows:

$$X = Y_{0;\alpha,\beta} \mathbb{1}_{\{W=w_0\}} + Y_{1;\alpha+1,\beta} \mathbb{1}_{\{W=w_1\}} + Y_{2;\alpha+2,\beta} \mathbb{1}_{\{W=w_2\}}, \quad (11)$$

where  $W$  is independent of  $Y_{0;\alpha,\beta}$ ,  $Y_{1;\alpha+1,\beta}$  and  $Y_{2;\alpha+2,\beta}$ .

**Proposition 4.1 (Stochastic representation for  $\delta < 0$ ):** *If  $X$  admits the form (11), then  $X \sim \text{Bbeta}(\boldsymbol{\theta}_\delta)$ . Conversely, if  $X \sim \text{Bbeta}(\boldsymbol{\theta}_\delta)$  then  $X$  is as in (11).*

**Proof:** Using the law of total probability and the definition of  $X$ , we get

$$\begin{aligned} \mathbb{P}(X \leq x) &= \mathbb{P}(Y_{0;\alpha,\beta} \leq x | W = w_0)\pi_0 + \mathbb{P}(Y_{1;\alpha+1,\beta} \leq x | W = w_1)\pi_1 \\ &\quad + \mathbb{P}(Y_{2;\alpha+2,\beta} \leq x | W = w_2)\pi_2 \\ &= \mathbb{P}(Y_{0;\alpha,\beta} \leq x)\pi_0 + \mathbb{P}(Y_{1;\alpha+1,\beta} \leq x)\pi_1 + \mathbb{P}(Y_{2;\alpha+2,\beta} \leq x)\pi_2, \end{aligned}$$

where in the last line we used the independence of  $W$  with respect to variables  $Y_{1;\alpha,\beta}$ ,  $Y_{2;\alpha+1,\beta}$  and  $Y_{3;\alpha+2,\beta}$ . Since  $\mathbb{P}(Y_{k;\alpha+k,\beta} \leq x) = I_x(\alpha + k, \beta)$ ,  $k = 0, 1, 2$ , the above equality becomes

$$\frac{1}{Z(\boldsymbol{\theta}_\delta)} \left[ (1 + \rho) I_x(\alpha, \beta) - 2\delta \frac{B_x(\alpha + 1, \beta)}{B(\alpha, \beta)} + \delta^2 \frac{B_x(\alpha + 2, \beta)}{B(\alpha, \beta)} \right].$$

But, by (5), the right-hand side is equal to the CDF  $F(x; \boldsymbol{\theta}_\delta)$ .

Then we have completed the proof. ■

#### 4.2. Identifiability

A simple observation shows that the bimodal beta PDF  $f(x; \boldsymbol{\theta}_\delta)$  in (1), with parameter vector  $\boldsymbol{\theta}_\delta = (\alpha, \beta, \rho, \delta)$ , can be written as a finite (generalized) mixture of three beta distributions with different shape parameters, i.e.

$$f(x; \boldsymbol{\theta}_\delta) = \pi_0 f(x; \alpha, \beta) + \pi_1 f(x; \alpha + 1, \beta) + \pi_2 f(x; \alpha + 2, \beta), \quad 0 < x < 1, \quad (12)$$

where  $\pi_0, \pi_1$  and  $\pi_2$  are constants (that depends only on  $\boldsymbol{\theta}_\delta$ ) given in Proposition 4.1, and  $f(x; \alpha, \beta) = x^{\alpha-1}(1-x)^{\beta-1}/B(\alpha, \beta)$ ,  $0 < x < 1$ ,  $(\alpha > 0, \beta > 0)$ , denotes the standard beta PDF.

Unlike Proposition 4.1, here  $\delta$  can be non-negative. In principle, mixing non-negative weights are not necessary since mixtures can be PDF even if some of weights are negative.

Let  $\mathcal{B}$  be the family of beta distributions, as follows:

$$\mathcal{B} = \left\{ F : F(x; \alpha, \beta) = \int_0^x f(y; \alpha, \beta) dy, \alpha > 0, \beta > 0, 0 < x < 1 \right\}.$$

Write  $\mathcal{H}_{\mathcal{B}}$  as the class of all finite mixtures of  $\mathcal{B}$ . It is well-known that  $\mathcal{H}_{\mathcal{B}}$  is not identifiable; see the main Theorem of Ahmad and Al-Hussaini [1]. Let  $\mathcal{H}_{\mathcal{B}^*}$  be the class of all finite mixtures of  $\mathcal{B}$  with the restriction that the shape parameters  $\beta$  are pairwise different (that is,  $\beta_i \neq \beta_j$  for  $i \neq j$ ). As a consequence of the main result of Atienza *et al.* [3], it is a simple task to prove that the class  $\mathcal{H}_{\mathcal{B}^*}$  is identifiable; see, e.g. Proposition 3.2.2 of de Alencar [5] or Proposition 1.2 in the Appendix of Alfaia [2].

