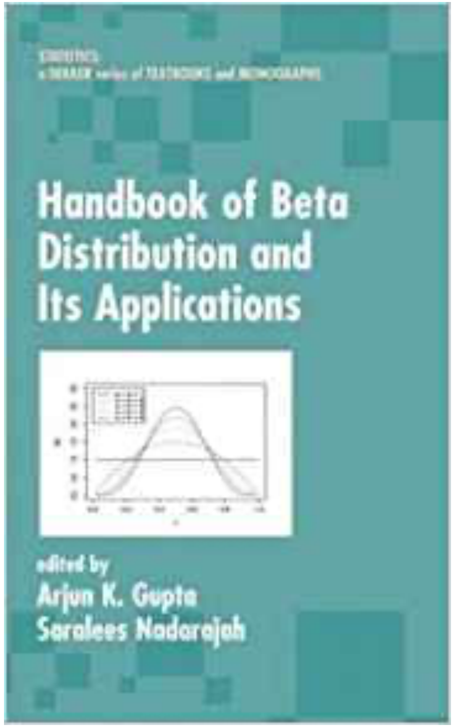
Beyond Beta: New developments and new applications for bounded response

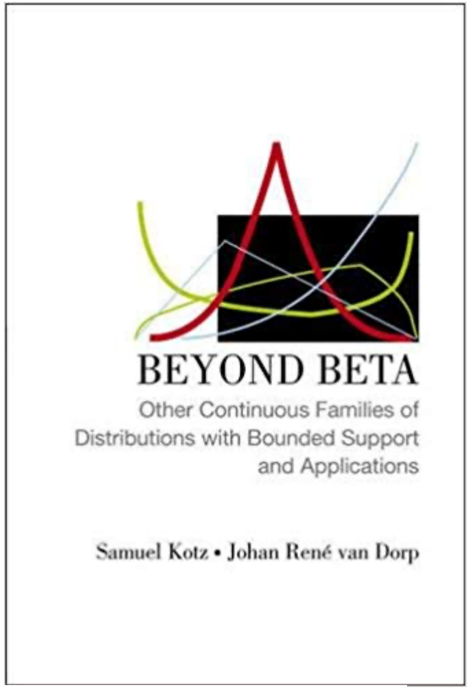
jlbazan@icmc.usp.br
https://jorgeluisbazan.weebly.com

XXII CONGRESO NACIONAL DE ESTUDIANTES DE ESTADÍSTICA-2021
Ica, UNICA, October 06, 2021

Dr. Jorge Luis Bazán
University of São Paulo
Brazil

2004





Bayesian Analysis (2012) 7, Number 4, pp. 841–866

A New Robust Regression Model for Proportions

Cristian L. Bayes*, Jorge L. Bazán† and Catalina García‡

1. Beta Regression Models
2. My experience
3. A framework to the formulation of bounded distributions
4. Application.
5. Extensions and applications
6. Final comments

1

Beta Regression Models

Beta regression model were introduced for several authors as

- Paolino, P. (2001). “Maximum likelihood estimation of models with beta-distributed dependent variables”. *Political Analysis*, 9: 325-346.
- Buckley, J. (2002). “Estimation of models with beta-distributed dependent variables: A replication and extension of Paolino (2001)”. *Political Analysis*, 11: 1–12.
- Kieschnick, R. and McCullough, B. D. (2003). “Regression analysis of variates observed on (0,1): percentages, proportions, and fractions.” *Statistical Modeling*, 3: 193–213.
- Ferrari, S. and Cribari-Neto, F. (2004). “Beta regression for modelling rates and proportions.” *Journal of Applied Statistics*, 31: 799–815.
- Smithson, M. and Verkuilen, J. (2006). “A Better Lemon Squeezer? Maximum- Likelihood Regression With Beta-Distributed Dependent Variables.” *Psychological Methods*, 11(1): 54–71.
- Branscum, A. J., Johnson, W. O., and Thurmond, M. C. (2007). “Bayesian Beta regression; application to household data and genetic distance between foot-and-mouth disease viruses.” *Australian & New Zealand Journal of Statistics*, 49(3): 287–301.

Beta density:

$$f(y;\mu,\phi) = \frac{\Gamma(\phi)}{\Gamma(\mu\phi)\Gamma((1-\mu)\phi)} y^{\mu\phi-1}(1-y)^{(1-\mu)\phi-1}, \quad 0 < y < 1,$$

where $0 < \mu < 1$ and $\phi > 0$. Note that

$$\mathrm{E}(y) = \mu$$

and

$$\mathrm{var}(y) = \frac{\mu(1-\mu)}{1+\phi}.$$

Hence, ϕ can be regarded as a precision parameter.

