A BAYESIAN APPROACH FOR MODELING INTERVAL-VALUED VARIABLES

Fidel Castro MORALES[1]
Eufrásio de Andrade LIMA NETO[2]

▪ *ABSTRACT: This paper proposes two Bayesian approaches to estimate the regression model coefficients considering interval-valued variables as response and explanatory variables. The first approach considers a more simple co-variance structure, while the second approach supposes a more general co-variance structure. The posterior distribution for the parameters was approximated considering Markov Chain Monte Carlo method (MCMC). A simulation study is presented and suggests the effectiveness of the sampling scheme in recovering the true values of the parameters and also indicates convergence of the parameter estimate algorithm. The new approaches are applied to real interval-valued data sets and their performance compared.*

▪ *KEYWORDS: Interval variables; regression; MCMC; Bayesian approach.*

1 Introduction

The presence of intervals data sets is becoming very common in data analysis problems. This new type of data represents the uncertainty present in an error measurement or the natural phenomenon variability present in the data. Another source of interval data sets derives from the aggregation of huge databases into a reduced number of groups which are described by symbolic data analysis (SDA), a domain related to multivariate analysis, pattern recognition and artificial intelligence (Bock and Diday, 2000; Billard and Diday, 2006). Interval data appear in some fields like engineering, economy, medicine, among other. Furthermore, technical specification, temperatures in meteorological stations, daily stock prices

[1]Federal University of Rio Grande do Norte – UFRN, Department of Statistics, Natal, RN, Brazil. E-mail: *fidel@ccet.ufrn.br*
[2]Federal University of Paraíba – UFPB, Department of Statistics, João Pessoa, PB, Brazil. E-mail: *eufrasio@de.ufpb.br*

are examples of possible interval-valued variables. Therefore, statistical tools to analyze interval-valued data are very much required. The use of Bayesian inference in linear regression models is widely studied (Gamerman, 1997; Gelman, 2006). The Bayesian techniques allow the use of the posterior function making possible to perform point estimation and inference over the parameters, differently from the frequentist approach where the distribution and variance of the estimators are not known.

Several regression methods to model interval data has been proposed in the literature. Some of these methods study the problem by an optimization point of view (Billard and Diday, 2000; Lima Neto and De Carvalho, 2008; Maia and De Carvalho, 2008; Lima Neto and De Carvalho, 2010). Other methods consider the principles of set arithmetic algebra to obtain a linear regression models for interval data (Gil *et al.*, 2007; Blanco-Fernandez *et al.*, 2011). Recently, some regression methods for interval data (Queiroz *et al.*, 2011; Brito and Silva, 2012; Fagundes *et al.*, 2013) take into account a probabilistic background for the response interval-valued variable Y. Lima Neto *et al.* (2011) presented an approach based on generalized linear models theory, called bivariate symbolic regression method (BSRM). They consider the interval response variable $Y = [Y_L, Y_U]$ as a bivariate random vector with joint distribution belongs to bivariate exponential family. Moreover, the BSRM method allows the use of inference techniques over the parameters estimate, goodness-of-fit tests, residual and diagnostic analysis.

In this paper, we propose a Bayesian approach to estimate the parameters for a regression model for interval-valued variables by considering equal and different variances. At this time, all proposed regression approaches for interval variables are based on frequentist inference point of view. The article is organized as follows: Section 2 presents the Bayesian approach and the estimation process. Section 3 takes a simulation study based on Monte Carlo Markov Chain (MCMC) framework and presents an application to real interval-valued data sets. Finally, in Section 4, we conclude the paper with some remarks.

2 Bayesian regression models for interval-valued variables

In recent years, some non-Bayesian approaches have been proposed to fit a linear regression model considering interval-valued variables in the response and explanatory variables (Billard and Diday, 2000; Coppi *et al.*, 2006; Gil *et al.*, 2007; Lima Neto and De Carvalho, 2008; Lima Neto and De Carvalho, 2010; Ferraro *et al.*, 2010; Xu, 2010; Blanco-Fernandez *et al.*, 2011). In this section we present a Bayesian approach for linear regression methods based on the bivariate Gaussian distribution. Both methods taking into consideration the midpoint and the range of intervals. The first approach considers a more simple co-variance structure, while the second approach supposes a more general co-variance structure. Methods based on the midpoint and range of intervals have been widely discussed in the literature. D'Urso and Gastaldi (2000) suggested that the dependence between the center and the range is often encountered in real world applications. Particularly, in the SDA field, Lauro and Palumbo (2000) discusses some principal component approaches based on midpoint of intervals and Maia and De Carvalho (2011) proposes a time series model for interval-valued data taking into account the midpoint and range of the intervals.

2.1 Bayesian linear regression method 1 - BLRM1

Let $E = \{e_1, \ldots, e_n\}$ be a set of examples that are described by $p+1$ interval-valued variables Y, $X_1, \ldots, X_p$. Each example $e_i \in E$ $(i = 1, \ldots, n)$ represents an interval quantitative feature vector $\mathbf{z}_i = (\mathbf{x}_i, y_i)$, $\mathbf{x}_i = (x_{i1}, \ldots, x_{ip})$, with $x_{ij} = [a_{ij}, b_{ij}] \in \Im = \{[a,b] : a, b \in \Re, a \leq b\}$ $(j = 1, \ldots, p)$ and $y_i = [y_{Li}, y_{Ui}] \in \Im$ being, respectively, the observed values of X_j and Y.

Let $Y_i = (Y_{1i}, Y_{2i})$ be a bivarate random vector where Y_{1i} and Y_{2i} are, respectively, the midpoint and the range of a interval-valued variables Y. Let X_{1j_i} and X_{2j_i}, $j = 1, 2, \ldots, p$, be explanatory variables representing the midpoint and the ranges of the interval-valued explanatory variables X_j, which are linearly related with the response variables Y_{1i} e Y_{2i}, respectively.

The BLRM1 approach supposes that the bivariate random vector Y_i have a bivariate Gaussian distribution with means $(\mu_{1i} = \mathbf{X}_{1i}\boldsymbol{\beta}_1,\ \mu_{2i} = \mathbf{X}_{2i}\boldsymbol{\beta}_2)$, $Corr(Y_{1i}, Y_{2i}) = \rho$ and $Var(Y_{1i}) = Var(Y_{2i}) = \sigma^2$, where $\mathbf{X}_1$ and $\mathbf{X}_2$ are known model matrices representing the observed values of the variables X_{1j} and X_{2j}, respectively, $\boldsymbol{\beta}_1$ and $\boldsymbol{\beta}_2$ are vectors of parameters to be estimated, μ_{1i} and μ_{2i} are the means of the response variables Y_{1i} and Y_{2i}, respectively, with $\boldsymbol{\beta}_1 = (\beta_{1_0}, \ldots, \beta_{1_p})^T$ and $\boldsymbol{\beta}_2 = (\beta_{2_0}, \ldots, \beta_{2_p})^T$. Then, the density probability function of the bivariate random vector Y_i is given by

$$
\begin{aligned}
f(y \mid \theta) = {} & \exp\left\{\lambda y_{1i}\left[\mu_{1i} - \rho\mu_{2i}\right] + \lambda y_{2i}\left[\mu_{2i} - \rho\mu_{1i}\right]\right\} \\
& \times \exp\left\{-\frac{1}{2}\lambda y_{1i}^2 - \frac{1}{2}\lambda y_{2i}^2 + \rho\lambda y_{1i} y_{2i}\right\} \\
& \times \exp\left\{-\frac{1}{2}\lambda(\mu_{1i})^2 - \frac{1}{2}\lambda(\mu_{2i})^2\right\} \\
& \times \exp\left\{\rho\lambda(\mu_{1i}\mu_{2i}) - \log\left(\frac{1}{\phi}\sqrt{1-\rho^2}\right) - \log(2\pi)\right\},
\end{aligned}
\tag{1}
$$

where: $\phi = \frac{1}{\sigma^2}$ and $\lambda = \frac{\phi}{1-\rho^2}$. In the BLRM1 the quantities β_1, β_2, ϕ e ρ are unknown parameters.