The following result proves the identifiability of bimodal beta distribution.

**Proposition 4.2:** *The mapping  $\theta_{\delta} = (\alpha, \beta, \rho, \delta) \mapsto f(\cdot; \theta_{\delta})$ , where the  $\beta$ 's are pairwise different, is one-to-one.*

**Proof:** Let us suppose that  $f(x; \theta_{\delta_i}) = f(x; \theta_{\delta_j})$  for all  $0 < x < 1$ , where  $\theta_{\delta_i} = (\alpha_i, \beta_i, \rho_i, \delta_i)$  and  $\theta_{\delta_j} = (\alpha_j, \beta_j, \rho_j, \delta_j)$ . In other words, by (12),

$$\begin{aligned} &\pi_{0;i}f(x; \alpha_i, \beta_i) + \pi_{1;i}f(x; \alpha_i + 1, \beta_i) + \pi_{2;i}f(x; \alpha_i + 2, \beta_i) \\ &= \pi_{0;j}f(x; \alpha_j, \beta_j) + \pi_{1;j}f(x; \alpha_j + 1, \beta_j) + \pi_{2;j}f(x; \alpha_j + 2, \beta_j), \end{aligned}$$

where  $\pi_{k;i}$  and  $\pi_{k;j}$ ,  $k = 0, 1, 2$ , are defined as in Proposition 4.1. Since  $\mathcal{H}_{\mathcal{B}^*}$  is identifiable, we have  $\pi_{k;i} = \pi_{k;j}$ , for  $k = 0, 1, 2$ , and  $\alpha_i = \alpha_j$ ,  $\beta_i = \beta_j$ . Hence, from equalities  $\pi_{k;i} = \pi_{k;j}$ ,  $k = 0, 1, 2$ , immediately follows that  $\rho_i = \rho_j$  and  $\delta_i = \delta_j$ . Therefore,  $\theta_{\delta_i} = \theta_{\delta_j}$ , and the proof follows. ■

### 5. Regression model, estimation and diagnostic analysis

Let  $X_1, \dots, X_n$  be  $n$  independent random variables, where each  $X_i$ ,  $i = 1, \dots, n$ , follows the PDF given in (1). We assume that the parameters  $\alpha_i$  and  $\beta_i$  satisfy the following functional relations:

$$g_1(\alpha_i) = \eta_{1i} = \mathbf{w}_i^{\top} \boldsymbol{\gamma} \quad \text{and} \quad g_2(\beta_i) = \eta_{2i} = \mathbf{z}_i^{\top} \boldsymbol{\zeta}, \tag{13}$$

where  $\boldsymbol{\gamma} = (\gamma_1, \dots, \gamma_p)^{\top}$  and  $\boldsymbol{\zeta} = (\zeta_1, \dots, \zeta_q)^{\top}$  are vectors of unknown regression coefficients which are assumed to be functionally independent,  $\boldsymbol{\gamma} \in \mathbb{R}^p$  and  $\boldsymbol{\zeta} \in \mathbb{R}^q$ , with  $p + q < n$ ,  $\eta_{1i}$  and  $\eta_{2i}$  are the linear predictors, and  $\mathbf{w}_i = (w_{i1}, \dots, w_{ip})^{\top}$  and  $\mathbf{z}_i = (z_{i1}, \dots, z_{iq})^{\top}$  are observations on  $p$  and  $q$  known regressors, for  $i = 1, \dots, n$ . Furthermore, we assume that the covariate matrices  $\mathbf{W} = (\mathbf{w}_1, \dots, \mathbf{w}_n)^{\top}$  and  $\mathbf{Z} = (\mathbf{z}_1, \dots, \mathbf{z}_n)^{\top}$  have rank  $p$  and  $q$ , respectively. The link functions  $g_1 : \mathbb{R} \rightarrow \mathbb{R}^+$  and  $g_2 : \mathbb{R} \rightarrow \mathbb{R}^+$  in (13) must be strictly monotone, positive and at least twice differentiable, such that  $\alpha_i = g_1^{-1}(\mathbf{w}_i^{\top} \boldsymbol{\gamma})$  and  $\beta_i = g_2^{-1}(\mathbf{z}_i^{\top} \boldsymbol{\zeta})$ , with  $g_1^{-1}(\cdot)$  and  $g_2^{-1}(\cdot)$  being the inverse functions of  $g_1(\cdot)$  and  $g_2(\cdot)$ , respectively. There are several possible choices for the link functions  $g_1(\cdot)$  and  $g_2(\cdot)$ . For example, one can use the logarithmic specification  $g_j(\cdot) = \log(\cdot)$ , square root  $g_j(\cdot) = \sqrt{\cdot}$ , or identity  $g_j(\cdot) = \cdot$  (with special attention to the sign of the estimates),  $j = 1, 2$ . In this paper, we consider the log link,  $g_j(\cdot) = \log(\cdot)$ , since it is the most used for positive parameters.