This is not the usual parameterization of the beta law, but is convenient for modeling purposes.

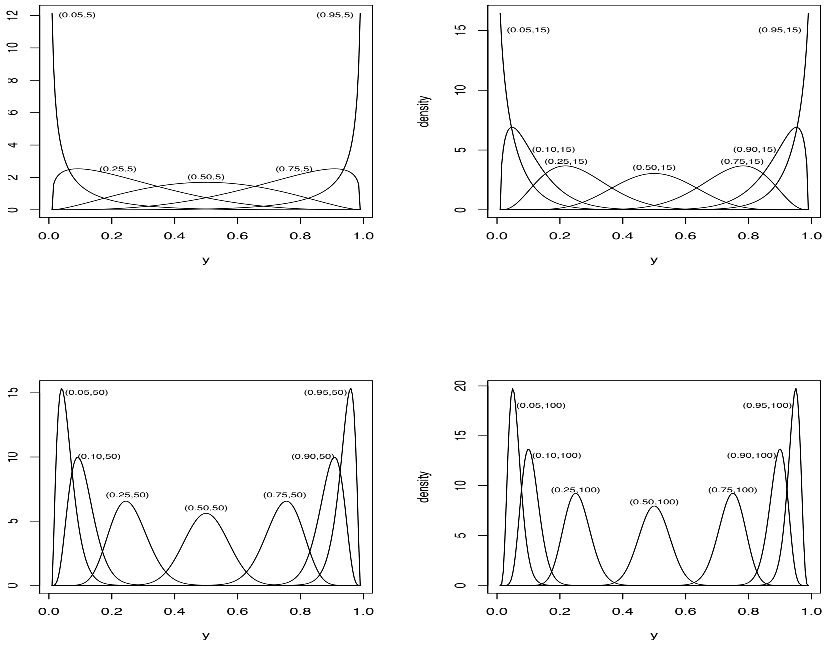


Figure 1. Beta densities for different combinations of (μ, ϕ).

Let $\mathbf{y} = (y_1, \ldots, y_n)^T$ be a vector of observed responses that takes values in (0,1). The doubly mixed beta regression model is given by:

$$y_{ij} \sim Beta(\mu_{ij}, \phi_{ij}).$$
$$g_1(\mu_{ij}) = \mathbf{x_{ij}}^T\beta + \mathbf{z_{ij}}^T\mathbf{b}_i, \quad g_2(\phi_{ij}) = -\mathbf{w}_{ij}^T\delta - \mathbf{h}_{ij}^T\mathbf{d}_i$$
$$\mathbf{b}_i \sim N_p(\mathbf{0}, \mathbf{\Sigma}_b) \quad \text{and} \quad \mathbf{d}_i \sim N_r(\mathbf{0}, \mathbf{\Sigma}_d)$$
$$j = 1, \ldots, n_i \quad i = 1, \ldots, n$$

where $\beta = (\beta_1, \ldots, \beta_k)^T$ is a vector of fixed effect regression coefficients associated with the location parameter and $\delta = (\delta_1, \ldots, \delta_l)^T$ is a vector of fixed effect regression coefficients associated with the shape parameter.

The random effects of the location and shape parameters are denoted, respectively, by $\mathbf{b}_i = (b_{i1}, \ldots, b_{ip})^T$ and $\mathbf{d}_i = (d_{i1}, \ldots, d_{ir})^T$

Since the dispersion decreases when the value of ϕ increases, we take the negative sign in the linear predictor to ease the interpretation of the coefficients.

Moreover, $\mathbf{x}_{ij} = (x_{ij1}, \ldots, x_{ijk})^T$, $\mathbf{w}_{ij} = (w_{ij1}, \ldots, w_{ijl})^T$, $\mathbf{z}_{ij} = (z_{ij1}, \ldots, z_{ijp})^T$ and

$\mathbf{h}_{ij} = (h_{ij1}, \ldots, h_{ijr})^T$

are covariate vectors, which do not need to be identical and they can be overlapping.

