Bayesian Beta Regression with the Bayesianbetareg R-Package

Edilberto Cepeda-Cuervo[a], Daniel Jaimes[b], Margarita Marín[c], Javier Rojas[d]

Departamento de Estadística, Facultad de Ciencias, Universidad Nacional de Colombia, Bogotá, Colombia[1]

Resumen

In this paper we summarize the main points of beta regression models under Bayesian perspective, including a presentation of the Bayesianbetareg R-package, used to fit the beta regression models under a Bayesian approach. Finally, beta regression models are fitted to a reading score database using, respectively, the Bayesianbetareg and betareg R-Packages for Bayesian and classic perspectives.

***Palabras clave*:** Beta regression, R-Package, Bayesian method, MCMC.

1. Introduction

In this paper we analyze situations where the variable of interest can be assumed to have a beta distribution. The beta $B(p, q)$ distribution function, can be re-parameterized as a function of the mean and dispersion parameters, μ and $\phi = p + q$, respectively. With this this re-parameterization, the beta regression models with regression structures in both mean and dispersion parameters were initially proposed by Cepeda-Cuervo (2001), p. 63, in the framework of the regression modeling in the biparametric exponential family of distributions. To fit this class of models, Cepeda-Cuervo (2001) proposed a Bayesian method, as a generalization of the Bayesian method proposed by Gamerman (1997), to fit generalized linear models.

Further papers have been published investigating beta regression models in recent years. Ferrari & Cribari-Neto (2004) proposed classic beta regression models, assuming that the dispersion parameter is constant through the range of the explanatory variables; Smithson & Verkuilen (2006) proposed mean and precision beta regression models under a classic method; Simas et al. (2010) proposed a generalization of beta regression models, including nonlinear regression structures in the mean and precision parameters, also proposed by Cuervo & Achcar (2010) in the context of double generalized nonlinear models; Gamerman & Cepeda-Cuervo (2013) proposed spatial beta regression models, with applications to land concentration and postnatal period screening analysis; and Rocha & Cribari-Neto (2009) proposed a beta autoregressive moving average process, including exogenous variables in the dynamics. Mixed beta regression models were proposed in Figueroa-Zuniga et al. (2013), under the Bayesian approach, and were subsequently extended by Bonat & Zeviani (2013) under a classic method. In the current literature, this regression has been applied in multiple fields of knowledge, to model multiple variables such as: ischemic stroke lesion volume (Swearingen et al. 2011), Gini index (Atkinson 1970), distance between body parts

[a]. E-mail: ecepedac@unal.edu.co
[b]
[c]
[d]

[1] The R-code and guide document were prepared by Daniel Jaimes, Margarita Marín and Javier Rojas, under the direction of the Dr. Edilberto Cepeda-Cuervo.

(Branscum et al. 2007), reading accuracy (Cribari-Neto & Zeileis 2010), election results (Chan 2006), among others. The present paper presents a new way to find consistent estimations for these inference problems.

In the Bayesian method proposed by Cepeda-Cuervo (2001) to fit the beta regression models, samples of the posterior distribution are obtained from the posterior conditional distributions of two blocks of parameters, one for the mean parameters and other for the dispersion parameter, assuming normal prior distributions for the mean and dispersion regression parameters (Cepeda & Gamerman 2005). These conditional posterior distributions are unknown and analytically intractable. Thus, two normal transitions kernel are built to apply a Metropolis Hastings algorithm to obtain the posterior samples.

With the purpose of making method more used-friendly along with the explanation of the key features of the models, we introduce the use of the Bayesianbetareg R-package, an R-code developed by us that contains all the necessary algorithms and customizable options for simulations - for the estimation of the beta regression model where both, mean and precision parameters are modeled. With the estimated parameters, standard deviations and credibility intervals, we present the residual analysis, the criteria for comparison (AIC-BIC) and the diagnostic for this model as useful alternative outputs for the researcher.

This paper is divided into five sections, including this introduction. Section 2 presents the beta distribution function and the beta regression models proposed in Cepeda-Cuervo (2001) and Cepeda & Gamerman (2005). Section 3 present the Bayeian method and applies it to fit beta regression models. Section 4 presentes the Bayesian betareg R-package. Section 5 presents the results of the Gini and per capita GDP databases, using the proposed R-package. Finally, Section 6 contains our concluding remarks.

2. Bayesian beta regression models

2.1. Beta distribution

A random variable Y follows a beta distribution if its probability density function is given by

$$f(y|p,q) = \frac{\Gamma(p+q)}{\Gamma(p)\Gamma(q)} y^{p-1}(1-y)^{1-q}\mathbf{I}_{(0,1)}(y), \tag{1}$$

where $p, q > 0$ and $\Gamma(.)$ denotes the gamma function. Additional, $I(0;1)(y) = 1$, if $y \in (0;1)$, and zero otherwise. With the re-parameterization of the beta distribution as a function of the mean, $\mu = E(Y)$, and the precision parameters, $\phi = p + q$, as proposed in Jorgensen (1997) or Cepeda-Cuervo (2001), the beta density function can be written as

$$f(y|\phi,\mu) = \frac{\Gamma(\phi)}{\Gamma(\mu\phi)\Gamma((1-\mu)\phi)} y^{\mu\phi-1}(1-y)^{(1-\mu)\phi-1}\mathbf{I}_{0,1}(y) \tag{2}$$

where, $\mu = p/(p+q)$ (Ferrari & Cribari-Neto 2004). In this re-parameterization $p = \mu\phi$, $q = \phi(1-\mu)$ and

$$\sigma^2 = \frac{\mu(1-\mu)}{\phi+1}. \tag{3}$$

From this re-parameterization of the beta distribution, the joint mean and precision beta regression were proposed in Cepeda-Cuervo (2001), as presented in the next section.