The log-likelihood function for  $\boldsymbol{\theta}_\delta = (\boldsymbol{\gamma}, \boldsymbol{\zeta}, \rho, \delta)$  based on a sample of  $n$  independent observations is given by

$$\ell(\boldsymbol{\theta}_\delta) = \sum_{i=1}^n \ell(\alpha_i, \beta_i, \rho, \delta), \quad (14)$$

where  $\ell(\alpha_i, \beta_i, \rho, \delta) = -\log Z(\boldsymbol{\theta}_\delta) - \log B(\alpha_i, \beta_i) + \log[\rho + (1 - \delta x_i)^2] + (\alpha_i - 1) \log x_i + (\beta_i - 1) \log(1 - x_i)$ ,  $i = 1, \dots, n$ , and  $Z(\boldsymbol{\theta}_\delta)$  is as in (2).

The maximum likelihood estimator (MLE)  $\widehat{\boldsymbol{\theta}}_\delta = (\widehat{\boldsymbol{\gamma}}^\top, \widehat{\boldsymbol{\zeta}}^\top, \widehat{\rho}, \widehat{\delta})^\top$  of  $\boldsymbol{\theta}_\delta = (\boldsymbol{\gamma}^\top, \boldsymbol{\zeta}^\top, \rho, \delta)^\top$  is obtained by the maximization of the log-likelihood function (14). However, it is not possible to derive analytical solution for the MLE  $\widehat{\boldsymbol{\theta}}$ , hence we resort to numerical solution using some optimization algorithm, such as Newton-Raphson and quasi-Newton.

Under mild regularity conditions and when  $n$  is large, the asymptotic distribution of the MLE  $\widehat{\boldsymbol{\theta}}_\delta = (\widehat{\boldsymbol{\gamma}}^\top, \widehat{\boldsymbol{\zeta}}^\top, \widehat{\rho}, \widehat{\delta})^\top$  is approximately multivariate normal (of dimension  $p + q + 2$ ) with mean vector  $\boldsymbol{\theta}_\delta = (\boldsymbol{\gamma}^\top, \boldsymbol{\zeta}^\top, \rho, \delta)^\top$  and variance covariance matrix  $\mathbf{K}^{-1}(\boldsymbol{\theta}_\delta)$  where  $\mathbf{K}(\boldsymbol{\theta}_\delta) = \mathbb{E}[-\partial \ell(\boldsymbol{\theta}_\delta) / \partial \boldsymbol{\theta}_\delta \partial \boldsymbol{\theta}_\delta^\top]$ , is the expected Fisher information matrix. Unfortunately, there is no closed form expression for the matrix  $\mathbf{K}(\boldsymbol{\theta}_\delta)$ . Nevertheless, a consistent estimator of the expected Fisher information matrix is given by  $\mathbf{J}(\widehat{\boldsymbol{\theta}}_\delta) = -\partial \ell(\boldsymbol{\theta}_\delta) / \partial \boldsymbol{\theta}_\delta \partial \boldsymbol{\theta}_\delta^\top |_{\boldsymbol{\theta}_\delta = \widehat{\boldsymbol{\theta}}_\delta}$ , which is the estimated observed Fisher information matrix. Therefore, for large  $n$ , we can replace  $\mathbf{K}(\boldsymbol{\theta}_\delta)$  by  $\mathbf{J}(\widehat{\boldsymbol{\theta}}_\delta)$ .

Let  $\theta_{\delta_r}$  be the  $r$ -th component of  $\boldsymbol{\theta}_\delta$ . The asymptotic  $100(1 - \varphi)\%$  confidence interval for  $\theta_{\delta_r}$  is given by  $\widehat{\theta}_{\delta_r} \pm z_{\varphi/2} SE(\widehat{\theta}_{\delta_r})$ ,  $r = 1, \dots, p + q + 2$ , where  $z_{\varphi/2}$  is the  $\varphi/2$  upper quantile of the standard normal distribution and  $SE(\widehat{\theta}_{\delta_r})$  is the asymptotic standard error of  $\widehat{\theta}_{\delta_r}$ . Note that  $SE(\widehat{\theta}_{\delta_r})$  is the square root of the  $r$ -th diagonal element of the matrix  $\mathbf{J}^{-1}(\widehat{\boldsymbol{\theta}}_\delta)$ .