Examples of link functions $g_1(\cdot)$ and $g_2(\cdot)$ are he logistic and logarithm functions, respectively.
It is a consolidated model

Verkuilen J, Smithson M. Mixed and Mixture Regression Models for Continuous Bounded Responses Using the Beta Distribution. *Journal of Educational and Behavioral Statistics*. 2012;37(1):82-113

It is a consolidated model. Some additional references

- Verkuilen J, Smithson M. Mixed and Mixture Regression Models for Continuous Bounded Responses Using the Beta Distribution. *Journal of Educational and Behavioral Statistics*. 2012;37(1):82-113

- Wang J, Luo S. Bayesian multivariate augmented Beta rectangular regression models for patient-reported outcomes and survival data. *Statistical Methods in Medical Research*. 2017;26(4):1684-1699

- Haiming Zhou, Xianzheng Huang, Bayesian beta regression for bounded responses with unknown supports, *Computational Statistics & Data Analysis*,Volume 167, 2022

2

My experience

- **1. The beta regression model is not a robust model. It is sensible for outliers (Bayes, et al, 2012; Lemonte an Bazán 2016).**

- Bayes, C., Bazán, J. L. García, C. (2012). A new robust regression model for proportions. *Bayesian Analysis*. 7(2), 771-796.

(Use of the beta rectangular distribution. Bayesian estimation.)

- Lemonte A.; B ; Bazán, J. L. (2016). New class of Johnson SB distributions and its associated regression model for rates and proportions. *Biometrical Journal* (1977), v. 58, p. 727-746.

(*Generalizations of the SB-U Johnson distribution based in the normal distribution using the class of the symmetrical distributions. ML estimation.*)

- **2. Zero-One Augmented versions of the Beta and Beta rectangular model can be interesting models for data with excesses of zeros and/ or ones (Nagarotto, et al, 2019; da Silva, et al, 2020)**

- Nogarotto, D. A.; Azevedo, C. L. N. ; Bazán, J. L.(2020). Bayesian Estimation, Residual Analysis and Prior Sensitivity Study for Zero-One Augmented Beta Regression Models with an Application to Psychometric Data. *Brazilian Journal of Probability and Statistics.* 34(2), 304-322.

(*Bayesian estimation using Jeffrey priors*)

- Silva, Ana R. S. ; Azevedo, C. L. N. ; Bazán, J. L. ; Nobre, S. J. (2021) Augmented-limited regression models with an application to the study of the risk perceived using continuous scales, Journal of Applied Statistics, 48:11, 1998-2021(*Bayesian estimation using a reparameterized version of the Beta rectangular*)

- **3. Sometimes we want model no only the mean as a location parameter but also a quantile of the distribution of the response variable (Lemonte and Bazan, 2016; Bayes et al 2019; da Paz et al, 2019, de Oliveira et al, 2019).**

- Lemonte A.; B ; Bazán, J. L. (2016). New class of Johnson SB distributions and its associated regression model for rates and proportions. *Biometrical Journal* 58, p. 727-746.

(*Median regression model using the symmetric SB distributions. ML estimation*)

- Bayes, C., Andrade, M. C. ; Bazán, J. L. (2017). A quantile parametric mixed regression model for bounded response variables. *Statistics and its Interface*, 10, p. 483-493.

(*Mixed quantile regression model using the Kumaraswamy distribution. Bayesian estimation*).

- Da Paz, R. F.; Balakrishnan, J. L. ; Bazán, J. L. (2019). L-Logistic Regression Models: Prior sensitivity analysis, robustness to outliers and applications. *Brazilian Journal of Probability and Statistics* , v. 33, p. 455-479.

(*Median regression model using the L-Logistic distribution. ML estimation*)

- De Oliveira, E. S. B.; de Castro, M.; Bayes, C.; Bazán, J. L. (2021). Bayesian parametric quantile models for heavy tailed bounded. Submitted for *Journal of Applied Statistics*. In Revision.

(*Mixed quantile regression model using the Gompertz limited distribution. Bayesian estimation*).

- **4. Mixture of Beta distributions no solve complex distributions with multiples modes (da Paz et al, 2017 ; da Paz et al, 2019).**

- Da Paz, R. F.; Bazán, J. L.; Milan, L. A. (2017). Bayesian estimation for a mixture of simplex distributions with an unknown number of components: HDI analysis in Brazil. *Journal of Applied Statistics*, v. 44, p. 1630-1643, 2017.

(*Use of the mixture of Simplex distributions. Bayesian estimation*.)

- Da Paz, R. F.; Bazán, J. L.; Lachos, V. H.; Dey. D (2020). A finite Mixture Mixed Proportion Regression Model for Classification Problems in Longitudinal Voting data. Submitted for publication in Journal of Applied Statistics. In revision.

(*Mixed quantile regression models using mixture of L-Logistic distribution. Bayesian estimation*.)

