BAYESIAN BETA REGRESSION MODELS
JOINT MEAN AND PRECISION MODELING

EDILBERTO CEPEDA CUERVO
Departamento de Estadística [1]
Universidad Nacional de Colombia

LILIANA GARRIDO
Departamento de Matemáticas [2]
Universidad de los Andes

Summary

This paper summarizes the beta regression models, with joint modeling of the mean and precision parameters, and the Bayesian methodology proposed by Cepeda (2001) and Cepeda and Gamenrman (2005) to fit these models. This Bayesian methodology is implemented and applied in the development of simulated and applied studies.

Key words: Beta regression, Bayesian methodology

[1]email: ecepedac@unal.edu.co
[2]email: bgarrido@uniandes.edu.co

1 Introduction

In this paper we analyze situations where the observations are associated with the beta distribution. A random variable with beta distribution, defined in equation (1), could represent random variation of a probability, fraction or prevalence, for example. Thus, this distribution has many applications in areas as finantial sciences or social sciences as education, where the random variables are continuous in a bounded interval which is isomorphic to the interval $[0, 1]$. To mention an example, in studies of the quality of education, a number between 0 and 5 (or any other positive integer bounds) is assigned as a measure of performance in the evaluation of school subjects as math, language, arts, natural sciences or any other scholar areas. In these cases, the measure assigned to each student can be expressed as a number between zero and one. Thus, it can be assumed that the level of student performance is a random variable with beta distribution.

The beta p, q distribution function, defined by equation (1) can be re-parametrized as a function of the mean and the so called dispersion parameter as in equation (4) or as function of the mean and variance. This characterizations of the beta distribution can be more appropriate. In the first re-parametrization, making $\phi = p + q$ we can see that $p = \mu\phi$, $q = \phi(1 - \mu)$ and $\sigma^2 = \frac{\mu(1-\mu)}{\phi+1}$. In this case, ϕ can be interpreted as a precision parameter in the sense that, for fixed values of μ, larger values of ϕ correspond to smaller values of the variance of Y. This reparametrization that is presented in Ferrari and Cribari-Neto (2004), has already appeared in the literature, for example in Jorgensen (1997) or in Cepeda (2001, pg 63).

In this case, the mean and dispersion parameters can be modeled as functions of explanatory variables. To cite a few examples, the educational level of mothers could influence the students school performance, the land concentration can be explained by random variables associated with social and political facts or the proportion of income spent monthly could be explained by the number of persons in the household. At the same time we can assume that the precision parameter changes as a function of the same or other random variables. With these ideas, Bayesian regression, with joint modeling of the mean and dispersion parameters, was initially proposed by Cepeda (2001, pg. 63), in the framework of joint modeling of parameters in the biparametric exponential family (see Cepeda and Gamerman 2001, 2005). In a later paper, Ferrari and Cribari-Neto (2004) proposed classical beta regression models, assuming that the dispersion parameter is constant through the rank of the explanatory variables. Further works have been published by Smithson and Verkuilen (2006), Simas et al. (2010) and, Cepeda-Cuervo and Achcar (2010a), the latter proposing nonlinear beta regression in the context of Double Generalized Nonlinear Models. This last model is extended in Cepeda et al (2011) and Cepeda and Nuñez-Anton (2011) where spatial correlation is assumed.

This paper summarizes the beta regression models, with joint modeling of the mean and precision parameters, and the Bayesian methodology proposed by Cepeda (2001) and Cepeda and Gamerman (2005) to fit these models. This Bayesian methodology is implemented and applied in the development of simulated and applied studies.

The rest of the paper is organized as follows. In section 2 general concepts about the beta distribution is presented. Section 3 presents the joint mean and precision beta regression models. Section 4 presents the Bayesian methodology proposed to fit the beta regression models. Section 5 includes simulations studies and section 6 includes an application of the percentage of income expended in food.