Residuals are widely used to check the adequacy of the fitted model. To check the goodness of fit of the Bbeta model, we propose to use the randomized quantile residuals introduced by Dunn and Smyth [7]. Let  $F(x_i; \boldsymbol{\theta}_\delta)$  be the cumulative distribution function of the Bbeta distribution, as defined in (5), in which the regression structures are assumed as in (13). The randomized quantile residual is given by

$$r_i = \Phi^{-1}(F(x_i; \widehat{\boldsymbol{\theta}}_\delta)), \quad i = 1, \dots, n,$$

where  $\Phi^{-1}(\cdot)$  is the standard normal distribution function. If the assumed model for the data is well adjusted, these residuals have standard normal distribution [7].

## 6. Simulation study

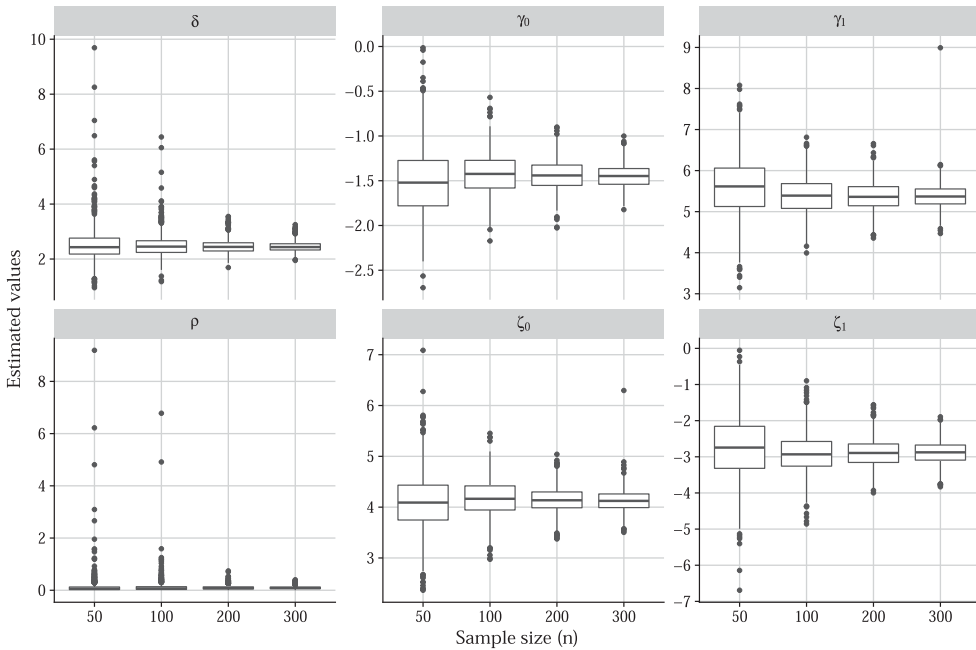
In this section, Monte Carlo simulations are performed (i) to evaluate the finite-sample behaviour of the maximum likelihood estimates of the regression coefficients and (ii) to investigate the empirical distribution of the randomized quantile residuals.

The Monte Carlo experiments were carried out by considering the following regression structure

$$\begin{aligned} \log(\alpha_i) &= \gamma_0 + \gamma_1 w_i, \\ \log(\beta_i) &= \zeta_0 + \zeta_1 z_i, \quad i = 1, \dots, n, \end{aligned}$$

**Table 3.** The estimated values for bias and mean squared error of the maximum likelihood estimators of  $\gamma_0, \gamma_1, \zeta_0, \zeta_1, \rho$  and  $\delta$ , and some values of sample size  $n$ .