The maximum likelihood method is a procedure based on likelihood function $L(\theta \mid y)$ to find the value of θ that maximizes $L(\theta \mid y)$ and corresponds to more well-known estimation method in statistics. Sometimes, to find the estimator of θ is necessary only obtain the maximum point of the likelihood function. However, for more complicated likelihood functions the solution is obtained through numerical methods. The maximum likelihood estimator (MLE) presents interesting asymptotic properties like unbiased, minimal variance and Gaussian sample distribution making easy the use of inferential procedures. Thus, let $y = (y_1, \ldots, y_n)$ be a random sample of (1), the likelihood function for $\theta = (\phi, \rho, \beta_1, \beta_2)$ is given by

$$
\begin{aligned}
L(\theta \mid y) = &\exp\left\{\lambda \sum_{i=1}^{n} y_{1i}\left[\mu_{1i} - \rho\mu_{2i}\right] + \lambda \sum_{i=1}^{n} y_{2i}\left[\mu_{2i} - \rho\mu_{1i}\right]\right\} \\
&\times \exp\left\{-\frac{1}{2}\lambda \sum_{i=1}^{n} y_{1i}^2 - \frac{1}{2}\lambda \sum_{i=1}^{n} y_{2i}^2 + \rho\lambda \sum_{i=1}^{n} y_{1i}y_{2i}\right\} \\
&\times \exp\left\{-\frac{1}{2}\lambda \sum_{i=1}^{n} (\mu_{1i})^2 - \frac{1}{2}\lambda \sum_{i=1}^{n} (\mu_{2i})^2\right\} \\
&\times \exp\left\{\rho\lambda \sum_{i=1}^{n} \mu_{1i}\mu_{2i} - n\log\left(\frac{1}{\phi}\sqrt{1-\rho^2}\right) - n\log(2\pi)\right\}.
\end{aligned}
\tag{2}
$$

2.1.1 Inferential aspects

Prior distributions

Under a Bayesian perspective, prior distributions for the parameters must be specified. Let suppose, for the vector of parameters $\boldsymbol{\beta}_1$ and $\boldsymbol{\beta}_2$, the following multivariate Gaussian distributions:

$$\boldsymbol{\beta}_1 \sim N_{p1}(m_1, V_1),$$

$$\boldsymbol{\beta}_2 \sim N_{p2}(m_2, V_2),$$

where m_1, m_2, V_1 and V_2 are known and reflect the prior knowledge of the research about these parameters. High values in the diagonal of the matrices V_1 e V_2 suggest low prior information about the parameters $\boldsymbol{\beta}_1$ and $\boldsymbol{\beta}_2$, respectively.

The prior specification for the parameter ϕ is given by a Gamma distribution with shape parameter a_1 and scale parameter b_1, e.g.,

$$\phi \sim G(a_1, b_1).$$

Values of $a_1 \to 0$ and $b_1 \to 0$ suggest low prior information about ϕ. Finally, we consider an uniform distribution for parameter ρ, e.g.,

$$\rho \sim U[-1, 1].$$

Then, the prior distribution for θ is given by $P(\theta) = P(\beta_1)P(\beta_2)P(\phi)P(\rho)$.

Posterior distribution

The posterior distribution for θ is obtained from Bayes theorem, considering the prior distribution of θ, previously specified, and the likelihood function (2). Thus,

$$P(\theta \mid y) \propto L(\theta \mid y)P(\theta). \tag{3}$$

The posterior distribution (3) has no closed form. In this way, is necessary the use of Monte Carlo Markov Chain (MCMC) methods in order to bring it.

2.1.2 Computational aspects

The Gibbs sample algorithm with Metropolis-Hasting step (Hastings, 1970) is used to obtain samples for the posterior distribution (3). The samples for the parameters β_1, β_2 and ϕ will be obtained from:

$$\beta_1 \mid \theta_{-\{\beta_1\}} \sim N(m_1^*, V_1^*),$$

$$\beta_2 \mid \theta_{-\{\beta_2\}} \sim N(m_2^*, V_2^*),$$

$$\phi \mid \theta_{-\{\phi\}} \sim G(a_1^*, b_1^*),$$

where: $m_1^* = (A + m_1' V_1^{-1})(D + V_1^{-1})^{-1}$, $V_1^* = (D + V_1^{-1})^{-1}$ with $A = \lambda \sum_{i=1}^{n} [y_{1i} X_{1i} - \rho y_{2i} X_{1i} + \rho \mu_{2i} X_{1i}]$ and $D = \lambda \sum_{i=1}^{n} X_{1i}' X_{1i}$; $m_2^* = (E + m_2' V_2^{-1})(H + V_2^{-1})$, $V_2^* = (H + V_2^{-1})^{-1}$ with $E = \lambda \sum_{i=1}^{n} [y_{2i} X_{2i} - \rho y_{1i} X_{2i} + \rho \mu_{1i} X_{2i}]$ and $H = \lambda \sum_{i=1}^{n} X_{2i}' X_{2i}$; $a_1^* = n + a_1$, $b_1^* = \sum_{i=1}^{n} Z_i' \Sigma^{-1} Z_i + b_1$, $Z_i' = (y_{1i} - \mu_{1i}, y_{2i} - \mu_{2i})$ and $\Sigma = \phi^{-1}$.

The marginal posterior density for ρ, based on Gibbs algorithm, is given by

$$P(\rho \mid \theta_{-\{\rho\}}) \propto \exp\left\{ -\frac{1}{2} \sum_{i=1}^{n} Z_i' \Sigma^{-1} Z_i - n \log \sqrt{1 - \rho^2} \right\}, \tag{4}$$

where $\theta_{-\{\rho\}}$ denotes the vector of parameter θ, excluding ρ. Note that this distribution has no closed form and the proposed ρ, denoted by (ρ^{prop}), will be generated from the uniform transfer density function g:

$$\rho^{prop} \sim U[\max\{-1, \rho^{act} - u_1\}, \min\{1, \rho^{act} + u_1\}],$$

where $U[a, b]$ is an uniform distribution in the interval $[a, b]$, ρ^{act} is the current ρ and u_1 is a value in the unit interval $(0, 1)$, such that, the acceptance rate will be around 40%. The proposed value for ρ will be accept with probability between 1 e p, where p is given by

$$p = \frac{P(\rho^{prop} \mid \theta_{-\{\rho\}}) g(\rho^{act} \mid \rho^{prop})}{P(\rho^{act} \mid \theta_{-\{\rho\}}) g(\rho^{prop} \mid \rho^{act})}. \tag{5}$$

2.2 Bayesian linear regression method 2 - BLRM2

Let $Y_i = (Y_{1i}, Y_{2i})$ be a bivarate random vector with Y_{1i} and Y_{2i} representing, again, the midpoint and the range of a interval-valued variables Y, respectively. Let X_{1j_i} and X_{2j_i} be explanatory variables representing the midpoint and the ranges of the interval-valued explanatory variables X_j and linearly related with the response variables Y_{1i} e Y_{2i}, respectively.

The BLRM2 approach supposes that the bivariate random vector Y_i have a bivariate Gaussian distribution with means ($\mu_{1i} = \mathbf{X}_{1i} \boldsymbol{\beta}_1$, $\mu_{2i} = \mathbf{X}_{2i} \boldsymbol{\beta}_2$) and $Corr(Y_{1i}, Y_{2i}) = \rho$. However, we consider now different variances for the bivarate random vector $Y_i = (Y_{1i}, Y_{2i})$, e.g., $Var(Y_{1i}) = \sigma_1^2$ and $Var(Y_{2i}) = \sigma_2^2$. The matrices $\mathbf{X}_1$ and $\mathbf{X}_2$ are the known model matrices representing the observed values of the variables X_{1j} and X_{2j}, respectively, $\boldsymbol{\beta}_1$ and $\boldsymbol{\beta}_2$ are vectors of parameters to be estimated, μ_{1i} and μ_{2i} are the means of the response variables Y_{1i} and Y_{2i}, respectively, with $\boldsymbol{\beta}_1 = (\beta_{1_0}, \ldots, \beta_{1_p})^T$ and $\boldsymbol{\beta}_2 = (\beta_{2_0}, \ldots, \beta_{2_p})^T$. Then, the density probability function of the bivariate random vector Y_i is given by

$$
\begin{aligned}
f(\boldsymbol{y} \mid \boldsymbol{\theta}) = & \exp\left\{ \frac{1}{1-\rho^2} y_{1i} \left[\phi_1 \mu_{1i} - \sqrt{\phi_1 \phi_2} \rho \mu_{2i} \right] \right\} \\
& \times \exp\left\{ \frac{1}{1-\rho^2} y_{2i} \left[\phi_2 \mu_{2i} - \sqrt{\phi_1 \phi_2} \rho \mu_{1i} \right] \right\} \\
& \times \exp\left\{ -\frac{1}{2} \lambda_1 y_{1i}^2 - \frac{1}{2} \lambda_2 y_{2i}^2 + \frac{\sqrt{\phi_1 \phi_2} \rho}{1-\rho^2} y_{1i} y_{2i} \right\} \\
& \times \exp\left\{ -\frac{1}{2} \lambda_1 (\mu_{1i})^2 - \frac{1}{2} \lambda_2 (\mu_{2i})^2 \right\} \\
& \times \exp\left\{ \frac{\sqrt{\phi_1 \phi_2} \rho}{1-\rho^2} \mu_{1i} \mu_{2i} - \log\left(\frac{\sqrt{1-\rho^2}}{\sqrt{\phi_1 \phi_2}} \right) - \log(2\pi) \right\},
\end{aligned} \tag{6}
$$

where: $\phi_1 = \frac{1}{\sigma_1^2}$, $\phi_2 = \frac{1}{\sigma_2^2}$, $\lambda_1 = \frac{\phi_1}{1-\rho^2}$ and $\lambda_2 = \frac{\phi_2}{1-\rho^2}$. In the BLRM2 the quantities $\boldsymbol{\beta}_1, \boldsymbol{\beta}_2, \phi_1, \phi_2$ and ρ are unknown parameters. Furthermore, let $\boldsymbol{y} = (\boldsymbol{y}_1, \ldots, \boldsymbol{y}_n)$ be a random sample of (6), the likelihood function for $\boldsymbol{\theta} = (\phi_1, \phi_2, \rho, \boldsymbol{\beta}_1, \boldsymbol{\beta}_2)$ is given by,