- **5. A extra parameter can give more flexibility to the distribution of the response (Cancho et al, 2020, Lemonte and Bazan, 2016; Rodrigues et al, 2020).**

- Lemonte A.; B ; Bazán, J. L. (2016). New class of Johnson SB distributions and its associated regression model for rates and proportions. Biometrical Journal 58, p. 727-746.

(*Median regression model using the symmetric SB distributions. ML estimation*)

- Cancho, V. G.; Bazán, J. L. ; Dey, D. K. (2020). A new class of regression model for a bounded response with application in the study of the incidence rate of colorectal cancer. *Statistical Methods in Medical Research*. 29(7), 2015-2033.

(*Median regression model using the Bounded Power Normal distribution. Bayesian estimation*).

- Rodrigues, J. ; Bazán, J. L.; Suzuki, A. K. (2020). A flexible procedure for formulating probability distributions on the unit interval with applications. *Communications in Statistics-Theory and Methods. 49(3),* 738- 754.

(*Median regression model using the generalized SB distributions. Classical and Bayesian estimation*).

- Piccirilli, G. P. ; De Bastiani, F. ; Bazán, J. L.; (2021). Bounded mixed regression models of Brazil's longitudinal mortality rate from bronchial and lung cancer. *International Statistical Review. Submitted.*

(*Regression model using distributions SB type. Classical estimation*).

3

A FRAMEWORK TO THE FORMULATION OF BOUNDED DISTRIBUTIONS

2. A composite quantile family of the cumulative distribution functions (cdf)

Let W be a random variable with cdf $G(w;\theta)$ and probability density function (pdf) $g(w;\theta)$ on the support R_W, where θ is an unknown parameter (or vector of parameters). Let X be another continuous random variable with cdf $F(x;\phi)$ and *quantile function*

$$Q_F(y;\phi) = F^{-1}(y;\phi), 0 \leq y \leq 1 \tag{1}$$

where $R_X = R_W$ and ϕ an unknown (known) parameter.

2.1. Definition and properties of GF-quantile distributions

In order to obtain a new flexible class of distributions on the (0, 1) interval, called *GF-quantile distribution*, we define the composite probability distribution function

$$GF(y;\theta,\phi) \doteq G^{\circ}Q_F(y;\theta,\phi) = G(Q_F(y;\phi);\theta) = \int_{-\infty}^{Q_F(y;\phi)} g(w;\theta)dw, y \in (0,1) \tag{2}$$

It can be clearly seen that $GF(y;\theta,\phi)$ is the cdf of the continuous random variable Y on the (0, 1) interval with parameters θ and ϕ. In fact, we generate by composition of the baseline distribution $G(w;\theta)$ with the quantile function $Q_F(y;\phi), 0 \leq y \leq 1$. The composite probability distribution in (2) will be denoted by $Y \sim GF(\theta,\phi)$.

The pdf of the composite quantile family of distributions in (2) can be obtained as

$$gf(y;\theta;\phi) = g(Q_F(y;\phi),\theta)q_F(y;\phi), y \in (0,1) \tag{4}$$

where $q_F(y;\phi)$ is the *quantile density function* (qdf) of X (Parzen 1979) defined by $q_F(y;\phi) = \frac{dQ_F(y;\phi)}{dy}$.

Next, we present some properties of the *GF*-quantile probability distribution and how to generate the data. The quantile function of the *GF*-quantile distribution is given by

$$Q_{GF}(y;\theta,\phi) = F\left(G^{-1}(y;\theta);\phi\right) = F(Q_G(y;\theta);\phi) = \int_{-\infty}^{Q_G(y;\theta)} f(x;\phi)dx, y \in (0,1) \tag{5}$$

If a simulation study of the *GF*-quantile distribution is needed the following procedure motivated by (5) could be useful. First of all, the parameter values θ and ϕ must be fixed and the y-values can be generated following these steps.

1. Generate u values from the (0, 1) uniform distribution,
2. Take $w = Q_G(u;\theta)$, where w are generated values of the distribution $G(w;\theta)$,
3. Take $y = F(w;\phi)$ which are the values generated from the *GF*-distribution.

Table 1. *GF*-quantile probability distributions with different transformed distributions on the bounded supports.