$n$	The estimated bias						The estimated MSE					
	$\gamma_0$	$\gamma_1$	$\zeta_0$	$\zeta_1$	$\delta$	$\rho$	$\gamma_0$	$\gamma_1$	$\zeta_0$	$\zeta_1$	$\delta$	$\rho$
50	0.212	0.106	0.132	0.299	0.177	1.306	0.234	0.634	0.417	0.839	0.488	0.235
100	0.213	0.099	0.114	0.254	0.120	0.938	0.192	0.475	0.276	0.558	0.183	0.091
200	0.202	0.093	0.095	0.215	0.081	0.543	0.157	0.390	0.181	0.381	0.068	0.006
300	0.195	0.091	0.088	0.200	0.061	0.414	0.139	0.353	0.152	0.313	0.037	0.003



**Figure 4.** Box plots from 5000 simulated estimates of  $\gamma_0, \gamma_1, \zeta_0, \zeta_1, \rho$  and  $\delta$  for different sample sizes.

i.e.  $g_j(\cdot) = \log(\cdot), j = 1, 2$ , where the true values of the parameters were chosen to be the same with the values of the estimated parameters for the case in which we use the application part of regression, i.e.  $\gamma_0 = -1.8, \gamma_1 = 5.9, \zeta_0 = 3.8, \zeta_1 = -2.4, \rho = 0.1$  and  $\delta = 2.4$ . The covariate values of  $w_i$  and  $z_i$  were generated from the standard uniform distribution. The sample size considered was  $n = 50, 100, 200$  and  $300$ . All simulations were conducted in R [22] using the BFGS algorithm available in the `optim()` function. For each scenario, the Monte Carlo experiment was repeated 5000 times.

The Bbeta distribution is easily simulated from (5) as follows: if  $U$  has a uniform  $U(0, 1)$  distribution, the solution of the non-linear equation  $X = F^{-1}(U; \theta_\delta)$  has the  $X \sim \text{Bbeta}(\theta_\delta)$  distribution, where  $F^{-1}(\cdot; \cdot)$  is the inverse functions of  $F(\cdot; \cdot)$ . To simulate data from this non-linear equation, we can use the programming language R through `f.inv()` function [22].

In the rest of this section, a small simulation study is presented to observe the finite sample performance of the proposed estimators from a regression approach. For such

**Table 4.** Descriptive measures of the randomized quantile residuals for the bimodal beta model for different sample sizes.

$n$	Mean	StdDev	Skewness	Kurtosis
50	-0.001	0.999	0.028	2.854
100	-0.002	0.999	0.054	2.976
200	-0.003	0.997	0.077	3.002
300	-0.003	0.997	0.084	3.025

evaluation, the estimated bias and the estimated mean squared error (MSE) were calculated. The results are presented in Table 3 and Figure 4.

Table 3 presents the bias and MSE for the MLEs of  $\gamma_0, \gamma_1, \zeta_0, \zeta_1, \rho$  and  $\delta$ . Based on the results at these tables, we find that the estimates are convergent to their corresponding values of parameters. As expected, increasing the sample size  $n$  reduces substantially both bias and MSE. The previous findings are confirmed by the box plots shown in Figure 4.

### 6.1. Residuals

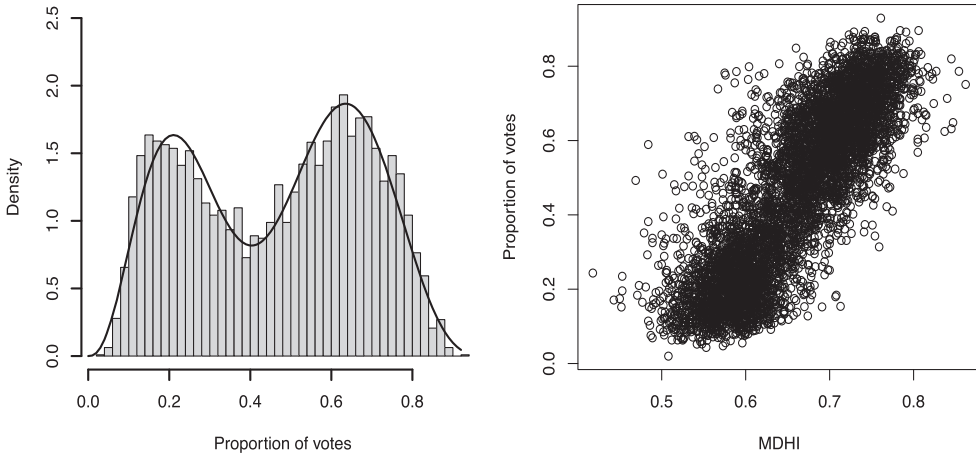
The second simulation study was performed to examine how well the distributions of the randomized quantile residuals are approximated by the standard normal distribution. The evaluation of the randomized quantile residuals were based on the normal probability plots of the mean order statistics and descriptive statistics. The results are presented in Table 4 and Figure 4 of the Supplementary Material.

In Table 4, we present the mean, standard deviation (StdDev), skewness and kurtosis of the randomized quantile residuals. For all scenarios, that is, the residuals have approximately zero mean and unit standard deviation, have skewness close to zero, and the kurtosis is near three.