2.2. Beta regression models

With the re-parameterization of the beta distribution as a function of μ and ϕ, in this section the joint mean and precision beta regression models are defined as in Cepeda-Cuervo (2001). Under a general framework, a random sample $Y_i \sim Beta(p_i, q_i)$, $i = 1, 2, \ldots, n$, is assumed, where the mean and precision parameters are modeled respectively as functions of explanatory variables by

$$g(\mu_i) = \mathbf{x}_i'\boldsymbol{\beta} \quad \text{and} \tag{4}$$

$$h(\phi_i) = \mathbf{z}_i'\boldsymbol{\gamma} \tag{5}$$

where g and h are appropriate real functions, $\boldsymbol{\beta} = (\beta_0, ..., \beta_k)$ and $\boldsymbol{\gamma} = (\gamma_0, ..., \gamma_k)$ are respectively the mean and precision parameter vectors, and $\mathbf{x}_i$ and $\mathbf{z}_i$ are respectively the mean and precision explanatory variables for the i-th observation. In the proposal developed by Cepeda-Cuervo (2001), g is the *logit* function and h the exponential function, but other link function can be assumed. Other link functions are: probit $g(\mu) = \phi^{-1}(\mu)$, where $\phi(.)$ is the standard normal distribution function; complementary loglog $g(\mu) = -log(-log(1-\mu))$; log-log $g(\mu) = log(-log(\mu))$; and Cauchy $g(\mu) = tan(\pi(\mu - 0.5))$. See Cribari-Neto & Zeileis (2010).

The beta regression model is naturally heteroscedastic. However in the joint mean and dispersion beta regression models, the variance results in complex and not easily interpretable expressions. In particular, if g is the logit function and h the exponential function, the variance is given by (6), which it is not easy to interpret in a practical framework.

$$Var(Y_i) = \frac{2 + \exp(\mathbf{x}_i\boldsymbol{\beta}) + \exp(-\mathbf{x}_i\boldsymbol{\beta})}{\exp(\mathbf{z}_i\boldsymbol{\gamma})}. \tag{6}$$

Sometimes it is possible to assume joint mean and variance models, sampling the regression parameters under a restricted subspace.

3. Bayesian method to fit beta regression models

In this section, we present the Bayesian method and the MCMC algorithm proposed in Cepeda-Cuervo (2001) and Cepeda et al. (2005), in the framework of double generalized regression model used to fit the beta regression model. As in these works, to implement a Bayesian approach to estimate the parameters of the joint beta regression model, we need to specify a prior distribution for the parameters. Thus, if $L(\boldsymbol{\Theta}|\text{ data})$ denotes the likelihood function and $p(\boldsymbol{\Theta})$ the joint prior distribution, where $\boldsymbol{\Theta} = (\boldsymbol{\beta}, \boldsymbol{\gamma})$, the posterior distribution is given by $\pi(\boldsymbol{\Theta}|\text{ data}) \propto L(\boldsymbol{\Theta})p(\boldsymbol{\Theta})$. However, given that when assuming normal prior distributions for the parameters, the posterior distribution $\pi(\boldsymbol{\Theta}|\text{ data})$ is analytically intractable, Cepeda-Cuervo (2001) proposed to get samples of $\boldsymbol{\Theta}$ using an alternative iterative algorithm, by sampling $\boldsymbol{\beta}$ and $\boldsymbol{\gamma}$ from the posterior conditional distributions $\pi(\boldsymbol{\beta}|\boldsymbol{\gamma}, \text{ data})$ and $\pi(\boldsymbol{\gamma}|\boldsymbol{\beta}, \text{ data})$, for which it is necessary to build normal transition kernels q_1 and q_2, given that also these conditional distributions are analytically intractable.

To build the kernel transition functions we need to define working observation variables to approximate $h(\mu_i)$ and $g(\phi_i)$ around the current values of μ and ϕ, respectively. They are defined as first order Taylor approximations of the real functions $h(t_1)$ and $g(t_2)$, where t_1 and t_2 are appropriate random variables such that $E(t_1) = \mu$ and $E(t_2) = \phi$. Thus, given that $E(t_1) = \mu$ for $t_1 = Y$, if the mean model is given by (4), the working observational variable is defined by

$$\tilde{y_i} = x_i^{'}\boldsymbol{\beta}^{(c)} + \frac{y_i - \mu_i^{(c)}}{(\mu_i^{(c)})(1-\mu_i^{(c)})},\ i = 1, 2, ..., n, \tag{7}$$

where $\mu^{(c)}$ and $\boldsymbol{\beta}^{(c)}$ are the current values of μ and $\boldsymbol{\beta}$. Thus, assuming that $\tilde{y}_i$, $i = 1, \ldots, n$, has a normal distribution and assuming conditional normal prior distribution $\boldsymbol{\beta}|\boldsymbol{\gamma} \sim N(b, B)$, the kernel transition function q_1 is given by the posterior distribution obtained from the combination of the prior distribution with the working observation model $\tilde{y_i} \sim N(x_i^{'}\boldsymbol{\beta}, \tilde{\sigma}_i^2)$, where $\tilde{\sigma}_i^2 = \text{Var}(\tilde{y}_i)$. That is, by

$$q_1(\boldsymbol{\beta}|\boldsymbol{\beta}^{(c)}, \boldsymbol{\gamma}^{(c)}) = N(\mathbf{b}^*\ , \mathbf{B}^*), \tag{8}$$

where

$$\begin{aligned}\mathbf{b}^* &= \mathbf{B}^*(\mathbf{B}^{-1}\mathbf{b} + \mathbf{X}^{'}\boldsymbol{\Sigma}^{-1}\tilde{Y}) \\ B^* &= (\mathbf{B}^{-1} + \mathbf{X}^{'}\boldsymbol{\Sigma}^{-1}\mathbf{X})^{-1}\end{aligned}$$

and where $\mathbf{\Sigma}$ is a diagonal matrix with diagonal entries $\tilde{\sigma}_i^2$, $i = 1, 2, ..., n$, (see Cepeda and Gamerman, 2001, 2005). Thus, the values of $\boldsymbol{\beta}$ from the posterior distribution sample of $\pi(\boldsymbol{\beta}, \boldsymbol{\gamma})$ are proposed from the transition kernel defined in equation (8).