Type of Support	Baseline distribution	Transformed distribution	Quantile function	Quantile density function	pdf on the unit interval
$R_X = R_W$	$W \sim G(w;\theta)$	$X \sim F(x;\phi)$	$Q_F(u;\phi)$	$q_F(u;\phi)$	$gf(y;\theta,\phi)$
(0, 1)	$g(w;\theta)$	$U(0,\ 1)$	u	1	$g(y;\theta)$
$(a,\ b)$	$g(w;\theta)$	$U(a,\ b)$	$(b-a)u+a$	(b-a)	$g((b-a)y+a;\theta)(b-a)$
$(0,\infty)$	$g(w;\theta)$	Exp (1)	$-\ln(1-u)$	$\frac{1}{1-u}$	$g(-\ln(1-y);\theta)\frac{1}{1-y}$
$(-\infty,0)$	$g(-w;\theta)$	RefExp(1)	$\ln(u)$	$\frac{1}{u}$	$g(-\ln(y);\theta)\frac{1}{y}$
$(0,\infty)$	$g(w;\theta)$	$\mathrm{W}(1,1/\phi)$	$[-\ln(1-u)]^{\phi}$	$\frac{\phi[-\ln(1-u)]^{\phi-1}}{1-u}$	$g([-\ln(1-y)]^{\phi};\theta)\frac{\phi[-\ln(1-y)]^{\phi-1}}{1-y}$
$(-\infty,0)$	$g(-w;\theta)$	$\mathrm{RefW}(1,\phi)$	$-[-\ln(u)]^{\phi}$	$\frac{\phi[-\ln(u)]^{\phi-1}}{u}$	$g([-\ln(y)]^{\phi};\theta)\frac{\phi[-\ln(y)]^{\phi-1}}{y}$
$(-\infty,\infty)$	$g(w;\theta)$	Logistic	$\ln(\frac{u}{1-u})$	$\frac{1}{u(1-u)}$	$g(\ln(\frac{y}{1-y});\theta)\frac{1}{y(1-y)}$
any	$f(x;\phi)$	$f(x;\phi)$	$Q_F(u;\phi)$	$q(u;\phi)$	1

3.4.1. The Logistic-Normal distributions

If the logistic quantile function of the transformed standard Logistic distribution $X \sim F(x; 0, 1)$ is applied to the baseline Normal distribution with mean μ and variance σ^2, we obtain the well-known Logistic-Normal distribution (Aitchison and Shen 1980) which is called as Johnson-S_B (Johnson 1949). This probability distribution has been used in many applications and its properties and generalizations on the simplex can be seen in Aitchison and Shen (1980). The pdf is given by

$$gf(y; \theta) = \frac{1}{\sqrt{2\pi\sigma^2} y(1 - y)} e^{-\frac{\left(\log\left(\frac{y}{1-y}\right) - \mu\right)^2}{2\sigma^2}}, y \in (0, 1), \theta = \left(\mu, \sigma^2\right) \tag{26}$$

3.4.2. GF real type distributions

The quantile approach can be used to generalize the example discussed before in the following way: Let X be any random variable with support $(-\infty, \infty)$ and the quantile function given by

$$Q_X(u;\phi) = \lambda + \eta S(u), u \in [0,1], \phi = (\lambda, \eta) \tag{27}$$

where $S(u)$ is called the basic quantile function (Gilchrist 2000). Some examples of this basic quantile function are shown in Table 1. Let $W \sim g(w;\theta)$ be the baseline distribution on the $(-\infty, \infty)$ interval. From (2), we have a general GF-quantile distribution with pdf given by

$$gf(u;\theta,\phi) = \eta g(\lambda + \eta S(u);\theta)|\frac{dS(u)}{du}|, u \in (0,1) \tag{28}$$

We call the probability distribution in (28) as GF real type distribution. It has a general representation and not restricted to the baseline normal distributions and the standard logistic quantile function. The quantile function of the GF-distribution is given by

$$Q_{gf}(u;\phi) = S^{-1}\left(\frac{w_u - \lambda}{\eta}\right) \tag{29}$$

where $w_u = Q_G(u)$.

3.4.3. Johnson-S_B type distributions

Let G be the baseline distribution of the random variable W with support $(-\infty, \infty)$ and $S(u) = \ln\left(\frac{u}{1-u}\right)$ the standard quantile function of the transformed logistic random variable X. From (28), the logit-Johnson-S_B distribution (denoted by $JSB(\theta, \phi)$) is given by

$$gf(y;\theta,\phi) = \frac{\eta g\left(\lambda + \eta S(y);\theta\right)}{y(1-y)}, y \in (0,1), \phi = (\lambda, \eta) \tag{30}$$

More details in

- Rodrigues, J. ; Bazán, J. L.; Suzuki, A. K. (2020). A flexible procedure for formulating pr distributions on the unit interval with applications. *Communications in Statistics-Theory and M* *49(3),* 738- 754.

(*Diferentes procedure to create different variable in the (0,1) interval*).