$$
\begin{aligned}
L(\boldsymbol{\theta} \mid \boldsymbol{y}) = & \exp\left\{ \frac{1}{1-\rho^2} \sum_{i=1}^{n} y_{1i} \left[\phi_1 \mu_{1i} - \sqrt{\phi_1 \phi_2} \rho \mu_{2i} \right] \right\} \\
& \times \exp\left\{ \frac{1}{1-\rho^2} \sum_{i=1}^{n} y_{2i} \left[\phi_2 \mu_{2i} - \sqrt{\phi_1 \phi_2} \rho \mu_{1i} \right] \right\} \\
& \times \exp\left\{ -\frac{1}{2} \lambda_1 \sum_{i=1}^{n} y_{1i}^2 - \frac{1}{2} \lambda_2 \sum_{i=1}^{n} y_{2i}^2 + \frac{\sqrt{\phi_1 \phi_2} \rho}{1-\rho^2} y_{1i} y_{2i} \right\} \\
& \times \exp\left\{ -\frac{1}{2} \lambda_1 \sum_{i=1}^{n} (\mu_{1i})^2 - \frac{1}{2} \lambda_2 \sum_{i=1}^{n} (\mu_{2i})^2 \right\} \\
& \times \exp\left\{ \frac{\sqrt{\phi_1 \phi_2} \rho}{1-\rho^2} \sum_{i=1}^{n} \mu_{1i} \mu_{2i} - n \log\left(\frac{\sqrt{1-\rho^2}}{\sqrt{\phi_1 \phi_2}} \right) - n \log(2\pi) \right\}.
\end{aligned} \tag{7}
$$

2.2.1 Inferential aspects

Prior and posterior distributions

The prior distributions for the vectors of parameters $\boldsymbol{\beta}_1$ and $\boldsymbol{\beta}_2$ follow the same structure defined in the Section 2.1.1, based on a multivariate Gaussian distribution. However, in the BLRM2 approach, is necessary to specify the priors for the dispersion parameters ϕ_1 and ϕ_2:

$$\phi_i \sim G(a_i, b_i), i = 1, 2.$$

Values of $a_i \to 0$ and $b_i \to 0$, suggest low prior information about ϕ_i, $i = 1, 2$. For the parameter ρ we continue considering an uniform distribution in the interval $(-1, 1)$, e.g.,

$$\rho \sim U(-1, 1).$$

Then, the prior distribution for $\boldsymbol{\theta}$ is given by

$$P(\boldsymbol{\theta}) = P(\boldsymbol{\beta}_1)P(\boldsymbol{\beta}_2)P(\phi_1)P(\phi_2)P(\rho).$$

The posterior distribution for $\boldsymbol{\theta}$ will be obtained from the prior distributions, previously specified, and the likelihood function (7) as follow:

$$P(\boldsymbol{\theta} \mid \boldsymbol{y}) \propto L(\boldsymbol{\theta} \mid \boldsymbol{y})P(\boldsymbol{\theta}). \tag{8}$$

The posterior distribution (8) has no closed form being necessary the use of Monte Carlo Markov Chain (MCMC) methods in order to bring it.

2.2.2 Computational aspects

The Gibbs sample algorithm with Metropolis-Hasting step (Hastings, 1970) also is used to obtain samples for the posterior distribution (8). The samples for the vector of parameters β_1 and β_2 will be obtained from:

$$\boldsymbol{\beta}_1 \mid \boldsymbol{\theta}_{-\{\beta_1\}} \sim N(\boldsymbol{m}_1^*, \boldsymbol{v}_1^*),$$
$$\boldsymbol{\beta}_2 \mid \boldsymbol{\theta}_{-\{\beta_2\}} \sim N(\boldsymbol{m}_2^*, \boldsymbol{v}_2^*),$$

where: $\boldsymbol{m}_1^* = (\boldsymbol{A} + \boldsymbol{m}_1'\boldsymbol{v}_1^{-1})(\boldsymbol{D} + \boldsymbol{v}_1^{-1})^{-1}$, $\boldsymbol{v}_1^* = (\boldsymbol{D} + \boldsymbol{v}_1^{-1})^{-1}$, with $\boldsymbol{A} = \frac{1}{1-\rho^2}\sum_{i=1}^n[\phi_1 y_{1i}\boldsymbol{X}_{1i} - \sqrt{\phi_1\phi_2}\rho y_{2i}\boldsymbol{X}_{1i} + \sqrt{\phi_1\phi_2}\rho\mu_{2i}\boldsymbol{X}_{1i}]$ and $\boldsymbol{D} = \lambda_1\sum_{i=1}^n \boldsymbol{X}_{1i}'\boldsymbol{X}_{1i}$; $\boldsymbol{m}_2 = (\boldsymbol{E} + \boldsymbol{m}_2'\boldsymbol{v}_2^{-1})(\boldsymbol{H} + \boldsymbol{v}_2^{-1})$, $\boldsymbol{v}_2^* = (\boldsymbol{H} + \boldsymbol{v}_2^{-1})^{-1}$, with $\boldsymbol{E} = \frac{1}{1-\rho^2}\sum_{i=1}^n[\phi_2 y_{2i}\boldsymbol{X}_{2i} - \sqrt{\phi_1\phi_2}\rho y_{1i}\boldsymbol{X}_{2i} + \sqrt{\phi_1\phi_2}\rho\mu_{1i}\boldsymbol{X}_{2i}]$ and $\boldsymbol{H} = \lambda_2\sum_{i=1}^n \boldsymbol{X}_{2i}'\boldsymbol{X}_{2i}$.

In BLRM2, the marginal posterior densities for the parameters ϕ_1 and ϕ_2 will be obtained taking into account the Gibbs algorithm, being given by

$$P(\phi_i \mid \boldsymbol{\theta}_{-\{\phi_i\}}) \propto \exp\left\{-\frac{1}{2}\sum_{i=1}^n \boldsymbol{Z}_i'\boldsymbol{\Sigma}^{-1}\boldsymbol{Z}_i - \frac{n}{2}\log(\phi_i) + (a_i - 1) \times \log(\phi_i) - b_i\phi_i\right\}, \tag{9}$$

where $\boldsymbol{Z}_i' = (y_{1i} - \mu_{1i}, y_{2i} - \mu_{2i})$ and $\boldsymbol{\Sigma} = \begin{pmatrix} \phi_1^{-1} & (\phi_1\phi_2)^{-\frac{1}{2}}\rho \\ (\phi_1\phi_2)^{-\frac{1}{2}}\rho & \phi_2^{-1} \end{pmatrix}$. This distribution has no closed form and the proposed ϕ_i, denoted by ϕ_i^{prop}, will be generated from the Gamma transfer density function g:

$$\phi_i^{prop} \sim G(\phi^{act} u, u),$$

where: ϕ_i^{act} is the current ϕ_i and u is a positive value, such that, the acceptance rate will be around 40%. The proposed value for ϕ_i will be accepted with probability between 1 and p, with p given by

$$p = \frac{P(\phi_i^{prop} \mid \boldsymbol{\theta}_{-\{\phi_i\}}) g(\phi_i^{act} \mid \phi_i^{prop})}{P(\phi_i^{act} \mid \boldsymbol{\theta}_{-\{\phi_i\}}) g(\phi_i^{prop} \mid \phi_i^{act})}. \quad (10)$$

Finally, the complete conditional for ρ is given by

$$P(\rho \mid \boldsymbol{\theta}_{-\{\rho\}}) \propto \exp\left\{ -\frac{1}{2} \sum_{i=1}^{n} Z_i' \Sigma^{-1} Z_i - n \log \sqrt{1-\rho^2} \right\}, \quad (11)$$

where $\boldsymbol{Z}_i' = (y_{1i} - \mu_{1i}, y_{2i} - \mu_{2i})$ and $\boldsymbol{\Sigma} = \begin{pmatrix} \phi_1^{-1} & (\phi_1\phi_2)^{-\frac{1}{2}}\rho \\ (\phi_1\phi_2)^{-\frac{1}{2}}\rho & \phi_2^{-1} \end{pmatrix}$. In the same way, this distribution has no closed form and the proposed ρ (ρ^{prop}) is generated from the uniform transfer density function g, given by

$$\rho^{prop} \sim U[\max\{-1, \rho^{act} - u_1\}, \min\{1, \rho^{act} + u_1\}],$$

as defined in the section 2.1.2.