As the full conditional distribution $\pi(\boldsymbol{\gamma}|\boldsymbol{\beta})$ is also analytically intractable and it is not easy to generate samples from it, we need to build a kernel transition q_2 to propose the values of $\boldsymbol{\gamma}$ from the posterior distribution of $\mathbf{\Theta}$. Given that $E(\boldsymbol{t_i}) = \phi_i$ for $\boldsymbol{t_i} = \frac{(p_i+q_i)^2}{p_i}\boldsymbol{Y}_i$, if the dispersion model is given by (5), the working observational variable (9) is obtained from the first order Taylor approximation around the current value of ϕ_i, given by the current values of the dispersion regression parameters $\boldsymbol{\gamma}^{(c)}$, in turn given by

$$\tilde{y}_i = \boldsymbol{z}_i^{'}\boldsymbol{\gamma}^{(c)} + \frac{y_i}{\mu_i} - 1, \ i = 1, 2, ..., n. \tag{9}$$

Thus, assuming that the observational working variable (9) has a normal distribution and since that the conditional prior distribution is given by $\boldsymbol{\gamma}|\boldsymbol{\beta} \sim N(\mathbf{g}, \mathbf{G})$, the normal transition kernel q_2 is given by the posterior distribution obtained from the combination of the prior distribution with the working observational model $\tilde{y}_i \sim N(z_i^{'}\boldsymbol{\gamma}, \tilde{\sigma}^2)$, where $\tilde{\sigma}_i^2 = \mathrm{Var}(\tilde{y}_i)$. That is,

$$q_2(\boldsymbol{\gamma}|\boldsymbol{\gamma}^{(c)}, \boldsymbol{\beta}^{(c)}) = N(\mathbf{g}^*\ , \mathbf{G}^*), \tag{10}$$

where

$$\begin{aligned} \mathbf{g}^* &= \mathbf{G}^*(\mathbf{G}^{-1}\mathbf{g} + \mathbf{Z}^{'}\Psi^{-1}\tilde{Y}), \\ \mathbf{G}^* &= (\mathbf{G}^{-1} + \mathbf{Z}^{'}\Psi^{-1}Z)^{-1}. \end{aligned}$$

and Ψ is a diagonal matrix with entries $\tilde{\sigma}_i^2$ for $i = 1, 2, ..., n$. Samples of $\boldsymbol{\gamma}$ from the posterior distribution $\pi(\boldsymbol{\beta}, \boldsymbol{\gamma})$, are obtained from the transition kernel function q_2.

With the transition kernels given by (8) and (10), the components $\boldsymbol{\beta}$ and $\boldsymbol{\gamma}$ of $(\boldsymbol{\beta}, \boldsymbol{\gamma})^{'}$ are updated as follows:

1. Begin the chain interaction counter at $j = 1$ and set initial values $(\boldsymbol{\beta}_0, \boldsymbol{\gamma}_0)$ to $(\boldsymbol{\beta}, \boldsymbol{\gamma})^{'}$.
2. Move the vector $\boldsymbol{\beta}$ to a new value ψ generated from the proposed density $q_1(\boldsymbol{\beta}^{(j-1)}, .)$.
3. Calculate the acceptance probability of movement, $\alpha(\boldsymbol{\beta}^{(j-1)}, \psi)$. If the movement is accepted, then $\boldsymbol{\beta}^{(j)} = \psi$. If it is not accepted, then $\boldsymbol{\beta}^{(j)} = \boldsymbol{\beta}^{(j-1)}$.
4. Move the vector $\boldsymbol{\gamma}$ to a new value ψ, generated from the proposed density $q_2(\boldsymbol{\gamma}^{j-1}, .)$.
5. Calculate the acceptance probability of movement, $\alpha(\boldsymbol{\gamma}^{(j-1)}, \psi)$. If the movement is accepted, then $\boldsymbol{\gamma}^{(j)} = \psi$. If it is not accepted, then $\boldsymbol{\gamma}^{(j)} = \boldsymbol{\gamma}^{(j-1)}$.
6. Finally, change the counter from j to $j + 1$ and go to 2 until convergence is reached.

As an assessment tool for the fit of the model, some types of residual are included. One is Pearson residuals, defined by

$$r_{P_i} = \frac{Y_i - \hat{\mu}_i}{\sqrt{\hat{V}ar(Y_i)}} \tag{11}$$

where $\hat{v}ar(Y_i) = \hat{\mu}_i(1 - \hat{\mu}_i)/(1 + \hat{\phi}_i)$, $\hat{\mu}_i = g^{-1}(\mathbf{x}_i'\hat{\boldsymbol{\beta}})$ and $\hat{\phi}_i = h^{-1}(\mathbf{z}_i'\hat{\boldsymbol{\gamma}})$. Another is a standardized weighted residual, defined as

$$r_i^w = \frac{y_i^* - \tilde{\mu}_i^*}{\sqrt{v_i(1 - h_{ii})}} \tag{12}$$

where

$$y_i^* = log(\frac{y_i}{1-y_i}) \tag{13}$$

$$\tilde{\mu}_i^* = \psi(\mu_i\phi) - \psi((1-\mu_i)\phi) \tag{14}$$

$$v_i = \psi'(\mu_i\phi) + \psi'((1-\mu_i)\phi) \tag{15}$$

and h_{ii} is the i-th diagonal element of the matrix $H = \tilde{W}^{1/2}X(X'\tilde{W}X)X'\tilde{W}^{1/2}$, where W is a diagonal matrix with elements $w_i = \phi v_i[1/\{g'(\mu_i\}^2]$. Among interesting properties of residual (12), is the fact that its distribution can be well approximated by a standard normal distribution when compared to the Pearson standardized residuals. In addition, given that it incorporates the observations leverage, it can more clearly identify influential observations for the parameters related to the linear predictor (See (Espinheira et al. 2008)).