## 7. Real data application

In this section, to evaluate the applicability of the proposed model, a real data set with bimodality is considered. In particular, a real-life application related to the proportion of votes that Jair Bolsonaro received in the second turn of Brazilian elections in 2018 is analysed. We compared the potentiality of the Bbeta regression with the traditional beta regression model. In order to estimate the parameters of model, we adopt the MLE method (as discussed in Section 5). The asymptotic standard errors were computed using the observed Fisher information matrix. The required numerical evaluations for data analysis were implemented using the R software [22].

The goal of this data analysis is to describe the proportion of votes that Jair Bolsonaro received in the second turn of Brazilian elections in 2018 for all 5.565 cities, and it is available at <https://dadosabertos.tse.jus.br>. The response variable  $X_i$  is the proportion of votes given the municipal human development (mhdi). The MHDI is used as explanatory variable since it is an important measure to guide authorities to assess progress and social reality as well as to define public policy priorities and comparisons of different cities [19]. Figure 5 plots the histogram with density estimated the response variable used in the application and the scatter plot of municipal human development against proportion



**Figure 5.** Estimated PDF and scatter plot of municipal human development against proportion of votes.

of votes. From Figure 5, we can see that the response variable has bimodality. Furthermore, there is evidence of a proportion of votes trend with increased municipal human development. The correlation coefficient between the proportion of votes and MHDI is 0.8290.

To explain this proportion of votes we consider the bimodal beta regression model, defined as

$$\begin{aligned}
 Y_i &\sim \text{Bbeta}(\boldsymbol{\theta}_\delta), \\
 \log(\alpha_i) &= \gamma_0 + \gamma_1 \text{mhdi}_i, \\
 \log(\beta_i) &= \zeta_0 + \zeta_1 \text{mhdi}_i,
 \end{aligned}$$

where  $i = 1, 2, \dots, 5.565$  cities and  $\text{mhdi}_i$  is municipal human development of cities  $i$ . For comparison purposes the beta regression model was fitted, assuming that

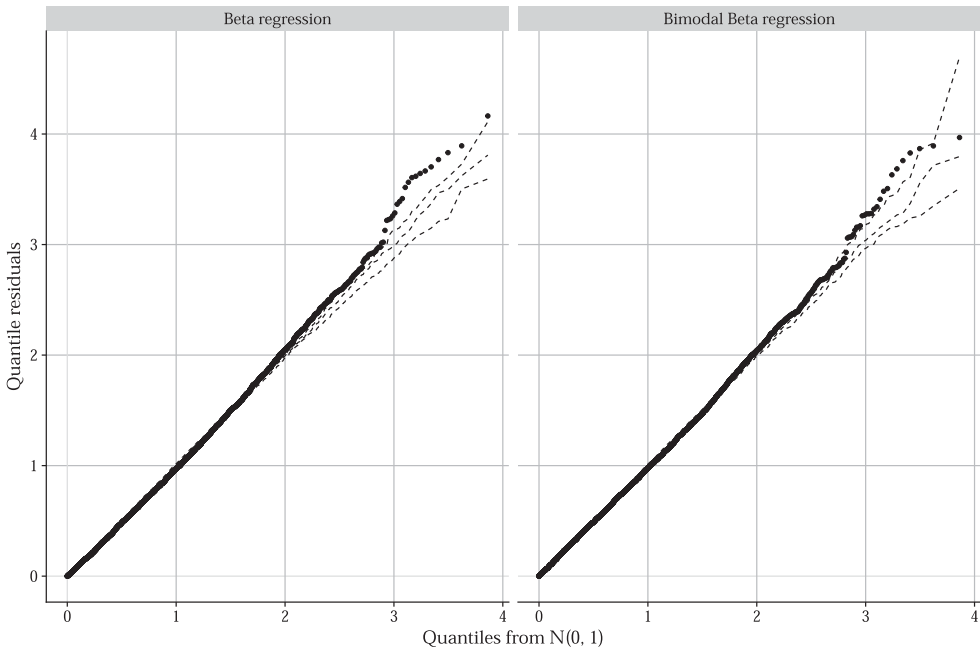
$$\begin{aligned}
 Y_i &\sim \text{beta}(\mu_i, \phi_i), \\
 \text{logit}(\mu_i) &= \beta_0 + \beta_1 \text{mhdi}_i, \\
 \log(\phi_i) &= \gamma_0 + \gamma_1 \text{mhdi}_i.
 \end{aligned}$$

and the unit-bimodal Birnbaum-Saunders (UBBS) regression model was fitted, assuming that

$$\begin{aligned}
 Y_i &\sim \text{UBBS}(\alpha_i, \beta_i, \delta), \\
 \log(\alpha_i) &= \nu_0 + \nu_1 \text{mhdi}_i, \\
 \log(\beta_i) &= \eta_0 + \eta_1 \text{mhdi}_i.
 \end{aligned}$$