3 Numerical study and applications

3.1 Synthetic data sets

Synthetic data set are considered for evaluate the parameters estimates of the BLRM1 and BLRM2 approaches. For BLRM1 approach, we built artificial data sets considering: $X_{1i} = (1, W_{1i})$ with $W_{1i} \sim N(2, 1)$, $X_{2i} = (1, W_{2i})$ with $W_{2i} \sim N(5, 1)$, $\beta_1 = (30, 13)$, $\beta_2 = (2.5,\ 1.5)$ and $\sigma^2 = 0.5$, e.g, $\phi = 2$. We also considered three different levels for the dependence between Y_1 and Y_2 $\rho = \{0.0, \pm 0.5, \pm 0.9\}$ and three different sample sizes $n = \{20, 100, 300\}$. The hyper-parameters of non-informative priors used in numerical study are: $\boldsymbol{\beta}_1 \sim N_2(\boldsymbol{m}_1^* = \boldsymbol{0}, \boldsymbol{v}_1^* = 1000 I_2)$ and $\boldsymbol{\beta}_2 \sim N_2(\boldsymbol{m}_2^* = \boldsymbol{0}, \boldsymbol{v}_2^* = 1000 I_2)$, where $\boldsymbol{m}_1^*$ and $\boldsymbol{m}_2^*$ are zero's vectors 2×1 and I_2 is an identity matrix 2×2; $\phi_i \sim Gamma(a_i = 0.0001, b_i = 0.0001)$, where $E(\phi_i) = a_i/b_i, Var(\phi_i) = a_i/(b_i)^2$, $i = \{1, 2\}$, and $\rho \sim U(-1, 1)$.

The samples for posterior distribution of the parameters in BLRM1 are generated as specified in Section 2.1.2. In the estimation of the parameter's posterior distribution we used 10,000 replicates, discarding the first 2,500 replicates (burn-in stage). The `R` software, version 2.15.0, was used as computational support. We have considered the Gelman and Rubin (1992) criterion to verify the convergence of the posterior distribution of the parameters. This criterion is available in the library CODA (Cowless et al., 1997) of the software R. Table 1 presents the results for BLRM1 in terms of the potential scale reduction factor (PSRF) and his credible interval (upper limit). It is possible verify the convergence of the chains since the criterion is close to 1, for all scenarios considered in the simulation study. The same occurs for BLRM2 model. Finally, the BLRM methods are compared with the Constrained Center and Range Method (CCRM), proposed by Lima Neto and De Carvalho (2011), in a Monte Carlo simulation scheme with 500 runs.

Table 1 - Potential scale reduction factor for the parameter estimates in BLRM1, according to correlation coefficient and sample size

	Parameter	PSRF	IC 95%	PSRF	IC 95%	PSRF	IC 95%
		$n = 20$		n=100		n=300	
	β_{01}	1.01	1.05	1.01	1.03	1.01	1.01
	β_{11}	1.01	1.01	1.01	1.01	1.01	1.01
$\rho = -0.9$	β_{02}	1.01	1.02	1.01	1.01	1.01	1.01
	β_{12}	1.01	1.01	1.01	1.01	1.01	1.01
	ϕ	1.01	1.01	1.01	1.01	1.01	1.02
	ρ	1.02	1.06	1.01	1.01	1.01	1.05
	β_{01}	1.01	1.01	1.01	1.01	1.01	1.01
	β_{11}	1.01	1.01	1.01	1.01	1.01	1.01
$\rho = -0.5$	β_{02}	1.01	1.01	1.01	1.01	1.01	1.01
	β_{12}	1.01	1.01	1.01	1.01	1.01	1.01
	ϕ	1.01	1.01	1.01	1.01	1.01	1.01
	ρ	1.04	1.18	1.01	1.02	1.01	1.01
	β_{01}	1.01	1.01	1.01	1.01	1.01	1.01
	β_{11}	1.01	1.01	1.01	1.01	1.01	1.01
$\rho = 0.0$	β_{02}	1.01	1.01	1.01	1.01	1.01	1.01
	β_{12}	1.01	1.01	1.01	1.01	1.01	1.01
	ϕ	1.01	1.03	1.01	1.01	1.01	1.01
	ρ	1.03	1.04	1.01	1.03	1.01	1.02
	β_{01}	1.01	1.04	1.01	1.01	1.01	1.01
	β_{11}	1.01	1.04	1.01	1.01	1.01	1.01
$\rho = 0.5$	β_{02}	1.01	1.01	1.01	1.01	1.01	1.01
	β_{12}	1.01	1.01	1.01	1.01	1.01	1.01
	ϕ	1.01	1.03	1.01	1.01	1.01	1.01
	ρ	1.1	1.37	1.01	1.01	1.01	1.01
	β_{01}	1.01	1.01	1.01	1.03	1.01	1.01
	β_{11}	1.01	1.01	1.01	1.01	1.01	1.01
$\rho = 0.9$	β_{02}	1.01	1.01	1.02	1.01	1.01	1.01
	β_{12}	1.01	1.01	1.01	1.01	1.01	1.01
	ϕ	1.01	1.01	1.01	1.01	1.01	1.01
	ρ	1.01	1.01	1.01	1.01	1.01	1.01

Figure 1 illustrates the posterior distributions of the BLRM1 parameters. The results demonstrate that the posterior distributions are closer to true parameter as higher is the sample size.

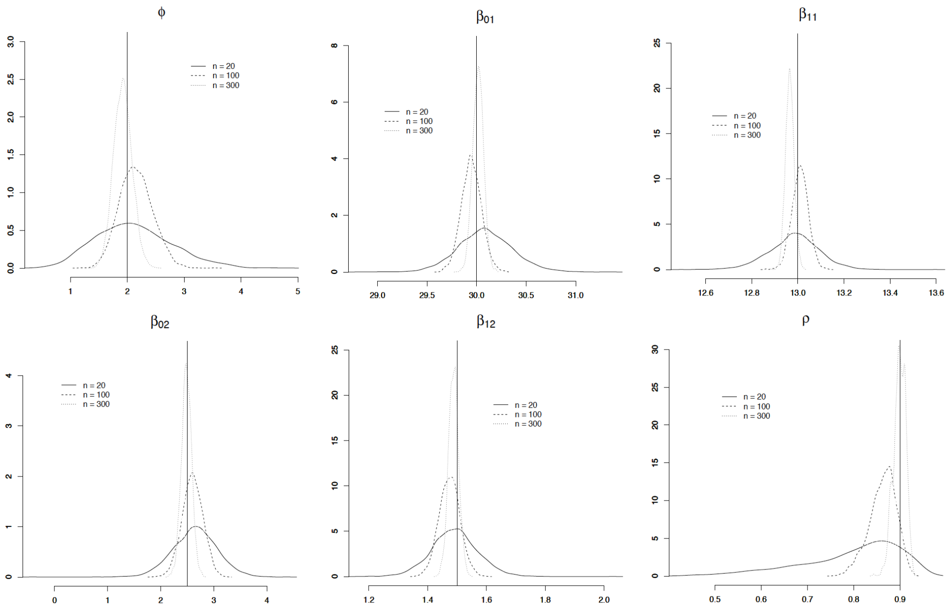

Figure 1 - Posterior distributions of the BLRM1 parameters ($\rho = 0.9$).

Table 2 presents the percentage of coverage for the true parameter in the credible (BLRM) and confidence (CCRM) intervals, after 500 Monte Carlo runs. The results demonstrated that the coverage of the BLRM methods outperform the CCRM method in the most of configurations, more precisely, in 82% of the times. Particularly, when $\rho = \{-0.9, +0.9\}$ and n = $\{20, 100\}$ the coverage of the BLRM is better than CCRM approach in 92% of the times. For the dispersion (ϕ) and correlation (ρ) parameters, we also verified a coverage rate close to the nominal level of 95%. in BLRM methods.

Tables 3 and 4 illustrate, more detailed, the credible intervals (BLRM1) and the confidence intervals (CCRM) of the parameter estimates for a particular Monte Carlo run.

For BLRM1 approach (Table 3), we observed that the precision of the credible interval (*Prec*) increases when the sample size increases. This corroborates the hypothesis that, as greater is the sample size as greater is the precision of the credible interval. Table 4 presents the confidence intervals for the parameters estimates of the CCRM (Lima Neto and De Carvalho, 2010).

We conclude, based on the results of the columns *Prec*, that the credible intervals (BLRM1) have presented a better accuracy when compared with

Table 2 - Coverage rate of the true parameters for the credible and confidence intervals according to method, correlation coefficient and sample size, after 500 Monte Carlo runs