4

APPLICATION

To illustration we formulate a new regression model based on $JSB(\theta, (0,1))$ responses in equation (30) and the baseline distribution $G(y, \theta)$ with support in the $\Re$. That is we consider a the Standard Logistic distribution as Transformed distribution.

The baseline two-parameter distributions considered were: *Gumbel* ($Gu(\mu, \sigma)$), *Logistic* ($Lo(\mu, \sigma)$), *Normal* ($N(\mu, \sigma)$) and the *Reverse Gumbel* ($RGu(\mu, \sigma)$).

The three-parameter continuous distributions considered were: *Exponential Gaussian* ($ExGauss(\mu; \sigma; \nu)$), *Power Exponential* ($PE(\mu; \sigma; \nu)$), *t family* ($TF(\mu; \sigma; \nu)$), *Skew Normal* ($SN1(\mu; \sigma; \nu)$ and $SN2(\mu; \sigma; \nu)$), for $\nu > 0$

We use GAMLSS parameterization for this distributions. See

https://www.gamlss.com/distributions/

we consider the D-poverty dataset (Y: proportion of poverty) and the human development index (HDI) as the covariate Z of 195 provinces of Peru.

Given the paired dataset $(Z_1, y_1), \ldots, (Z_{195}, y_{195})$, we consider the $JSB(\theta, (0,1))$ regression models with

$\theta = (\mu, \sigma)$ and $\theta = (\mu, \sigma, \nu)$ given by

$$Y_i \sim JSB(\theta, (0,1)), i = 1, ..., 195 \tag{35}$$

with the following link functions given in Table 4.

$$\begin{aligned} \mu_{link}(\mu_i) &= \beta_0 + \beta_1 Z_i, i = 1, 2, ..., 195, \\ \sigma_{link}(\sigma) &= \delta_0, \\ \nu_{link}(\nu) &= \phi_0 \end{aligned}$$

- Model comparison criteria indicated that the best model is the *Gumbel-Logit* regression model

Table 4. Johnson-S_B regression model for the poverty dataset.

Models	Parameters	μ_{link}	σ_{link}	ν_{link}	Global Deviance:	AIC:	SBC:
Beta	3	logit	logit	–	−308.48	−302.48	−292.66
Lo	3	identity	log	–	−324.41	−318.41	−308.59
Gu	3	identity	log	–	−348.13	−342.13	−332.31
RGu	3	identity	log	–	−220.59	−214.59	−204.78
ExGauss	4	identity	log	log	−311.55	−303.55	−290.46
PE	4	identity	log	log	−325.24	−317.24	−304.14
SN1	4	identity	log	log	−312.05	−304.05	−290.95
SN2	4	identity	log	log	−341.94	−333.94	−320.85
T	4	identity	log	log	−326.04	−318.04	−304.95

Table 5. l-Gumbel Regression model versus Beta regression model for the poverty dataset.

Model	Coefficients	Estimate	Std. Error	t value	$Pr(>\|t\|)$
Beta	β_0	9.84	0.53	18.68	$< 2e-16$
	β_1	−17.08	0.91	−18.77	$< 2e-16$
	δ_0	−1.17	0.06	−19.11	$< 2e-16$
GU	β_0	11.69	0.42	27.78	$< 2e-16$
	β_1	−19.88	0.72	−27.45	$< 2e-16$
	δ_0	−0.88	0.06	−15.66	$< 2e-16$

More details in Section 4.2 Regression models: Likelihood approach from

- Rodrigues, J. ; Bazán, J. L.; Suzuki, A. K. (2020). A flexible procedure for formulating pr distributions on the unit interval with applications. *Communications in Statistics-Theory and M* *49(3),* 738- 754.

(*Diferentes procedure to create different variable in the (0,1) interval*)

- Piccirilli, G. P. ; De Bastiani, F. ; Bazán, J. L.; (2021). Bounded mixed regression models of longitudinal mortality rate from bronchial and lung cancer. *International Statistical Review. Submitte*

(*Regression model using distributions SB type. Classical estimation*).

5

Extensions and applications

Political analysis. Votes are proportions!

- da Paz, R. F. ; Ehlers, R. S. ; Bazán, J. L. . *A Weibull Mixture Model for the Votes of a Brazilian Political Party*. Springer Proceedings in Mathematics & Statistics. 1 ed. : Springer International Publishing, 2015, v. 118, p. 229-241.

- Bazán, J. L. ; Sulmont, D. ; Calderón, A. B (2014). Las Organizaciones políticas en las elecciones presidenciales peruanas de 2011 usando análisis de componentes principales. *Revista de Estudos Sociais*. UFMT 14(1), 10-27.