2 The beta distribution

A random variable Y has beta distribution if its density function is given by

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

where $p > 0$, $q > 0$, $\Gamma(.)$ denotes the gamma function and $I_{(0,1)}(y)$ the indicator function in the open interval $(0,1)$. The mean and variance of Y, $\mu = E(Y)$ and $\sigma^2 = Var(Y)$, are given by

$$\mu = \frac{p}{p+q} \tag{2}$$

$$\sigma^2 = \frac{p\,q}{(p+q)^2(p+q+1)} \tag{3}$$

Many random variables can be assumed to have beta distribution. For example, the income inequality or the land distribution when it is measured using the Gini index proposed by Atkinson(1970) and the performance of the students in subjects as mathematics, natural sciences or literature. In the last case, if the performance X takes values in the interval (a,b), the random variable $Y = (X-a)/(b-a)$ can be assumed to have beta distribution. This performance can be explained by household socioeconomic variables, that have fundamental impact on the cognitive achievement of students. For example, the level of student achievement is closely related to educational levels of their parents and the number of hours devoted to study a subject. Thus, the beta regression model could be appropriate to explain the behavior of the school performance as a function of associated factors. In these applications however, the reparametrization of the beta distribution given in (4) can be more appropriate. In the first one, doing $\phi = p + q$ we can see that $p = \mu\phi$, $q = \phi(1-\mu)$ and $\sigma^2 = \frac{\mu(1-\mu)}{\phi+1}$. In this case, ϕ can be interpreted as a precision parameter in the sense that, for fixed values of μ, larger values of ϕ correspond to smaller values of the variance of Y. This reparametrization that is presented in Ferrari and Cribari-Neto (2004), had already appeared in the literature, for example in Jorgensen (1997) or in Cepeda (2001). With this reparametrization, the density of the beta distribution (1) can be rewritten as

$$f(y|\alpha, \beta) = \frac{\Gamma(\phi)}{\Gamma(\mu\phi)\Gamma((1-\mu)\phi)} y^{\mu\phi-1}(1-y)^{(1-\mu)\phi-1} I_{(0,1)}(y) \tag{4}$$

In this case, the mean and dispersion parameters can be modeled as function of explanatory variables, for example, as was proposed in Cepeda(2001), given that changes in the precision parameter can be explained by explanatory variables, such as mothers educational level in the case of the student's school performance.

The beta distribution given in (1) can also be reparametrized as a function of the mean and variance, with

$$p = \frac{(1-\mu)\mu^2 - \mu\sigma^2}{\sigma^2} \tag{5}$$

$$q = \frac{(1-\mu)[\mu - \mu^2 - \sigma^2]}{\sigma^2} \tag{6}$$

Although writing (1) as a function of μ and σ^2 can result in a complex expression, joint modeling of the mean and variance can be easily achieved applying the Bayesian methodology proposed in Cepeda(2001) and Cepeda and Gamerman (2005). Sometimes the joint modeling of the mean and variance could be more appropriate than the joint modeling of the mean and the so called precision parameter, given that the parameters of the regression models would be more easily interpreted.

3 Joint modeling in beta regression

With the reparametrization of the beta distribution as a function of μ and ϕ we can define a double generalized beta regression model as is proposed in Cepeda (2001) and in Cepeda and Gamerman (2005). In that research work the joint modeling of the mean and dispersion parameters in the beta regression model and a Bayesian methodology to fit the parameters of the proposed model, was defined. In a more general frame, for example they assume a random sample $Y_i \sim Beta(p_i, q_i)$, $i = 1, 2, \ldots, n$, where both, the mean and the precision parameter, are not constant for all observations and are modeled as regression models. That is,

$$\begin{aligned} \text{logit}(\mu_i) &= \mathbf{x}_i^t \boldsymbol{\beta} \qquad (7) \\ \log(\phi_i) &= \mathbf{z}_i^t \boldsymbol{\gamma} \end{aligned}$$

where $\boldsymbol{\beta} = (\beta_0, \beta_1, ..., \beta_k)$ and $\boldsymbol{\gamma} = (\gamma_0, \gamma_1, \ldots, \gamma_p)$ are the vectors of the mean and dispersion regression models and, $\mathbf{x}_i$ and $\mathbf{z}_i$ are the vectors of the mean and dispersion explanatory variables, at the i-th observation, respectively. In a later paper, Ferrari and Cribari-Neto (2004) proposed the same reparametrization of the beta distribution, that is $\mu = p/(p+q)$ and $\phi = p+q$. In that paper, they assumed that $g(\mu_i) = \mathbf{x}_i^t\boldsymbol{\beta}$, where g is a strictly monotonic and twice differentiable real valued link function, defined on $(0,1)$, assuming that the dispersion parameter is constat through the range of the explanatory variables. Although they consider many possible link functions, in the applications they take the logit link function, given that the mean can be interpreted as a function of the odds ratio. The joint beta regression models proposed by Cepeda(2001), was later studied by Smithson and Verkuilen (2006) and then by Simas et al. (2010). At the same time a nonlinear beta regression was proposed by Cepeda and Achcar (2010), assuming the model