Parameter		BLRM1			CCRM			BLRM2			CCRM		
		20	100	300	20	100	300	20	100	300	20	100	300
β_{01}	30	96.1	96.6	96.2	93.0	95.2	95.4	96.8	94.4	94.8	94.2	93.0	96.4
β_{11}	13	97.6	96.4	95.4	92.4	94.0	95.0	97.8	97.8	94.8	93.2	92.0	96.4
β_{02}	2.5	97.8	95.4	95.2	93.8	94.0	94.2	97.4	95.4	93.8	96.6	95.2	93.2
β_{12}	1.5	97.7	96.6	95.4	93.3	93.4	94.2	97.4	96.2	93.6	97.0	95.2	92.8
ϕ	2	95.9	95.8	96.4	-	-	-	-	-	-	-	-	-
ρ	-0.9	88.3	94.0	93.6	-	-	-	84.8	94.2	92.8	-	-	-
$\phi1$	4	-	-	-	-	-	-	93.8	94.4	91.0	-	-	-
$\phi2$	0.25	-	-	-	-	-	-	95.8	95.2	90.4	-	-	-
β_{01}	30	95.2	95.8	95.2	93.4	94.8	95.4	95.4	94.2	95.6	93.0	93.6	94.8
β_{11}	13	94.4	96.2	94.0	91.8	94.8	94.8	94.4	95.4	93.4	92.6	94.6	94.8
β_{02}	2.5	94.8	94.4	95.2	91.0	95.6	94.8	96.0	94.6	95.2	98.0	93.4	94.4
β_{12}	1.5	95.0	95.0	94.4	91.0	96.4	95.2	96.0	94.2	95.2	97.0	93.8	94.4
ϕ	2	96.2	94.6	93.4	-	-	-	-	-	-	-	-	-
ρ	-0.5	90.6	94.6	93.8	-	-	-	93.2	94.6	94.6	-	-	-
$\phi1$	4	-	-	-	-	-	-	92.6	96.0	95.0	-	-	-
$\phi2$	0.25	-	-	-	-	-	-	93.2	96.0	92.2	-	-	-
β_{01}	30	95.0	95.4	94.8	94.0	95.2	95.2	94.8	93.2	95.2	94.6	93.4	95.8
β_{11}	13	95.2	94.8	94.4	93.4	94.0	94.6	94.4	94.2	95.4	94.2	93.4	94.6
β_{02}	2.5	94.6	94.0	93.8	93.8	94.8	94.0	94.6	95.8	94.8	96.0	96.4	94.4
β_{12}	1.5	95.2	94.4	95.0	93.4	94.4	94.6	94.8	95.6	94.2	95.0	96.2	94.8
ϕ	2	95.0	94.6	96.8	-	-	-	-	-	-	-	-	-
ρ	0.0	92.2	93.2	93.8	-	-	-	94.0	93.4	93.4	-	-	-
$\phi1$	4	-	-	-	-	-	-	93.8	94.2	93.2	-	-	-
$\phi2$	0.25	-	-	-	-	-	-	93.2	95.6	93.4	-	-	-
β_{01}	30	94.2	95.6	94.2	93.8	95.4	95.0	95.6	94.2	94.6	92.8	92.8	93.6
β_{11}	13	94.6	94.8	94.0	94.2	94.0	95.0	95.0	94.4	94.0	91.6	92.8	95.2
β_{02}	2.5	94.6	96.2	94.6	91.8	95.2	94.8	95.6	95.0	95.0	96.8	94.8	94.8
β_{12}	1.5	95.2	95.6	96.0	93.2	95.0	95.4	95.8	94.6	94.8	96.6	94.2	94.8
ϕ	2	96.6	93.6	94.4	-	-	-	-	-	-	-	-	-
ρ	0.5	91.6	95.4	96.0	-	-	-	93.4	95.6	93.4	-	-	-
$\phi1$	4	-	-	-	-	-	-	95.4	93.0	95.4	-	-	-
$\phi2$	0.25	-	-	-	-	-	-	95.6	94.2	92.2	-	-	-
β_{01}	30	97.4	94.2	94.8	93.8	94.8	93.8	95.6	92.6	95.2	91.2	94.4	94.6
β_{11}	13	98.2	95.2	93.4	93.6	94.6	94.2	98.4	94.6	96.0	93.4	94.2	94.6
β_{02}	2.5	98.6	95.0	96.4	93.6	94.8	95.2	97.0	95.4	96.2	97.0	94.4	93.8
β_{12}	1.5	98.6	95.6	95.6	93.0	94.8	94.8	97.4	94.8	97.0	97.4	94.0	93.8
ϕ	2	95.8	95.2	94.6	-	-	-	-	-	-	-	-	-
ρ	0.9	89.8	93.4	93.4	-	-	-	89.2	93.2	93.4	-	-	-
$\phi1$	4	-	-	-	-	-	-	94.6	95.0	91.6	-	-	-
$\phi2$	0.25	-	-	-	-	-	-	94.6	94.2	93.0	-	-	-

confidence intervals (CCRM). In this way, the Bayesian estimation process (BLRM1) presented an improvement in the parameters estimates if compared with the frequentist approach (CCRM).

Table 5 presents the credible intervals for the parameter estimates of the BLRM2 and Table 6 presents the confidence intervals for the parameter estimates of the CCRM method. In this case, the synthetic data set were built considering different variances for Y_1 and Y_2, e.g., $\sigma_1^2 = 0.25$ and $\sigma_2^2 = 4$ or, respectively, $\phi_1 = 4$ and $\phi_2 = 0.25$. The other simulation parameters were the same used in BLRM1. We obtained the same conclusions verified when we compared the BLRM2 and CCRM approaches. The credible intervals (BLRM2) also presented a better accuracy when compared with confidence intervals (CCRM). In this way, the Bayesian estimation process (BLRM2) also presented an improvement, in the parameter estimates, if compared with the frequentist approach (CCRM).

Moreover, the BLRM approaches present estimate and the credible intervals for the parameters ϕ and ρ, in contrast with CCRM method. The parameters chain

Table 3 - Credible interval for the parameters estimates in BLRM1, according to correlation coefficient ρ and sample size

	θ	50%	2.5%	97.5%	Prec	50%	2.5%	97.5%	Prec	50%	2.5%	97.5%	Prec
			n=20				n=100				n=300		
β_{01}	30	29.40	28.92	29.84	0.92	30.19	30.01	30.38	0.37	30.02	29.91	30.12	0.21
β_{11}	13	13.27	13.09	13.46	0.38	12.96	12.90	13.02	0.12	13.00	12.97	13.03	0.07
β_{02}	2.5	2.12	1.07	3.17	2.10	2.56	2.20	2.91	0.71	2.51	2.33	2.70	0.37
β_{12}	1.5	1.56	1.38	1.75	0.36	1.47	1.41	1.54	0.13	1.50	1.47	1.53	0.07
ϕ	2	2.76	1.31	4.64	3.34	1.73	1.28	2.26	0.98	1.85	1.56	2.16	0.60
ρ	-0.9	-0.78	-0.91	-0.51	0.40	-0.92	-0.95	-0.87	0.08	-0.92	-0.93	-0.89	0.04
β_{01}	30	30.27	29.50	31.09	1.59	30.07	29.77	30.35	0.57	29.97	29.81	30.14	0.33
β_{11}	13	12.91	12.58	13.22	0.64	13.01	12.89	13.14	0.24	13.00	12.93	13.08	0.15
β_{02}	2.5	3.02	1.65	4.38	2.72	2.58	3.22	1.97	1.25	2.45	2.14	2.77	0.64
β_{12}	1.5	1.40	1.13	1.66	0.54	1.48	1.36	1.60	0.24	1.51	1.45	1.57	0.12
ϕ	2	1.93	1.02	3.12	2.10	1.81	1.44	2.24	0.80	2.10	1.84	2.37	0.53
ρ	-0.5	-0.53	-0.79	-0.09	0.70	-0.53	-0.65	-0.37	0.28	-0.51	-0.59	-0.42	0.17
β_{01}	30	30.06	29.48	30.63	1.15	29.80	29.48	30.11	0.63	30.07	29.88	30.26	0.37
β_{11}	13	12.91	12.66	13.17	0.52	13.12	12.97	13.26	0.29	12.99	12.91	13.08	0.17
β_{02}	2.5	2.36	1.10	3.36	2.57	2.51	1.82	3.24	1.42	2.58	2.17	2.98	0.81
β_{12}	1.5	1.49	1.22	1.75	0.52	1.50	1.36	1.64	0.28	1.49	1.41	1.57	0.16
ϕ	2	3.10	1.83	4.80	2.97	2.00	1.63	2.42	0.79	2.08	1.85	2.33	0.48
ρ	0.0	-0.12	-0.49	0.25	0.74	-0.03	-0.21	0.16	0.38	0.07	-0.05	0.18	0.23
β_{01}	30	29.32	28.58	30.09	1.51	29.98	29.70	30.26	0.56	29.95	29.79	30.12	0.33
β_{11}	13	13.23	12.81	13.62	0.82	13.02	12.89	13.15	0.26	13.05	12.97	13.12	0.16
β_{02}	2.5	2.48	0.74	4.16	3.42	2.70	2.09	3.30	1.20	2.57	2.19	2.95	0.76
β_{12}	1.5	1.53	1.23	1.85	0.63	1.47	1.36	1.59	0.23	1.49	1.42	1.57	0.15
ϕ	2	1.71	0.98	2.69	1.71	1.80	1.43	2.21	0.78	1.96	1.72	2.22	0.51
ρ	0.5	0.32	0.03	0.66	0.63	0.46	0.31	0.60	0.29	0.51	0.41	0.59	0.17
β_{01}	30	30.03	29.48	30.52	1.04	30.16	29.97	30.36	0.39	30.03	29.92	30.14	0.22
β_{11}	13	12.98	12.77	13.21	0.44	12.95	12.88	13.02	0.14	12.99	12.95	13.03	0.08
β_{02}	2.5	3.63	2.45	4.88	2.44	2.25	1.91	2.60	0.69	2.43	2.22	2.63	0.41
β_{12}	1.5	1.28	1.03	1.52	0.49	1.57	1.50	1.63	0.13	1.51	1.48	1.55	0.07
ϕ	2	1.74	0.78	3.01	2.23	1.80	1.31	2.33	1.02	1.92	1.64	2.23	0.59
ρ	0.9	0.88	0.67	0.96	0.29	0.91	0.85	0.94	0.09	0.89	0.87	0.91	0.05