- Da Paz, R., Bazán, J. L.; Lachos, V. H.; Dey. D (2020). A finite Mixture Mixed Proportion Regression Model for Classification Problems in Longitudinal Voting data. Submitted for publication in *Journal of Applied Statistics*.

Educational analysis. Response are bounded!

- Flores, S.; Prates, M. O. ; Bazán, J.L ; Bolfarine, H. (2021). Spatial regression models for bounded response variables with evaluation of the degree of dependence. *Statistics and Its Interface*. v. 14, p. 95-107

- De Oliveira, E. S. B.; Wang, X.; Bazán, J. L. (2020). Bayesian Cognitive Diagnosis Model for Bounded Responses. Submitted for *Statistics and Its Interface*.

Psychometric analysis. Times and Response are bounded!

- Nogarotto, D. A.; Azevedo, C. L. N. ; Bazán, J. L.(2020). Bayesian Estimation, Residual Analysis and Prior Sensitivity Study for Zero-One Augmented Beta Regression Models with an Application to Psychometric Data. *Brazilian Journal of Probability and Statistics.* 34(2), 304-322.

- Flores, S.; Bazán, J. L.; Bolfarine, H. (2020). A Hierarchical Joint Model for Bounded Response Time and Response Accuracy. In: Wiberg M., Molenaar D., González J., Böckenholt U., Kim JS. (eds) Quantitative Psychology. IMPS 2019. Springer Proceedings in Mathematics & Statistics, vol 322. Springer, Cham. 95-109

- Silva, Ana R. S. ; Azevedo, C. L. N. ; Bazán, J. L. ; Nobre, S. J. (2021) Augmented-limited regression models with an application to the study of the risk perceived using continuous scales. *Journal of Applied Statistics*, 48:11, 1998-2021

Psychometric analysis. Times and Response are bounded!

- Silva, Ana R. S. ; Azevedo, C. L. N. ; Bazán, J. L. ; Nobre, S. J. (2021) Bayesian inference for zero-and/or-one augmented rectangular beta regression models. Accepted. *Brazilian Journal of Probability and Statistics*.

- Molenaar, D. ; Curi, M. ; Bazán, J. L. ; Nobre, S. J. (2021) Item Response Theory Models for Bounded Continuous Data. Submitted *Journal of Educational and Behavioral Statistics*

- **Health analysis. Incidences rates are proportions**

- Cancho, V. G.; Bazán, J. L. ; Dey, D. K. (2020). A new class of regression model for a bounded response with application in the study of the incidence rate of colorectal cancer. *Statistical Methods in Medical Research*. *49(3),* 738- 754*.*

- Piccirilli, G. P. ; De Bastiani, F. ; Bazán, J. L.; (2021). Bounded mixed regression models of Brazil's longitudinal mortality rate from bronchial and lung cancer. *International Statistical Review. Submitted.*

(*Regression model using distributions SB type. Classical estimation*).

- **Economy, Finances, Business. Index are proportions**

6

COMMENTS

- Beyond Beta: The Beta in the “(0,1) world” is as the Normal in the “line world.” It is the Queen. But it not cover the most part of the problems in this world. We need new distributions as claimed by Kotz and van Dorp (2014).

- However, it is not about of create new distributions, it is about of explain new mechanism to explain the randomization in the interval (0,1). Then is it open and contribution are welcomed. We need study mathematical and probabilistic properties of the new distributions formulated in (0,1). (Rodrigues et al, 2020).

- The majority of propose emphasize in a quantile parameterization considering a median as a local parameter and additional parameters as shape parameters. It is more convenient than the mean-dispersion parameterization of the Beta distribution because in the (0,1) we characterize better the distribution using a robust approach.

- We showed different applications which exhibited that the original data are bounded. In special in Humanities, Psychology, Education, Social sciences, Political Analysis and Health.

- Proportions, index, time and some continuous response are often limited and it was ignored in the modeling. The different applications show that it is possible modeling this response without transform data.

- We need development new Exploratory Data Analysis and Data Visualization techniques for bounded data.

- We need multivariate version of the distributions proposed using different correlational structure.

- We require censured, truncated, mixtures and inflated or augmented versions of the distributions studied.

- New Bounded-Binomial distributions can be formulated to Count data.

- Model comparison criteria, checking model methods, diagnostic and residual analysis must be reexamined for bound response and eventually new development bust be proposed.