4. Bayesianbetareg R-package

The Bayesianbetareg R-package has the computational implementation of the Bayesian method defined in Section 3. The main function of this package is the **Bayestianbetareg()**, which allows the user to calculate the mean and dispersion regression parameters in a beta regression model under Bayesian perspective. The general formula for this function is

$$Bayesianbetareg(Y, X, Z, nsim, bpri, Bpri, gpri, G_pri, burn, jump, graph1 = T, graph2 = T), \tag{16}$$

where Y is a vector whit the dependent variable, X is a matrix of the explanatory variables of the mean, Z is a matrix of the explanatory variables of the dispersion, *nsim* is the number of iterations, *bpri* is the mean of the prior distribution of $\boldsymbol{\beta}$, *Bpri* is the variance-covariance matrix of the prior distribution of $\boldsymbol{\beta}$, *gpri* is the mean of the prior distribution of γ, *Gpri* is the variance-covariance matrix of the prior distribution of $\boldsymbol{\gamma}$, *burn* is a number that indicates the proportion of iterations for burn-in at the beginning of the chains, *jump* is a number that indicate the distance (number of iterations) between samples and *graph* is a indicator to present or not the graphic representations of the chains.

The returned fitted-model object of the Byesianbetareg class provides to the user the regression parameter estimates, $\hat{\boldsymbol{\beta}}$ and $\hat{\boldsymbol{\gamma}}$, and their standard deviations. It also provides the fitted values of Y, the residuals, variance and a matrix with the posterior samples.

The Bayesianbetareg R-package also has nine other functions which allow the user, among other things, to obtain AIC, BIC and deviance criterion values, plots of four types of residuals and residual diagnostic plots for beta regression models. The functions of this package are described in the Table 1.

5. An application of the Bayesianbetareg code

Here we carry out a brief example of the capabilities of this package using the dyslexia data set presented in Smithson & Verkuilen (2006), in which the response variable represents the scores obtained by 44 children in a reading accuracy test. The explanatory variables correspond to the dyslexia or lack of status (coded as 0 or 1 respectively), the standardized non-verbal IQ and their interaction. This database is available in the **betareg** package.

We plot reading score versus dyslexia (left plot), reading score versus non-verbal intelligence (center plot) and the reading score versus interaction between dyslexia and reading score (right plot), to explore the relationship between the variable of interest and regression variables.

Tabla 1: Bayesianbetareg functions

Function	Description
Bayesianbetareg	estimates the media and dispersion regression parameters
betaresiduals	calculates the Deviance, Pearson and Standarized Pearson residuals
criteria	calculates the AIC, BIC and Deviance information criteria for comparison of Bayesian beta regression models
diagnostics	allows to plot different residuals diagnostics (like the cook distance and the leverage)
dpostb	gives a value of the mean regression parameters, $\boldsymbol{\beta}$, from the posterior distribution
dpostg	gives a value for the dispersion parameter, $\boldsymbol{\gamma}$, from the posterior distribution
gammakernel	calculates the likelihood function at a value of $\boldsymbol{\gamma}$, given the current values of $\boldsymbol{\beta}$
gammaproposal	provides parameter values from the proposal of $\boldsymbol{\gamma}$ parameter vector
mukernel	calculates likelihood function at a $\boldsymbol{\beta}$, given the current value of $\boldsymbol{\gamma}$
muproposal	provides parameter values from the proposal of $\boldsymbol{\beta}$ parameter vector
plotresiduals	allows to plot the residuals calculated whit the betaresiduals function
print.Bayesianbetareg	prints the estimates coefficients and the credibility intervals of a bayesian beta regression
print.summary. Bayesianbetareg	prints the summary of a bayesian beta regression
summary.Bayesianbetareg	is the standard regression output (coefficient estimates, standard errors, criterions); returns an object of class "summary.bayesianbetareg" containing the relevant summary statistics (which has a print() method)

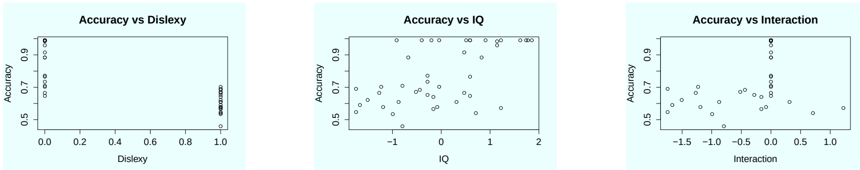


Figure 1: Plot (reading score vs variables)

5.1. Beta regression models with constant precision

To illustrate the use of the **Bayesianbetareg** package, we first fit a regression assuming a constant dispersion parameter, assuming the reparameterization $\phi = \exp(\gamma_0)$.

```
> library(betareg)
> data(ReadingSkills)
> Y <- as.matrix(ReadingSkills[,1])
> n <- length(Y)
> X1 <- as.matrix(ReadingSkills[,2])
> for(i in 1:length(X1)){
+   X1 <- replace(X1,X1=="yes",1)
+   X1 <- replace(X1,X1=="no",0)
+ }
> X0 <- rep(1, times=n)
> X1 <- as.numeric(X1)
> X2 <- as.matrix(ReadingSkills[,3])
> X3 <- X1*X2
> X <- cbind(X0,X1,X2,X3)
> Z0 <-  X0
> Z <- cbind(X0,X1,X2)
>
> burn <- 0.2
> jump <- 30
> nsim <- 100000
>
> b_pri <- c(0,0,0,0)
> B_pri <- diag(100,nrow=ncol(X),ncol=ncol(X))
> g_pri <- c(0)
> G_pri <- diag(10,nrow=ncol(Z0),ncol=ncol(Z0))
>
> reading_skills<- Bayesianbetareg (Y,X,Z0,nsim,b_pri,B_pri,g_pri,G_pri,
burn,jump,graph1=T, graph2=T )
```