$$\mu_i = \frac{\beta_0}{1+\beta_1 \exp(\beta_2 x_i)} \tag{8}$$

$$\log(\phi_i) = \mathbf{z_i}^t\boldsymbol{\gamma} \tag{9}$$

in the context of Double Generalized Nonlinear Models (Cepeda and Gamerman, 2005). This model was applied to the schooling rate data analysis in Colombia, for the period ranging from 1991 to 2003.

4 Bayesian methodology

To implement a Bayesian approach to estimate the parameters of the model (7) we need to specify prior distributions for the parameters of the model. For simplicity, we assume independent normal prior distributions given by $\boldsymbol{\beta} \sim N(\mathbf{b}, \mathbf{B})$ and $\boldsymbol{\gamma} \sim N(\mathbf{g}, \mathbf{G})$ where $\mathbf{b}, \mathbf{B}, \mathbf{g}, \mathbf{G}$ are given by the researcher, based on prior knowledge. Thus, if $L(\boldsymbol{\beta}, \boldsymbol{\gamma}|$ data) denotes the likelihood function and $p(\boldsymbol{\beta}, \boldsymbol{\gamma})$ the joint prior distribution, the posterior distribution is given by

$\pi(\boldsymbol{\beta},\boldsymbol{\gamma}|\text{ data}) \propto L(\boldsymbol{\beta},\boldsymbol{\gamma}|\text{ data})p(\boldsymbol{\beta},\boldsymbol{\gamma})$. Given that the posterior distribution $\pi(\boldsymbol{\beta},\boldsymbol{\gamma}|\text{ data})$ is analytically intractable and it is not easy to generate samples from it, Cepeda(2001) proposed to sample these parameters using an iterative alternating process, that is, sampling $\boldsymbol{\beta}$ and $\boldsymbol{\gamma}$ from the posterior conditional distributions $\pi(\boldsymbol{\beta}|\boldsymbol{\gamma},\text{ data})$ and $\pi(\boldsymbol{\gamma}|\boldsymbol{\beta},\text{ data})$, respectively. But given that these distributions are analytically intractable, the Bayesian methodology proposed in Cepeda (2001) Cepeda and Gamerman (2001), in which it is necessary to build normal transition kernels, is applied. Specifically, to fit the beta regression models where $h(\mu) = \mathbf{X}\beta$ and $g(\tau) = Z\gamma$, where $\tau = \phi$ (or $\tau = \sigma^2$), we build working variables to approximate $h(\mu)$ and $g(\tau)$ around the current values of μ and τ, respectively. Given that $E(y_i) = \mu_i$ and h has Taylor representation in some neighborhood of $\mu_i = h^{-1}(\mathbf{x}_i'\boldsymbol{\beta})$, to obtain the posterior samples of $\boldsymbol{\beta}$, we define the working variable as the first order Taylor approximation

$$h(y_i) \approx h(\mu_i) + h'(\mu_i)(y_i - \mu_i) = \tilde{y}_i. \tag{10}$$

This new random variable has $E(\tilde{y}_i) = \mathbf{x}_i'\boldsymbol{\beta}$ and $Var(\tilde{y}_i) = [h'(\mu_i)]^2 Var(y_i)$. Thus, if $\boldsymbol{\beta}^{(c)}$ and $\boldsymbol{\gamma}^{(c)}$ are the current values of $\boldsymbol{\beta}$ and $\boldsymbol{\gamma}$, the appropriate working observation variable (10), defined to build the transition kernel to get samples of $\boldsymbol{\beta}$, can be rewritten as

$$\tilde{y}_i = \mathbf{x}_i'\boldsymbol{\beta}^{(c)} + h'[h^{-1}(\mathbf{x}_i'\boldsymbol{\beta}^{(c)})][y_i - h^{-1}(\mathbf{x}_i'\boldsymbol{\beta}^{(c)})], \text{ for } i = 1, 2, ..., n, \tag{11}$$

and its associated working observational variance as

$$\tilde{\sigma}_i^2 = \{h'[h^{-1}(\mathbf{x}_i'\boldsymbol{\beta}^{(c)})]\}^2 Var(y_i). \tag{12}$$