Table 4 - Confidence Intervals (95%) for the parameters estimates in CCRM method, according to correlation coefficient ρ and sample size. Comparative analysis with BLRM1

ρ		θ	Mean	Lower	Upper	Prec	Mean	Lower	Upper	Prec	Mean	Lower	Upper	Prec
				n=20				n=100				n=300		
-0.9	β_{01}	30	29.24	28.67	29.81	1.13	30.28	29.94	30.61	0.66	30.13	29.96	30.30	0.34
	β_{11}	13	13.35	13.09	13.61	0.52	12.93	12.78	13.07	0.30	12.94	12.86	13.03	0.16
	β_{02}	2.5	2.14	0.52	3.76	3.24	2.88	2.03	3.72	1.70	2.41	1.99	2.84	0.85
	β_{12}	1.5	1.56	1.28	1.85	0.57	1.41	1.24	1.57	0.34	1.52	1.43	1.60	0.17
-0.5	β_{01}	30	30.58	29.79	31.37	1.58	30.06	29.73	30.40	0.67	29.99	29.81	30.17	0.37
	β_{11}	13	12.77	12.45	13.10	0.65	13.01	12.87	13.16	0.29	13.00	12.91	13.08	0.17
	β_{02}	2.5	3.00	1.37	4.63	3.26	3.13	2.45	3.81	1.36	2.35	1.98	2.72	0.74
	β_{12}	1.5	1.40	1.07	1.72	0.65	1.37	1.24	1.50	0.26	1.53	1.46	1.60	0.15
0.0	β_{01}	30	30.07	29.55	30.59	1.05	29.80	29.46	30.13	0.67	30.08	29.90	30.26	0.36
	β_{11}	13	12.91	12.67	13.14	0.47	13.12	12.97	13.27	0.31	12.99	12.91	13.07	0.16
	β_{02}	2.5	2.31	1.00	3.62	2.62	2.53	1.87	3.19	1.32	2.57	2.16	2.99	0.82
	β_{12}	1.5	1.50	1.23	1.76	0.53	1.49	1.37	1.62	0.26	1.49	1.41	1.57	0.16
0.5	β_{01}	30	29.49	28.95	30.02	1.06	30.05	29.72	30.38	0.66	29.96	29.76	30.16	0.40
	β_{11}	13	13.13	12.84	13.43	0.59	12.98	12.82	13.14	0.32	13.04	12.95	13.14	0.19
	β_{02}	2.5	2.18	0.09	4.27	4.18	2.82	2.21	3.44	1.23	2.76	2.37	3.16	0.79
	β_{12}	1.5	1.59	1.21	1.97	0.76	1.45	1.33	1.57	0.24	1.45	1.38	1.53	0.16
0.9	β_{01}	30	29.91	29.04	30.78	1.74	30.24	29.89	30.59	0.70	30.00	29.81	30.18	0.37
	β_{11}	13	13.05	12.60	13.50	0.90	12.91	12.75	13.07	0.33	13.00	12.92	13.09	0.17
	β_{02}	2.5	4.22	1.82	6.61	4.80	2.20	1.46	2.94	1.48	2.33	1.91	2.76	0.85
	β_{12}	1.5	1.16	0.67	1.65	0.98	1.58	1.43	1.73	0.30	1.53	1.45	1.61	0.16

Table 5 - Credible interval for the parameters estimates in BLRM2, following correlation coefficient ρ and sample size

	θ	50%	2,5%	97,5%	Prec	50%	2,5%	97,5%	Prec	50%	2,5%	97,5%	Prec
			n=20				n=100				n=300		
β_{01}	30	29.94	29.63	30.23	0.60	29.98	29.86	30.10	0.24	30.02	29.95	30.10	0.15
β_{11}	13	13.01	12.90	13.12	0.22	13.03	12.99	13.07	0.08	13.00	12.98	13.03	0.05
β_{02}	2.5	2.25	0.41	5.10	4.69	3.34	2.33	4.48	2.15	2.40	1.85	2.97	1.12
β_{12}	1.5	1.65	1.08	2.16	1.08	1.29	1.09	1.48	0.40	1.51	1.41	1.60	0.20
ϕ_1	4	5.49	2.88	9.65	6.78	4.56	3.27	5.88	2.61	4.51	3.85	5.17	1.32
ϕ_2	0.25	0.45	0.20	0.84	0.64	0.23	0.17	0.30	0.13	0.28	0.23	0.31	0.08
ρ	-0.9	-0.64	-0.86	-0.31	0.54	-0.89	-0.93	-0.83	0.10	-0.89	-0.91	-0.85	0.06
β_{01}	30	29.81	29.42	30.22	0.80	29.99	29.74	30.25	0.52	29.97	29.85	30.09	0.24
β_{11}	13	13.11	12.93	13.27	0.34	12.99	12.88	13.11	0.23	13.03	12.98	13.08	0.10
β_{02}	2.5	3.43	0.25	7.02	6.77	1.25	0.42	2.96	2.54	2.15	1.12	3.23	2.11
β_{12}	1.5	1.36	0.62	2.12	1.50	1.72	1.38	2.03	0.65	1.54	1.33	1.74	0.42
ϕ_1	4	6.05	2.86	10.28	7.42	3.10	2.27	4.09	1.82	4.18	3.50	4.90	1.40
ϕ_2	0.25	0.37	0.18	0.65	0.47	0.25	0.19	0.33	0.14	0.26	0.22	0.31	0.09
ρ	-0.5	-0.39	-0.70	-0.07	0.63	-0.42	-0.56	-0.26	0.31	-0.43	-0.53	-0.34	0.20
β_{01}	30	29.87	29.32	30.40	1.08	29.75	29.53	30.01	0.48	29.98	29.85	30.11	0.27
β_{11}	13	13.08	12.83	13.34	0.51	13.09	12.99	13.19	0.20	13.02	12.96	13.09	0.12
β_{02}	2.5	0.57	5.36	6.44	11.80	2.35	0.09	4.46	4.38	2.68	1.40	3.97	2.57
β_{12}	1.5	1.88	0.79	3.02	2.23	1.57	1.15	2.00	0.86	1.46	1.20	1.71	0.51
ϕ_1	4	4.75	1.97	9.13	7.16	3.67	2.66	4.85	2.19	4.00	3.33	4.64	1.31
ϕ_2	0.25	0.23	0.11	0.42	0.30	0.27	0.20	0.36	0.15	0.25	0.21	0.29	0.08
ρ	0.0	0.10	-0.44	0.52	0.96	0.05	-0.13	0.22	0.35	0.05	-0.05	0.17	0.22
β_{01}	30	30.21	29.75	30.68	0.93	29.81	29.60	30.01	0.42	30.07	29.95	30.18	0.24
β_{11}	13	12.91	12.70	13.12	0.42	13.10	13.00	13.19	0.19	12.98	12.92	13.03	0.11
β_{02}	2.5	1.42	-3.52	6.00	9.52	1.96	0.29	3.64	3.36	2.59	1.47	3.73	2.26
β_{12}	1.5	1.63	0.75	2.60	1.85	1.59	1.27	1.92	0.65	1.49	1.27	1.71	0.43
ϕ_1	4	5.86	2.81	11.26	8.45	4.44	3.38	5.76	2.38	4.10	3.44	4.77	1.33
ϕ_2	0.25	0.22	0.11	0.40	0.29	0.26	0.20	0.34	0.14	0.25	0.21	0.28	0.07
ρ	0.5	0.35	0.05	0.66	0.61	0.51	0.35	0.66	0.31	0.51	0.41	0.58	0.17
β_{01}	30	29.99	29.64	30.32	0.68	30.00	29.88	30.13	0.25	30.00	29.92	30.08	0.16
β_{11}	13	12.96	12.80	13.10	0.31	13.01	12.96	13.05	0.09	13.02	12.99	13.04	0.05
β_{02}	2.5	2.99	-1.52	7.10	8.61	2.94	2.04	3.87	1.83	2.55	2.04	3.12	1.07
β_{12}	1.5	1.36	0.59	2.22	1.63	1.42	1.25	1.59	0.33	1.51	1.42	1.60	0.19
ϕ_1	4	3.14	1.60	5.64	4.05	4.46	3.39	6.05	2.66	3.67	3.12	4.24	1.12
ϕ_2	0.25	0.21	0.10	0.40	0.29	0.24	0.18	0.33	0.15	0.22	0.19	0.26	0.07
ρ	0.9	0.85	0.62	0.94	0.33	0.88	0.82	0.92	0.10	0.90	0.88	0.92	0.04

presented a fast convergence for the true values in the majority of cases.