In the results, the credibility intervals show that the estimates have a 95 % probability of being different from zero.

```
> summary(reading_skills)

################################################################
###                  Bayesian Beta Regression              ###
################################################################
```

```
 Call:
Bayesianbetareg.default(Y = Y, X = X, Z = Z0, nsim = nsim, b_pri = b_pri,
    B_pri = B_pri, g_pri = g_pri, G_pri = G_pri, burn = burn,
    jump = jump, graph1 = T, graph2=T)

         Estimate Est.Error L.CredIntv U.CredIntv
beta.X0    2.0863    0.1276     1.8338      2.335
beta.X1   -1.7041    0.1709    -2.0309     -1.358
beta.X2    0.5202    0.1463     0.2321      0.799
beta.X3   -0.5786    0.1824    -0.9396     -0.219
gamma.X0   2.1975    0.1533     1.8905      2.486

 Deviance:
[1] 180.3878

 AIC:
[1] 188.3878

 BIC:
[1] 195.5246
```

Using the Coda package, we tests the convergence of the chains. In all of the cases, the chains pass the three test of convergence (the geeks diagnostic, rafters diagnostic and Heidelberger and Welch diagnostic), showing the performance of the Bayesian method in fitting the model to the data originally presented in Heidelberger & Welch (1981)

```
library (coda)

> geweke.diag(reading_skills$beta.mcmc, frac1=0.1,frac2=0.5)

Fraction in 1st window = 0.1
Fraction in 2nd window = 0.5

Series 1 Series 2 Series 3 Series 4
 -0.1890   1.8223   0.2704   1.0401

>
> raftery.diag(reading_skills$beta.mcmc, q = 0.025, r = 0.005, s = 0.95)

Quantile (q) = 0.025
Accuracy (r) = +/- 0.005
Probability (s) = 0.95

          Burn-in  Total Lower bound  Dependence
          (M)      (N)   (Nmin)       factor (I)
 Series X0 14       14554 3746         3.89
 Series X1 21       22911 3746         6.12
 Series X2 17       18600 3746         4.97
 Series X3 13       14382 3746         3.84
```

```
>
> heidel.diag(reading_skills$beta.mcmc)

         Stationarity start     p-value
         test         iteration
Series X0 passed       1         0.878
Series X1 passed       1         0.571
Series X2 passed       1         0.242
Series X3 passed       1         0.412

         Halfwidth Mean   Halfwidth
         test
Series X0 passed     2.084 0.00147
Series X1 passed    -1.705 0.00250
Series X2 passed     0.522 0.00249
Series X3 passed    -0.583 0.00314
>
> geweke.diag(reading_skills$gamma.mcmc, frac1=0.1,frac2=0.5)

Fraction in 1st window = 0.1
Fraction in 2nd window = 0.5

Series 1
 -0.3239

>
> raftery.diag(reading_skills$gamma.mcmc, q = 0.025, r = 0.005, s = 0.95)

Quantile (q) = 0.025
Accuracy (r) = +/- 0.005
Probability (s) = 0.95

          Burn-in  Total  Lower bound  Dependence
          (M)      (N)    (Nmin)       factor (I)
 Series Z0 120      132240 3746         35.3

>
> heidel.diag(reading_skills$gamma.mcmc)

         Stationarity start     p-value
         test         iteration
Series Z0 passed       1         0.301

         Halfwidth Mean Halfwidth
         test
Series 1 passed    2.2  0.0126
>
>
> geweke.diag(reading_skills$beta.mcmc.short, frac1=0.1,frac2=0.5)

Fraction in 1st window = 0.1
Fraction in 2nd window = 0.5
```

```
Series X0 Series X1 Series X2 Series X3
  1.0552  -0.6642  -0.1377   0.1387

>
> raftery.diag(reading_skills$beta.mcmc.short, q = 0.025, r = 0.005, s = 0.95)

Quantile (q) = 0.025
Accuracy (r) = +/- 0.005
Probability (s) = 0.95

You need a sample size of at least 3746 with these values of q, r and s
>
> heidel.diag(reading_skills$beta.mcmc.short)

         Stationarity start     p-value
         test         iteration
Series X0 passed       1         0.834
Series X1 passed       1         0.920
Series X2 passed       1         0.875
Series X3 passed       1         0.378

         Halfwidth Mean   Halfwidth
         test
Series X0 passed     2.086 0.00426
Series X1 passed    -1.704 0.00591
Series X2 passed     0.520 0.00593
Series X3 passed    -0.579 0.00760
>
> geweke.diag(reading_skills$gamma.mcmc.short, frac1=0.1,frac2=0.5)

Fraction in 1st window = 0.1
Fraction in 2nd window = 0.5

Series 1
  0.9876

>
> raftery.diag(reading_skills$gamma.mcmc.short, q = 0.025, r = 0.005, s = 0.95)

Quantile (q) = 0.025
Accuracy (r) = +/- 0.005
Probability (s) = 0.95

You need a sample size of at least 3746 with these values of q, r and s
>
> heidel.diag(reading_skills$gamma.mcmc.short)

         Stationarity start     p-value
         test         iteration
Series Z0 passed       1         0.202

         Halfwidth Mean Halfwidth
         test
```

```
Series Z0 passed    2.2  0.014
```

Finally, here we use the Bayesianbetareg package to plot the chains of the posterior parameter samples. Figures 4 and 5 shows a very small transient period giving a strongly intuitive indication of the convergence, agreeing with the theoretical result of convergence.