Thus, assuming that (11) has normal distribution and given the normal conditional prior distribution $\boldsymbol{\beta}|\boldsymbol{\gamma} \sim N(b, B)$, the normal transition kernel Q_1 is given by the posterior distribution obtained from the combination of the prior distribution with the working observation model $\tilde{y}_i \sim N(\mathbf{x}_i\boldsymbol{\beta}, \tilde{\sigma}_i^2)$. This is,

$$Q_1(\boldsymbol{\beta}|\hat{\boldsymbol{\beta}}, \hat{\boldsymbol{\gamma}}) = N(\mathbf{b}^* \ , \mathbf{B}^*), \tag{13}$$

where

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

with $\boldsymbol{\Sigma}$ a diagonal matrix with diagonal entries $\tilde{\sigma}_i^2$, $i = 1, 2, ..., n$. Thus, the values of $\boldsymbol{\beta}$ from the posterior distribution sample of $\pi(\boldsymbol{\beta}, \boldsymbol{\gamma})$ are proposed from the transition kernel (13).

To obtain posterior samples of $\boldsymbol{\gamma}$, a kernel transition function is built assuming that there exists a random variable t_i such that $E(t_i) = \tau_i$, where $\tau_i = g^{-1}(\mathbf{z}_i'\boldsymbol{\gamma})$. The working observational variable $\hat{y}_i$ is given by the first order Taylor approximation of $g(t_i)$

$$g(t_i) = g(\tau_i) + g'(\tau_i)(t_i - \tau_i) = \hat{y}_i, \quad i = 1, \ldots, n. \tag{14}$$

Thus, $E(\hat{y}_i) = \mathbf{z}_i'\boldsymbol{\gamma}$ and $Var(\hat{y}_i) = [g'(\mu_i)]^2 Var(t_i)$. In consequence, if $\boldsymbol{\beta}^{(c)}$ and $\boldsymbol{\gamma}^{(c)}$ are the current values of $\boldsymbol{\beta}$ and $\boldsymbol{\gamma}$, the working observational variables (14) can be rewritten as

$$\hat{y}_i = \mathbf{z}_i'\boldsymbol{\gamma}^{(c)} + g'[g^{-1}(\mathbf{z}_i'\boldsymbol{\gamma}^{(c)})][t_i - h^{-1}(\mathbf{z}_i'\boldsymbol{\gamma}^{(c)})], \text{ for } i = 1, 2, ..., n, \tag{15}$$

and their associated observational working variances as

$$\hat{\sigma}_i^2 = \{g'[g^{-1}(\mathbf{z}_i^{'}\boldsymbol{\gamma}^{(c)})]\}^2 Var(t_i). \tag{16}$$

Assuming that the observational working variables (15) have independent normal distributions and given that the conditional prior distribution is given by $\boldsymbol{\gamma}|\boldsymbol{\beta} \sim N(\mathbf{g}, \mathbf{G})$, the normal transition kernel to obtain the posterior samples is given by the posterior distribution obtained from the combination of the prior distribution with the working observational models $\hat{y}_i \sim N(\mathbf{z}_i^{'}\boldsymbol{\gamma}, \hat{\sigma}_i^2)$, $i = 1, \dots, n$. This is,

$$Q_2(\boldsymbol{\gamma}|\hat{\boldsymbol{\gamma}}, \hat{\boldsymbol{\beta}}) = N(\mathbf{g}^*\ , \mathbf{G}^*), \tag{17}$$

where

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

with Ψ a diagonal matrix with entries $\hat{\sigma}_i^2$ for $i = 1, 2, ..., n$. From the transition kernel Q_2, samples of the posterior distribution $\pi(\boldsymbol{\beta}, \boldsymbol{\gamma})$ are proposed.

With the transition kernels given by (13) and (17), the parameters vector $(\boldsymbol{\beta}, \boldsymbol{\gamma})^{'}$ is updated as follows:

1. Begin the chain interactions counter $j = 1$ and give 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 $\boldsymbol{\beta}^{(j)} = \boldsymbol{\beta}^{(j-1)}$.