3.2 Application to real data set 1 - Mushroom species

The data set presented in Table 7 gives the values of the pileus cap width (Y), stipe length (T_1) and stipe thickness (T_2) for 23 mushroom species (Billard and Diday, 2006). Our aim is to predict the interval values of the dependent variable Y in terms of the explanatory variables T_j ($j = 1, 2$). We assume the random component $\mathbf{Y} = (Y_1, Y_2)$ structured in terms of the midpoint Y^m and the half-range Y^r of the interval-valued variable Y, respectively, and consider the two Bayesian linear regression models (BLRM1 and BLRM2) to fit a regression model for this interval-valued data set.

Tables 8 and 9 present the parameter estimates for the BLRM1 and BLRM2, respectively. The deviance information criterion (DIC) is an important Bayesian model selection measure. Based on this measure, the BLRM2 (DIC_{BLRM2}=73.3469) presented a best fit when compared with BLRM1 (DIC_{BLRM1}=80.5369). Corroborates with this statement the fact that the BLRM2

Table 6 - Confidence Intervals (95%) for the parameters estimates in CCRM method, according to correlation coefficient ρ and sample size. Comparative analysis with BLRM2

ρ	θ		Mean	Lower	Upper	Prec	Mean	Lower	Upper	Prec	Mean	Lower	Upper	Prec
				n=20				n=100				n=300		
-0.9	β_{01}	30	29.92	29.58	30.26	0.67	29.95	29.75	30.15	0.40	30.01	29.89	30.14	0.25
	β_{11}	13	13.02	12.88	13.15	0.27	13.05	12.96	13.14	0.18	13.01	12.95	13.06	0.10
	β_{02}	2.5	2.34	-1.17	5.85	7.02	2.52	0.26	4.78	4.53	2.45	1.33	3.58	2.25
	β_{12}	1.5	1.62	0.94	2.31	1.37	1.46	1.01	1.91	0.90	1.50	1.28	1.71	0.43
-0.5	β_{01}	30	29.88	29.50	30.26	0.76	30.00	29.74	30.28	0.54	29.96	29.83	30.09	0.26
	β_{11}	13	13.07	12.91	13.24	0.33	12.98	12.86	13.11	0.24	13.03	12.98	13.09	0.12
	β_{02}	2.5	2.71	-1.06	6.47	7.53	1.59	-0.22	3.40	3.62	1.99	0.81	3.17	2.36
	β_{12}	1.5	1.51	0.74	2.27	1.54	1.65	1.30	2.00	0.70	1.57	1.34	1.80	0.47
0.0	β_{01}	30	29.86	29.39	30.33	0.94	29.76	29.54	29.99	0.45	29.99	29.85	30.12	0.27
	β_{11}	13	13.09	12.86	13.31	0.44	13.09	12.99	13.19	0.19	13.02	12.96	13.08	0.12
	β_{02}	2.5	0.56	-5.49	6.61	12.10	2.36	0.20	4.53	4.33	2.70	1.37	4.03	2.66
	β_{12}	1.5	1.88	0.74	3.03	2.29	1.56	1.14	1.99	0.85	1.45	1.19	1.72	0.52
0.5	β_{01}	30	30.23	29.74	30.72	0.98	29.71	29.48	29.94	0.47	30.03	29.89	30.16	0.28
	β_{11}	13	12.90	12.67	13.13	0.46	13.14	13.04	13.25	0.21	12.99	12.93	13.06	0.13
	β_{02}	2.5	1.99	-2.98	6.96	9.94	1.33	-0.61	3.27	3.89	2.47	1.20	3.75	2.55
	β_{12}	1.5	1.52	0.55	2.50	1.94	1.72	1.34	2.10	0.75	1.51	1.26	1.76	0.50
0.9	β_{01}	30	30.11	29.64	30.57	0.93	30.02	29.81	30.22	0.41	29.95	29.82	30.09	0.27
	β_{11}	13	12.88	12.62	13.14	0.52	13.00	12.91	13.10	0.19	13.04	12.98	13.10	0.12
	β_{02}	2.5	3.00	-4.80	10.80	15.60	2.90	1.06	4.75	3.69	3.24	2.08	4.40	2.32
	β_{12}	1.5	1.36	-0.15	2.87	3.02	1.43	1.07	1.79	0.72	1.38	1.15	1.60	0.45

Table 7 - Mushroom interval-valued data set

Species	Y	T_1	T_2	Species	Y	T_1	T_2
1	[3.0-8.0]	[4.0-9.0]	[0.50-2.50]	13	[3.5-8.0]	[4.0-10.0]	[1.00-2.00]
2	[6.0-21.0]	[4.0-14.0]	[1.00-3.50]	14	[7.0-14.0]	[8.0-14.0]	[1.50-2.50]
3	[4.0-8.0]	[5.0-11.0]	[1.00-2.00]	15	[8.0-20.0]	[9.0-19.0]	[3.00-5.00]
4	[6.0-7.0]	[4.0-7.0]	[3.00-4.50]	16	[2.5-4.0]	[2.5-4.5]	[0.40-0.70]
5	[5.0-12.0]	[2.0-5.0]	[1.50-2.50]	17	[7.0-19.0]	[8.0-15.0]	[2.00-3.50]
6	[5.0-15.0]	[4.0-10.0]	[2.00-4.00]	18	[5.0-15.0]	[6.0-15.0]	[2.50-3.50]
7	[4.0-11.0]	[3.0-7.0]	[0.40-1.00]	19	[8.0-12.0]	[6.0-12.0]	[1.50-2.00]
8	[5.0-10.0]	[3.0-6.0]	[1.00-2.00]	20	[2.0-6.0]	[3.0-7.0]	[0.40-0.80]
9	[2.5-4.0]	[3.0-5.0]	[0.40-0.70]	21	[6.0-12.0]	[6.0-12.0]	[1.50-2.00]
10	[2.5-6.0]	[1.5-3.5]	[1.00-1.50]	22	[6.0-12.0]	[6.0-16.0]	[1.00-2.00]
11	[1.5-2.5]	[3.0-6.0]	[0.25-0.35]	23	[5.0-17.0]	[4.0-14.0]	[1.00-3.50]
12	[4.0-15.0]	[4.0-15.0]	[1.50-2.50]				

also presented an higher value for the correlation coefficient ρ. Moreover, the difference between ϕ_1 and ϕ_2 suggests that the variances for the response interval-valued variable $Y = [Y_1, Y_2]$ are not the same, being the BLRM2 a more appropriate approach.

3.3 Application to real data set 2 - Soccer teams

This data set gives the records of the weight (Y), height (T_1) and age (T_2) for 531 soccer players grouped in 20 teams of the French Football Professional Championship. We use the **BLRM1** and **BLRM2** to predict the dependent

Table 8 - Fitted BLRM1 for the mushroom interval-valued data set

Parameter	Mean	50%	2.5%	97.5%
β_{11}	1.2861	1.2822	-0.2611	2.9388
β_{12}	0.6194	0.6174	0.3873	0.8567
β_{13}	1.1624	1.1542	0.4697	1.8617
β_{21}	0.3573	0.3465	-0.9973	1.7216
β_{22}	0.7710	0.7742	0.2726	1.2673
β_{23}	1.2172	1.1918	-0.7898	3.4143
ϕ	0.4475	0.4394	0.2435	0.7056
ρ	0.5887	0.6080	0.2604	0.8270

Table 9 - Fitted BLRM2 for the mushroom interval-valued data set

Parameter	Mean	50%	2.5%	97.5%
β_{11}	1.3082	1.3110	-0.6859	3.2715
β_{12}	0.6321	0.6321	0.3514	0.9232
β_{13}	1.0998	1.1100	0.2537	1.9235
β_{21}	0.1764	0.1652	-0.9012	1.2331
β_{22}	0.7838	0.7868	0.4282	1.1334
β_{23}	1.4717	1.4833	-0.0642	2.9225
ϕ_1	0.3187	0.3067	0.1545	0.5381
ϕ_2	0.8281	0.8110	0.3866	1.4106
ρ	0.6404	0.6550	0.3381	0.8403

variable Y from the explanatory variables T_j ($j = 1, 2$). Table 10 gives the data which can be accessed in Lima Neto *et al.* (2011).

Tables 11 and 12 present the parameter estimates for the BLRM1 and BLRM2, respectively. Based on DIC measure, the BLRM1 (DIC_{BLRM1}=121.5029) presented a best fit when compared with BLRM2 (DIC_{BLRM2}=124.0437) for this interval data set. This real interval-valued data set illustrates the case where the BLRM1 is preferred in relation to BLRM2. Corroborates with this statement the very similar estimates between ϕ_1 and ϕ_2 in BLRM2 approach. Then, the inclusion of an additional parameter results in a loss of adjustment for the model, being the BLRM2 method inappropriate for this data set. Moreover, the BLRM1 presented a higher correlation coefficient when compared with BLRM2 method.