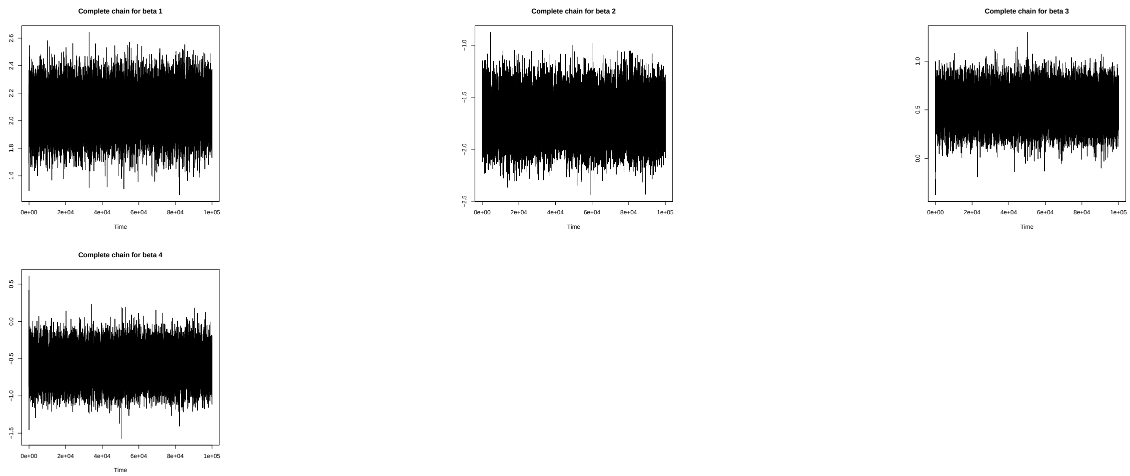


FIGURE 2: Plot - Chain of the posterior parameter samples of the $\boldsymbol{\beta}$ parameter.

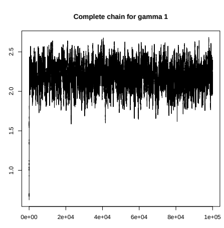


FIGURE 3: Plot - Chain of the posterior samples of γ_0.

5.2. Joint mean and dispersion beta regression models

Cribari-Neto & Zeileis (2010) fitted beta regression models to this score data set, applying the classic method, using the scores on the test as dependent variables and the other three variables as independent variables for the mean and the precision parameters. The results obtained by Cribari-Neto & Zeileis (2010) are:

```
> rs_beta <- betareg(accuracy ~ dyslexia * iq | dyslexia + iq, data = ReadingSkills,
hessian = TRUE)
> summary (rs_beta)

Call:
betareg(formula = accuracy ~ dyslexia * iq | dyslexia + iq, data = ReadingSkills,
    hessian = TRUE)

Standardized weighted residuals 2:
```

```
    Min      1Q  Median      3Q     Max
-2.3900 -0.6416  0.1572  0.8524  1.6446

Coefficients (mean model with logit link):
            Estimate Std. Error z value Pr(>|z|)
(Intercept)   1.1232     0.1509   7.444 9.76e-14 ***
dyslexia     -0.7416     0.1515  -4.897 9.74e-07 ***
Iq            0.4864     0.1671   2.911 0.003603 **
dyslexia:iq  -0.5813     0.1726  -3.368 0.000757 ***

Phi coefficients (precision model with log link):
            Estimate Std. Error z value Pr(>|z|)
(Intercept)   3.3044     0.2265  14.589  < 2e-16 ***
dyslexia      1.7466     0.2940   5.941 2.83e-09 ***
Iq            1.2291     0.4596   2.674  0.00749 **
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Type of estimator: ML (maximum likelihood)
Log-likelihood:  65.9 on 7 Df
Pseudo R-squared: 0.5756
Number of iterations in BFGS optimization: 25
```

5.3. Bayesian joint mean and precision beta regression models

A beta regression model, where the mean is a function of the dyslexia status, non-verbal IQ and their interaction, and the dispersion parameter is modeled as a function of the dyslexia status and non-verbal IQ, was fitted to the score data using the **Bayesianbetareg** function. The model assumes a *logit* link function for the mean and a logarithm function for the precision, including the intercepts, in order to compare the result obtained using the Bayesian method proposed by Cepeda-Cuervo(2001) with the result obtained under the classic method.

```
> data(ReadingSkills)
>
>
> Y <- as.matrix(ReadingSkills[,1])
> n <- length(Y)
> X1 <- as.matrix(ReadingSkills[,2])
> for(i in 1:length(X1)){
+   X1 <- replace(X1,X1=="yes",1)
+   X1 <- replace(X1,X1=="no",0)
+ }
> X0 <- rep(1, times=n)
> X1 <- as.numeric(X1)
> X2 <- as.matrix(ReadingSkills[,3])
> X3 <- X1*X2
> X <- cbind(X0,X1,X2,X3)
> Z0 <-  X0
> Z <- cbind(X0,X1,X2)
>
> burn <- 0.3
```

```
> jump <- 30
> nsim <- 300000
>
> plot(X1,Y, main="Accuracy vs Dislexy", xlab="Dislexy", ylab="Accuracy")
> plot(X2,Y, main="Accuracy vs IQ", xlab="IQ", ylab="Accuracy")
> plot(X3,Y, main="Accuracy vs Interaction", xlab="Interaction", ylab="Accuracy")
>
> plot(Y, main="Accuracy")
>
> b_pri <- c(0,0,0,0)
> B_pri <- diag(100,nrow=ncol(X),ncol=ncol(X))
> g_pri <- c(0,0,0)
> G_pri <- diag(10,nrow=ncol(Z),ncol=ncol(Z))
>
> reading_skills<- Bayesianbetareg (Y,X,Z,nsim,b_pri,B_pri,g_pri,G_pri,
burn,jump,graph1=T, graph2=T)
```