Table 5 shows the estimated parameters and standard errors. Table 6 shows Akaike information criterion (AIC) and Bayesian information criterion (BIC) for the fitted models. In general, it is expected that the better model fitting the data presents the smallest values for the quantities which are AIC and BIC. Based on the AIC and BIC criteria, the model





**Figure 6.** Half-normal plot of randomized quantile residuals with simulated envelope for the fit of beta and bimodal beta.

**Table 5.** Maximum likelihood estimates and standard errors (SE) for the fit of the bimodal beta, beta and unit-bimodal Birnbaum-Saunders models in the proportion of votes.

Model	Parameter	Estimate	SE
Bbeta	$\gamma_0$	-1.8999	0.1963
	$\gamma_1$	5.9471	0.3044
	$\zeta_0$	3.8341	0.1915
	$\zeta_1$	-2.4232	0.2862
	$\rho$	0.1096	0.0090
	$\delta$	2.4092	0.0351
beta	$\beta_0$	-7.5343	0.0749
	$\beta_1$	11.1820	0.1105
	$\gamma_0$	1.0029	0.1675
UBBS	$\gamma_1$	2.5214	0.2528
	$\nu_0$	-0.5721	0.1001
	$\nu_1$	-0.1035	0.1436
	$\eta_0$	5.0120	0.0257
	$\eta_1$	-8.0601	0.0381
	$\delta$	0.6405	0.0990

which provides a better fit in this data set is the Bbeta regression model. This claim is also supported by the residual plots with simulated envelopes shown in Figure 6.

### 8. Concluding remarks

When modeling responses with bimodal bounded to the unit interval, despite its broad sense applicability in many fields, the beta distribution is not suitable. In this paper,

**Table 6.** Information criteria for the fitted models.

Models	AIC	BIC
Bbeta	−8786	−8746
beta	−8238	−8212
UBBS	−8115	−8082

the well-known two-parameter beta distribution is extended by introducing two extra parameters, thus defining the bimodal beta (Bbeta) distribution, based on a quadratic transformation technique used to generate bimodal functions [8,28], which generalizes the beta distribution. We provide a mathematical treatment of the new distribution, including bimodality, moments, entropies, stochastic representation and identifiability. We allow a regression structure for the parameters  $\alpha$  and  $\beta$ . The estimation of the model parameters is approached by maximum likelihood and its good performance has been evaluated by means of Monte Carlo simulations. Furthermore, we have proposed residuals for the proposed model and conducted a simulation study to establish their empirical properties in order to evaluate their performances. The proposed model was fitted to the proportion of votes that Jair Bolsonaro received in the second turn of Brazilian elections in 2018. As expected, the Bbeta model outperforms the beta regression in the presence of bimodality. Further, Bbeta is capable to fit well when compared with UBBS.

## Acknowledgments

The authors would like to thank the reviewers for all useful and helpful comments on an earlier version of our manuscript, which resulted in this improved version.

## Disclosure statement

No potential conflict of interest was reported by the author(s).

## ORCID

Roberto Vila  <http://orcid.org/0000-0003-1073-0114>

André F.B. Menezes  <http://orcid.org/0000-0002-3320-9834>

Mehmet N. Çankaya  <http://orcid.org/0000-0002-2933-857X>

Marcelo Bourguignon  <http://orcid.org/0000-0002-1182-5193>

## References

- [1] K.E. Ahmad and E.K. Al-Hussaini, *Remarks on the non-identifiability of mixtures of distributions*, Ann. Inst. Stat. Math. 34 (1982), pp. 543–544.
- [2] L.M. Alfaia, *A Distribuição Beta Bimodal: Propriedades e Aplicações*, UnB, Brasília, 2021.
- [3] N. Atienza, J. Garcia-Heras, and J.M. Munoz-Pichardo, *A new condition for identifiability of finite mixture distributions*, Metrika 63 (2006), pp. 215–221.
- [4] C.L. Bayes, J.L. Bazán, and G. Catalina, *A new robust regression model for proportions*, Bayesian Anal. 7 (2012), pp. 841–866.
- [5] E.R. de Alencar, *Discriminante não-linear Para Mistura De Distribuições Beta*, UnB, Brasília, 2018.