Conclusions

In this paper we proposed a Bayesian approach to estimate the parameters of regression models for interval variables. As an interval is an infinity list of values, it is difficult to take into account the role information inside it. Despite the loss of information, we represented an interval-valued variable Y as a two-dimensional feature vector $\mathbf{y} = (y_1, y_2)$, where the variables Y_1 and Y_2 are one-dimensional random variables represented by their midpoint and range, respectively. We considered a Bayesian approach for the parameters estimate of two linear regression

Table 10 - Soccer interval data set

Team	Y	T_1	T_2	Team	Y	T_1	T_2
A	[58-85]	[164-192]	[21-35]	K	[62-86]	[164-191]	[18-34]
B	[67-84]	[171-190]	[20-30]	L	[62-80]	[168-189]	[19-35]
C	[65-88]	[170-186]	[18-36]	M	[63-85]	[167-190]	[18-31]
D	[60-83]	[162-188]	[19-31]	N	[65-95]	[168-196]	[20-35]
E	[60-84]	[170-189]	[18-34]	O	[63-83]	[170-187]	[18-35]
F	[67-83]	[173-190]	[18-36]	P	[60-87]	[170-197]	[18-37]
G	[69-90]	[176-193]	[19-34]	Q	[67-85]	[168-190]	[18-32]
H	[65-85]	[170-193]	[19-31]	R	[62-83]	[169-192]	[18-35]
I	[63-84]	[168-188]	[18-34]	S	[63-84]	[172-192]	[18-33]
J	[58-88]	[167-197]	[19-35]	T	[63-85]	[169-194]	[20-34]

Table 11 - Fitted BLRM1 for the soccer interval-valued data set

Parameter	Mean	50%	2,5%	97,5%
β_{11}	-2,94	-2,83	-21,94	16,07
β_{12}	0,44	0,43	0,28	0,58
β_{13}	-0,40	-0,38	-1,44	0,72
β_{21}	-2,88	-3,02	-21,99	15,29
β_{22}	0,40	0,41	0,26	0,55
β_{23}	0,33	0,32	-0,34	0,91
ϕ	0,14	0,14	0,08	0,22
ρ	0,39	0,37	-0,05	0,70

Table 12 - Fitted BLRM2 for the soccer interval-valued data set

Parameter	Mean	50%	2.5%	97.5%
β_{11}	-3.32	-3.22	-21.70	15.47
β_{12}	0.43	0.44	0.29	0.58
β_{13}	-0.37	-0.39	-1.47	0.68
β_{21}	-2.72	-2.79	-21.35	16.56
β_{22}	0.40	0.40	0.24	0.55
β_{23}	0.34	0.36	-0.35	1.03
ϕ_1	0.15	0.15	0.07	0.28
ϕ_2	0.13	0.12	0.06	0.22
ρ	0.35	0.40	-0.05	0.71

models, so-called BLRM1 and BLRM2, which can be useful in statistical analysis of interval-valued data. Particularly, the BLRM2 considers different variances for the components Y_1 and Y_2 of the two-dimensional feature vector Y.

The results presented in the numerical study's section showed the effectiveness in the estimation of true parameters considering different sample sizes. The credible (BLRM) interval showed a percentage of coverage of the true parameter better than the confidence (CCRM) interval. This result demonstrated that the BLRM methods have outperformed the CCRM method in the majority of the configurations. We also observed that the credible intervals (BLRM) have presented more accuracy when compared with the confidence intervals (CCRM). Moreover, the credible intervals covered the true parameters for all simulation scenarios, in contrast with CCRM approach. The BLRM also allows interval estimation for the parameters ϕ and ρ. We also verified the convergence of the MCMC chains obtained from the parameter's estimate algorithm. Applications to real interval-valued data sets illustrated the importance of the BLRM1 and BLRM2 methods in practical situations.

MORALES, F.C; LIMA NETO, E.A. Uma abordagem Bayesiana para modelagem de variáveis tipo-intervalo. *Rev. Bras. Biom.,* São Paulo, v.32, n.3, p.360-378, 2014.

- RESUMO: Neste artigo propomos duas abordagens bayesianas para estimar os parâmetros de um modelo de regressão consideriando variáveis tipo-intervalo como variávies resposta e explicativas. A primeira abordagem considera uma estrutura de covariância mais simples, enquanto que a segunda abordagem considera uma estrutura de covariância mais geral. A distribuição a posteriori dos parâmetros foi aproximada considerando o método MCMC. O estudo de simulação apresentado sugere a eficiência do esquema amostral em recuperar os verdadeiros valores dos parâmetros e também indicam a convergência do algoritmo de estimação dos parâmetros. As novas abordagens são aplicads a dados reais e suas performances comparadas.

- PALAVRAS-CHAVE: Variáveis intervalares; regressão; MCMC; abordagem Bayesiana.

References

BILLARD, L.; DIDAY, E. Regression Analysis for Interval-Valued Data, In: DATA ANALYSIS, CLASSIFICATION AND RELATED METHODS: PROCEEDING OF THE SEVENTH CONFERENCE OF THE INTERNATIONAL FEDERATION OF CLASSIFICATION SOCIETIES (IFCS'00), 2000, Belgium. Springer-Verlag, 2000. p.369-374.

BILLARD, L.; DIDAY, E. *Symbolic Data Analysis: Conceptual Statistics and Data Mining*. John Wiley: New York, 2006. 330p.

BLANCO-FERNANDEZ, A.; CORRAL, N.; GONZÁLEZ-RODRUÍGUEZ, G. Estimation of a flexible simple linear regression model for interval data based on set arithmetic. *Computational Statistics and Data Analysis*, v.55, p.2568-2578, 2011.

BOCK, H. H.; DIDAY, E. *Analysis of Symbolic Data, Exploratory methods for extracting statistical information from complex data*. Springer-Verlag: Heidelberg, 2000. 425p.

BRITO, P.; SILVA, A. P. D. Modeling Interval Data with Normal and Skew-Normal Distributions. *Journal of Applied Statistics*, v.39, p.3-20, 2012.

COPPI, R.; D'URSO P.; GIORDANI P.; SANTORO, A. Least squares estimation of a linear regression model with LR fuzzy response. *Computational Statistics and Data Analysis*, v.51, p.267-286, 2006.

COWLESS, M. K. BEST, N., VINES, K. AND PLUMMER, M. *Convergence and Diagnostics Output Analysis software for Gibbs sampling output*, Cambridge: MRC Biostatistics Unit, 1997.

D'URSO, P; GASTALDI, T. A least-square approach to fuzzy linear regression analysis. *Computational Statistics and Data Analysis*, v.34, p.427-440, 2000.

FAGUNDES, R. A. A.; DE SOUZA, R. M. C. R.; CYSNEIROS, F. J. A. Robust Regression with Application to Symbolic Interval Data. *Engineering Applications of Artificial Intelligence*, v.26, p.564-569, 2013.

FERRARO, M. B.; COPPI, R.; GONZÁLEZ-RODRUÍGUEZ, G.; COLUBI, A. A linear regression model for imprecise response. *International Journal of Approximate Reasoning*, v.51, p.759-770, 2010.

GAMERMAN, D. Sampling from the posterior distribution in generalized linear mixed models. *Statistics and Computing*, v.7, p.57-68, 1997.

GELMAN, A. Prior distributions for variance parameters in hierarchical models. *Bayesian Analysis*, v.1, p.515-534, 2006.

GELMAN, A. and RUBIN, D. B Inference from iterative simulation using multiple sequence. *Statistical Science*, v.7, p.457-511, 1992.

GIL, M. A.; GONZÁLEZ-RODRUÍGUEZ, G.; COLUBI, A.; MONTENEGRO, M. Testing linear independence in linear models with interval-valued data. *Computational Statistics and Data Analysis*, v.51, p.3002-3015, 2007.

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

LAURO, N. C.; PALUMBO, F. Principal component analysis of interval data: A symbolic data analysis approach. *Computational Statistics*, v.15, p.73-87, 2000.

LIMA NETO, E. A.; DE CARVALHO, F. A. T. Centre and range method to fitting a linear regression model on symbolic interval data. *Computational Statistics and Data Analysis*, v.52, p.1500-1515, 2008.

LIMA NETO, E. A.; DE CARVALHO, F. A. T. Constrained linear regression models for symbolic interval-valued variables. *Computational Statistics and Data Analysis*, v.54, p.333-347, 2010.

LIMA NETO, E. A.; CORDEIRO, G. M.; DE CARVALHO, F. A. T. Bivariate symbolic regression models for interval-valued variables. *Journal of Statistical Computation and Simulation*, v.81, p.1727-1744, 2011.

MAIA, A. L. S.; DE CARVALHO, F. A. T. Holt's exponential smoothing and neural network models for forecasting interval-valued time series. *International Journal of Forecasting*, v.27, p.740-759, 2011.

MAIA, A. L. S.; DE CARVALHO, F. A. T. Fitting a least absolute deviation regression model on interval-valued data. In: 6th BRAZILIAN SYMPOSIUM ON NEURAL NETWORKS, ADVANCES IN ARTIFICIAL INTELLIGENCE, 2008. p.3742-3747.

QUEIROZ, D. F. C.; DE SOUZA, R. M. C. R.; CYSNEIROS, F. J. A. Logistic Regression-Based Pattern Classifiers for Symbolic Interval Data. *Pattern Analysis and Applications*, v.14, p.273-282, 2011.

XU, W. *Symbolic Data Analysis: Interval-valued data regression*. 148p. Ph.D. Thesis, University of Georgia, Athens, 2010.

Received in 21.05.2014.

Approved after revised in 15.10.2014.