Just like for the previous beta regression models, the 95 % credibility intervals show that the mean and dispersion regression parameters are different from zero.

```
> summary(reading_skills)

################################################################
###                  Bayesian Beta Regression                ###
################################################################

 Call:
Bayesianbetareg.default(Y = Y, X = X, Z = Z, nsim = nsim, b_pri = b_pri,
    B_pri = B_pri, g_pri = g_pri, G_pri = G_pri, burn = burn,
    jump = jump, graph = T)

         Estimate Est.Error L.CredIntv U.CredIntv
beta.X0    1.44213   0.15216    1.15326      1.743
beta.X1   -1.10655   0.15438   -1.41132     -0.810
beta.X2    1.77584   0.10673    1.56610      1.979
beta.X3   -1.82596   0.10920   -2.03789     -1.618
gamma.Z0   1.16032   0.13011    0.89799      1.409
gamma.Z1   4.14779   0.34107    3.43596      4.780
gamma.Z2   3.29898   0.07004    3.16263      3.443

 Deviance:
[1] 528.5511

 AIC:
[1] 536.5511

 BIC:
[1] 543.6878
```

In all the cases, we obtained the same positive/negative sign for the estimated coefficients, and values close to the results reported by Cribari-Neto & Zeileis (2010).

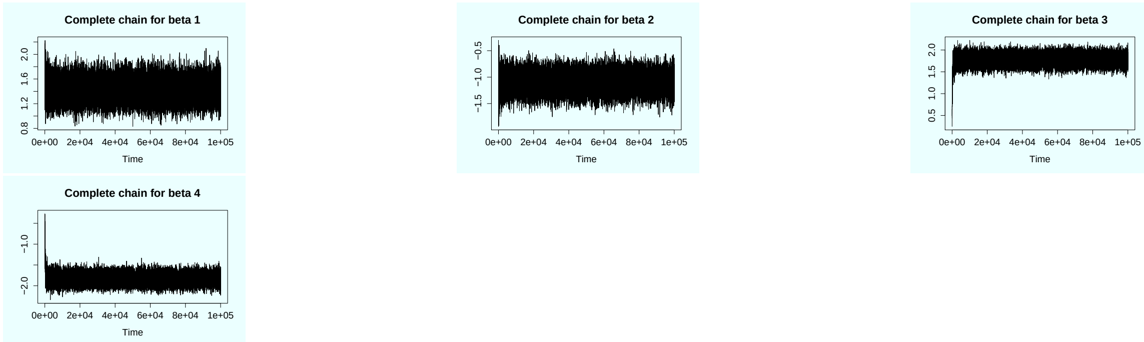


FIGURA 4: Plot - Chain of beta parameter

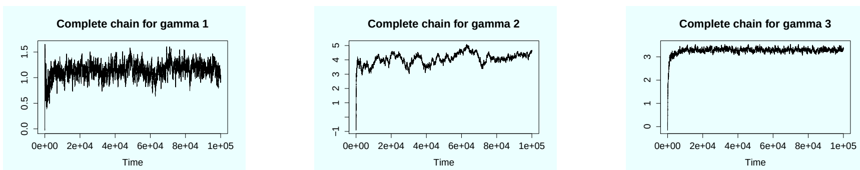


FIGURA 5: Plot - Chain of gamma parameter

The posterior chains seem to show good behavior, a small transient period and small standard deviations. Results of the Geweke, Heidelberger & Welch and Rifter & Lewis test indicate convergence of the chains for the mean regression parameters, at a level of 95 %. For precision regression parameters, the tests indicate convergence, except for the parameter associated with dyslexia.

The chains were cleaned using a burn in process excluding the first 30 % of the sample and a jump of 30 steps for recollect the final posterior samples. The results were:

```
> library (coda)

> geweke.diag(reading_skills$beta.mcmc.short, frac1=0.1,frac2=0.5)

Fraction in 1st window = 0.1
Fraction in 2nd window = 0.5

Series 1 Series 2 Series 3 Series 4
 1.25697 -1.09921 -0.18970  0.06014

>
> raftery.diag(reading_skills$beta.mcmc.short, q = 0.025, r = 0.005, s = 0.95)

Quantile (q) = 0.025
Accuracy (r) = +/- 0.005
Probability (s) = 0.95

You need a sample size of at least 3746 with these values of q, r and s
>
> heidel.diag(reading_skills$beta.mcmc.short)

         Stationarity start     p-value
```

```
         test         iteration
Series X0 passed       1         0.622
Series X1 passed       1         0.558
Series X2 passed       1         0.935
Series X3 passed       1         0.907

         Halfwidth Mean  Halfwidth
         test
Series X0 passed     1.44 0.00632
Series X1 passed    -1.11 0.00658
Series X2 passed     1.78 0.00411
Series X3 passed    -1.83 0.00442
>
> geweke.diag(reading_skills$gamma.mcmc.short, frac1=0.1,frac2=0.5)

Fraction in 1st window = 0.1
Fraction in 2nd window = 0.5

Series 1 Series 2 Series 3
 -0.5886  -0.8054   0.4717

>
> raftery.diag(reading_skills$gamma.mcmc.short, q = 0.025, r = 0.005, s = 0.95)

Quantile (q) = 0.025
Accuracy (r) = +/- 0.005
Probability (s) = 0.95

You need a sample size of at least 3746 with these values of q, r and s
>
> heidel.diag(reading_skills$gamma.mcmc.short)

         Stationarity start     p-value
         test         iteration
Series Z0 passed         1       0.1892
Series Z1 passed       801       0.0715
Series Z2 passed         1       0.4600

         Halfwidth Mean Halfwidth
         test
Series Z0 passed    1.16 0.01934
Series Z1 passed    4.23 0.07621
Series Z2 passed    3.30 0.00913
```

In the case of the short chains of $\boldsymbol{\beta}$ and $\boldsymbol{\gamma}$, all pass the three tests of convergence.

Finally, assessing the fit of both models, it can be clearly seen that the information criteria suggest that the best model is the one with constant precision.