   For the acceptance of the movement an observation u is drawn from the uniform distribution $\mathrm{U}(0, 1)$. If $\alpha(\boldsymbol{\beta}^{(j-1)}, \psi) < u$ the movement is accepted. Otherwise, the movement is rejected.

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 a specified number of draws. If convergence has not been reached, the number of draws should be incremented until convergence is achieved.

In the case where the data come from the beta distribution, $Y_i \sim B(p_i, q_i)$, $i = 1, 2, \ldots, n$, with mean and precision models given by (7), the working observational variable associated with the mean is obtained from (11), with $t_i = Y_i$, and the working observational variable associated with the precision model is obtained from (15) with $t_i = \frac{(p_i+q_i)^2}{p_i} Y_i$. These working observational variables are given by:

1. For $Y_i \sim B(p_i, q_i)$, the mean $\mu_i = p_i/(p_i + q_i)$ can be modeled as $logit(\mu_i) = \boldsymbol{x}_i' \boldsymbol{\beta}$ and $p_i + q_i$ can be modeled as $log(p_i + q_i) = \boldsymbol{z}_i' \boldsymbol{\gamma}$.

Then the appropriate working observational variable defined to build the transition kernel to get samples of $\boldsymbol{\beta}$ is

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

where $\mu^{(c)}$ and $\boldsymbol{\beta}^{(c)}$ are the current values of μ and $\boldsymbol{\beta}$.

2. For $\phi_i = p_i + q_i$, we propose the model $\tau_i = \exp(\boldsymbol{z}_i\boldsymbol{\gamma})$. In this case, given that $E(t_i) = p_i + q_i$ for $t_i = \frac{(p_i+q_i)^2}{p_i}Y_i$, the working observational variable obtained from (15) is

$$\begin{aligned} \tilde{Y}_i &= \boldsymbol{z}_i'\boldsymbol{\gamma}^{(c)} + \frac{\frac{(p_i^{(c)}+q_i^{(c)})^2}{p_i^{(c)}}Y_i - (p_i^{(c)} + q_i^{(c)})}{p_i^{(c)} + q_i^{(c)}}, \\ &= \boldsymbol{z}_i'\boldsymbol{\gamma}^{(c)} + \frac{Y_i}{\mu_i^{(c)}} - 1, \quad i = 1, 2, ..., n. \end{aligned} \tag{18}$$

The normal transition kernels are given by the posterior distributions obtained from the combination of the prior distributions with the working observational models given by the working observational variables as in equations (13) and (17).

5 Simulation studies

In this section we include the results of two simulation studies in which we analyze the performance of the Bayesian methodology proposed to fit the joint modeling of the mean and precision parameters in the beta regression models. In both cases 40 values of two independent explanatory variables X and Z were generated from the uniform distribution $U(0, 20)$. With these data, denoted by x_i and z_i, $i = 1, \ldots, n$, respectively, the mean and precision parameters of the beta distribution were generated from the mean and precision models

$$\begin{aligned} \text{logit}(\mu_i) &= \beta_0 + \beta_1 x_i \text{ and} && (19) \\ \log(\phi_i) &= \gamma_0 + \gamma_1 z_i && (20) \end{aligned}$$

respectively, where the true parameter values are given in each of the simulations. Then, values of the interest variable Y were generated from the beta distribution $B(\mu_i, \phi_i)$.

5.1 First simulation

To apply Bayesian methodology, normal prior distributions of the form $N(a, 10^k I)$ were assigned to the mean and dispersion parameters, where I stands for the 2×2 identity matrix, $a = (0, 0)$ and $k = 4$. The number $k = 4$ was chosen to impose large prior variances but, as we have checked in our analysis, increasing this value to larger orders of magnitude made no effective difference in the estimation process. Larger values could have also been used leading to very small changes in the posterior distributions.

With the data generated assuming the true parameter values (t.v) given in Table (5.1), the posterior summary of the parameters obtained using the usual MCMC algorithms, implemented in MatLab software, is given in Table (5.1), where B.E. denote the Bayesian parameter estimates and s.d their corresponding standard deviation. The estimates were obtained with a sample of size 400 obtained after a burn-in period of 6000 steps with a sampling gap of size 10. In the Table (5.1) we can see that the estimates are close to the corresponding true values, that are at less than one standard deviation from their estimates.