- [6] F. Domma, B.V. Popović, and S. Nadarajah, *An extension of Azzalini's method*, J. Comput. Appl. Math. 278 (2015), pp. 37–47.
- [7] P.K. Dunn and G.K. Smyth, *Randomized quantile residuals*, J. Comput. Graph. Stat. 5 (1996), pp. 236–244.
- [8] D. Elal-Olivero, *Alpha-skew-normal distribution*, Proyecciones J. Math. 29 (2010), pp. 224–240.
- [9] S. Ferrari and F. Cribari-Neto, *Beta regression for modelling rates and proportions*, J. Appl. Stat. 31 (2004), pp. 799–815.
- [10] L. Griffiths, *Introduction to the Theory of Equations*, J. Wiley, 1947.
- [11] E.D. Hahn, *Regression modelling with the tilted beta distribution: A Bayesian approach*, Can. J. Stat. 49 (2021), pp. 262–282.
- [12] Y. Ji, C. Wu, P. Liu, J. Wang, and K.R. Coombes, *Applications of beta-mixture models in bioinformatics*, Bioinformatics 21 (2005), pp. 2118–2122.
- [13] N.L. Johnson, S. Kotz, and N. Balakrishnan, *Continuous Univariate Distributions*, Vol. 2nd ed., 2, John Wiley & Sons Inc., New York, 1995.
- [14] T.I. Lin, J.C. Lee, and W.J. Hsieh, *Robust mixture models using the skew-t distribution*, Stat. Comput. 17 (2007a), pp. 81–92.
- [15] T.I. Lin, J.C. Lee, and S.Y. Yen, *Finite mixture modeling using the skew-normal distribution*, Stat. Sin. 17 (2007b), pp. 909–927.
- [16] Z. Ma and A. Leijon, *Beta mixture models and the application to image classification*. Proceedings of IEEE International Conference on Image Processing (ICIP), 2045–2048, 2009.
- [17] M. Martínez-Flórez, E. Martínez, R. Tovar-Falón, and H.W. Gómez, *A family of bimodal distributions generated by distributions with positive support*, J. Appl. Stat. 49 (2022), pp. 3614–3637.
- [18] M. Martínez-Flórez, N.M. Olmos, and O. Venegas, *Unit-bimodal Birnbaum-Saunders distribution with applications*, Commun. Stat. – Simul. Comput. (2022), pp. 1–20. <https://doi.org/10.1080/03610918.2022.2069260>
- [19] A.F.B. Menezes and W.O. Furrriel, *Beta and simplex regression models in the analysis of the municipal human development index 2010*, Rev. Bras. Biom. 37 (2019), pp. 394–408.
- [20] N.M. Olmos, G. Martínez-Flórez, and H. Bolfarine, *Bimodal Birnbaum-Saunders distribution with applications to non-negative measurements*, Commun. Stat. – Theory Methods 46 (2017), pp. 6240–6257.
- [21] R. Ospina and S.L.P. Ferrari, *Inflated beta distributions*, Stat. Pap. 51 (2008), pp. 111–126.
- [22] R Core Team, R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. URL <https://www.R-project.org/> (2021).
- [23] C.R. Rao, *Quadratic entropy and analysis of diversity*, Sankhya Ser. A. 72 (2010), pp. 70–80.
- [24] C.E. Shannon, *A mathematical theory of communication*, Bell Syst. Tech. J. 27 (1948). 379–423. 623–656.
- [25] M. Smithson, E.C. Merkle, and J. Verkuilen, *Beta regression finite mixture models of polarization and priming*, J. Educ. Behav. Stat. 36 (2011), pp. 804–831.
- [26] M. Smithson and C. Segale, *Partition priming in judgments of imprecise probabilities*, J. Stat. Theory Pract. 3 (2009), pp. 169–181.
- [27] C. Tsallis, *Possible generalization of Boltzmann-Gibbs statistics*, J. Stat. Phys. 52 (1988), pp. 479–487.
- [28] R. Vila and M.N. Çankaya, *A bimodal Weibull distribution: properties and inference*, J. Appl. Stat. 49 (2022), pp. 3044–3062.
- [29] R. Vila, L. Ferreira, H. Saulo, F. Pratavia, and E.M.M. Ortega, *A bimodal gamma distribution: properties, regression model and applications*, Statistics 54 (2020), pp. 469–493.
- [30] M.C. Wong, *Bubble Value At Risk: A Countercyclical Risk Management Approach*, John Wiley & Sons, Singapore, 2013.
- [31] J. Xue, *Loop Tiling for Parallelism*, The Springer International Series in Engineering and Computer Science, Springer, New York, 2012.