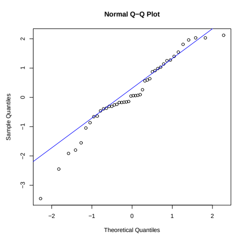
With this in mind, we perform a brief inspection of the residuals with the functions **beta.residuals**, **plotresiduals** and **diagnostics**, as shown above.

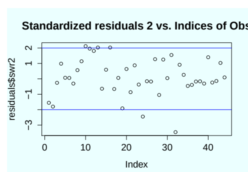
```
\begin{Schunk}
```

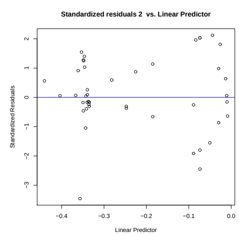
```
reading_skills.residuals<- beta.residuals(Y,reading_skills)
plotresiduals(X,Y,reading_skills.residuals,type=5)
diagnostics(Y,reading_skills,reading_skills.residuals)

\end{Schunk}
```

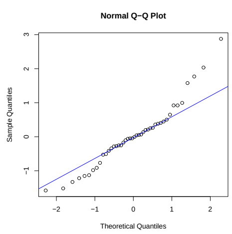
This reveals that the model that includes beta regression structures in both mean and precision provides better fit.

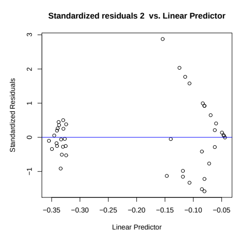
Regarding the divergence form the standard normal quartiles, it is important to take into account the low overall estimated precision in the chosen model. Thus, further inspection is needed of the points outside of the confidence bands for the errors, since their influence on the mean response may be due to errors outside of the sampling process.

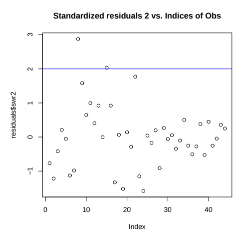












6. Conclusions

This paper is introduced the Bayesianbetareg R-package, that can be used to fit beta regression models applying Bayesian method proposed by Cepeda-Cuervo (2001) and summarize the results. We use the Dyslexia and IQ Predicting Reading Accuracy database to illustrate the use of the different functions of this package.

As a suggestions for future works and practical issues is possible introduce two key extensions. One is the use of alternative link functions that could adjust in a better way for different database and the second is the formulation of a model that have as target parameter the variance parameter.

Referencias

Atkinson, A. (1970), 'On the measurement oj inequality', *Joutnal of economic theory* **2**, 244–263.

Bonat, W. H., R. J. P. J. & Zeviani, W. M. (2013), 'Likelihood analysis for a class of beta mixed models. not published.'.

Branscum, A., Johnson, W. & Thurmond, M. (2007), 'Bayesian beta regression: applications to household expenditure data and genetic distance between foot-and-mouth disease viruses', *Australian & New Zealand Journal of Statistics* **49**, 287–301.

Cepeda-Cuervo, E. (2001), Modelagem da variabilidade em modelos lineares generalizados., PhD thesis, Instituto de Matemática. Universidade Federal do Río do Janeiro. URL: *http : //www.bdigital.unal.edu.co/9394/1/modelagem$_v$ariabilidade$_m$odelos.pdf*.

Cepeda, E. & Gamerman, D. (2005), 'Bayesian methodology for modeling parameters in the two-parameter exponential family', *Estadistica* **57**, 93–105.

Chan, S. (2006), Beta regression analysis of singapore general election results, Technical report, SIM University.

Cribari-Neto, F. & Zeileis, A. (2010), 'Beta regression in r.', *Journal of Statistical Software* **34**, 1–24.

Cuervo, E. C. & Achcar, J. A. (2010), 'Heteroscedastic nonlinear regression models.', *Communications in StatisticsŮSimulation and Computation* **39(2)**, 405–419.

Espinheira, P., Ferrari, S. L. & Cribari-Neto, F. (2008), 'On beta regression residuals', *Journal of Applied Statistics* **35**(4), 407–419.
*http://www.tandfonline.com/doi/abs/10.1080/02664760701834931

Ferrari, S. & Cribari-Neto, F. (2004), 'Beta regression for modelling rates and proportions', *Journal of Applied Statistics* **31**(7), 799–815.

Figueroa-Zuniga, J., Arellano-Valle, R. B. & S.L.P., F. (2013), 'Beta mixed regression: a bayesian perspective.', *Computational Statistics and Data Analysis.* **61**, 137–147.

Gamerman, D. (1997), 'Sampling from the posterior distribution in generalized linear mixed models.', *Statistics and Computing* **7**(1), 57–68.

Gamerman, D. & Cepeda-Cuervo, E. (2013), Generalized spatial dispersion models, Technical report, Universidade Federal do Rio de Janeiro.

Heidelberger, P. & Welch, P. D. (1981), 'A spectral method for confidence interval generation and run length control in simulations', *Comm. ACM.* **24**, 233–245.

Rocha, A. V. & Cribari-Neto, F. (2009), 'Beta autoregressive moving average models.', *Test* **18**(3), 529–545.

Simas, A., Barreto-Souza, W. & Rocha, A. (2010), 'Improved estimators for a general class of beta regression models.', *Computational Statistics & Data Analysis* **54**(2), 348–366.

Smithson, M. & Verkuilen, J. (2006), 'A better lemon squeezer? maximum-likelihood regression with beta-distributed dependent variables.', *Psychol Methods* **1**(11), 54–71.

Swearingen, C., Tilley, B., Adams, R., Rumboldt, Z., Nicholas, J., Bandyopadhyay, D. & Robert, F. (2011), 'Application of beta regression to analyze ischemic stroke volume in ninds rt-pa clinical trials', *Neuroepidemiology* **37**(2), 73–82.