- We need use this distributions in problems of Times Series, Dynamic and sequential problems, Spatial and Temporal data, Pooled data, Survivor Analysis and Conjoint Analysis, Funcional data, Latent variables models, (Factor Analysis, SEM, Item Response Theory, Cognitive Diagnostic models), etc.

- In a Bayesian approach, to models with parameters in the interval (0,1) new class of priors can be formulated considering the new distributions alternative to the Beta distribution.

- Regression models using Beta and Simplex are implemented in R packages using ML estimation.

https://cran.r-project.org/web/packages/betareg/betareg.pdf
https://cran.r-project.org/web/packages/simplexreg/simplexreg.pdf

- Regression models using Beta, L-Logistic and Kumaraswamy regression are implemented in R packages using Bayesian estimation.

https://cran.r-project.org/web/packages/llogistic/llogistic.pdf
https://www.r-inla.org

- Beta is a consolidate model and it is available in different commercial software as SPSS, Stata and SAS

- Other generic R packages as GAMLSS could implement other distributions as showed in the Rodrigues et al (2020)

- Any model listed here could be implemented using bayesian approach in WinBUGS, OpenBugs, JAGS, Stan and then could be fitted using interface with R

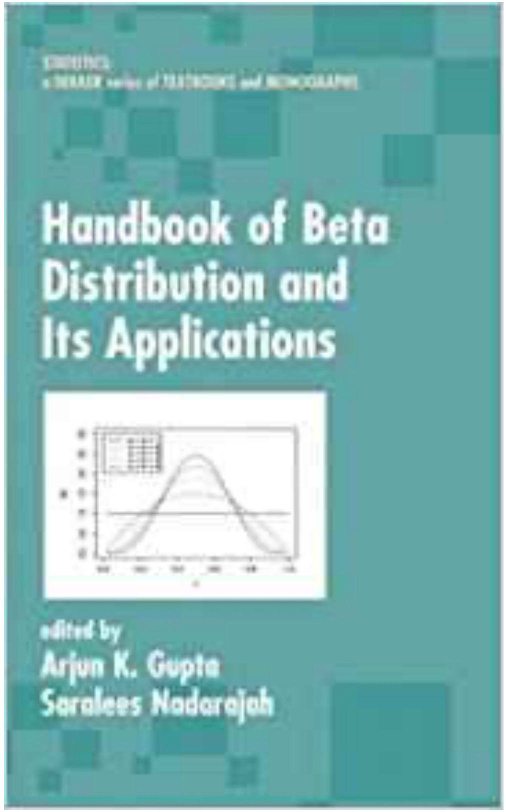
- Using PyStan and PyMC3 we can implement any model listed here under Bayesian approach in Python.

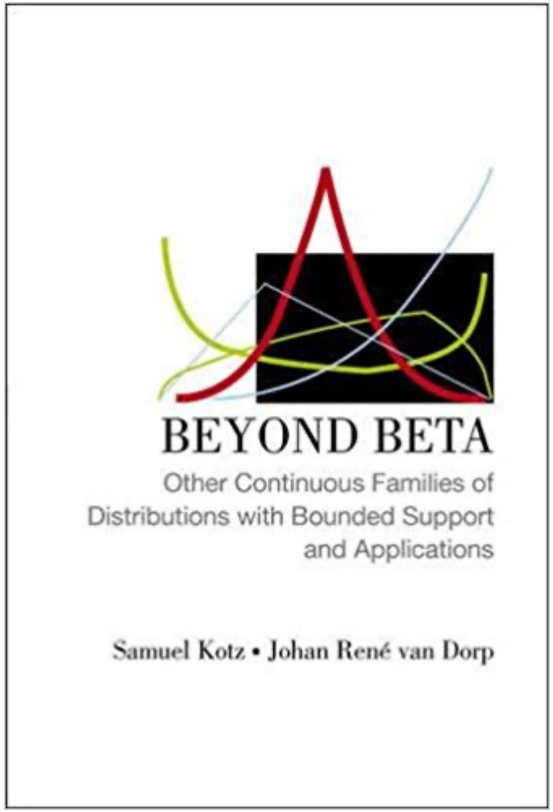
- We need a General Bounded package following https://cran.r-project.org/web/packages/quantreg/quantreg.pdf

- Use combined of Statistical techniques with Data Science and Machine Learning using new algorithm will be developed to complex and big data.

17 years later.

New versions?





Let’s go to work.

Thank you for your attention!

!TODOS ESTÁN INVITADOS A PRODUCIR NUEVOS RESULTADOS!

jlbazan@icmc.usp.br
https://jorgeluisbazan.weebly.com