	Mean model		Precision model	
Parameters	β_0	β_1	γ_0	γ_1
t.v.	0.75	-0.055	0.15	-0.04
B.E.	0.437	-0.047	0.017	-0.046
s.d.	0.410	0.036	0.303	0.025

Table 1: Posterior parameter estimates, first simulation.

Figure 5.1 includes the chains and the histograms of the posterior samples. Although the chains are shown from iteration 6000, in general, they showed a small transient period, indicating a good performance of the proposed Bayesian methodology. The horizontal line in the graphs of the chains represent the true parameter values. The histograms seem to show that the posterior samples of the parameters come from normal distributions.

5.2 Second simulation

The second simulated study was developed assuming the mean and variance models given by (19) and (20). With the same prior distribution as in the first simulated study and with the data set generated with the true parameter values given in Table (5.2), samples of the posterior parameter distribution are obtained using usual MCMC algorithm. The posterior parameter estimates and their respective standard deviation, obtained with a sample of size

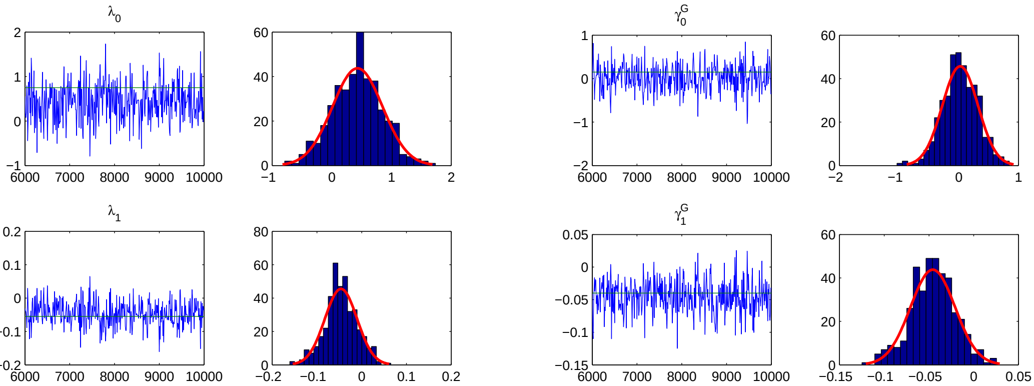

Figure 1: Chains and histograms of the posterior samples.

400 obtained after a burn-in period of 6000 steps with a sampling gap of size 10 are given in Table (5.2). As in the first simulation, we can see that the estimates are close to the corresponding true values, that are at less than one standard deviation from their estimates.

	Mean model		variance model	
Parameters	β_0	β_1	γ_0	γ_1
t.v.	0.450	-0.035	-0.350	0.025
B.e.	0.465	-0.056	-0.356	0.043
s.d	0.416	0.036	0.310	0.026

Table 2: Posterior parameter estimates, second simulation.

Figure 6 includes the chains and histograms of the posterior samples. In all cases, many chains were simulated starting at different initial values, which converged to the same values after a short transition period. Thus, the chains exhibit the same qualitative behavior, providing a rough indication of stationarity. For these models, the estimates of the precision parameters (and their respective standard deviation) are in Table 5.2.

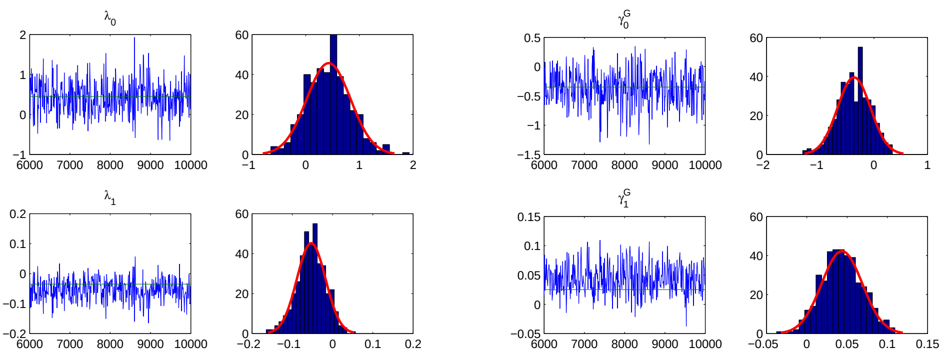

Figure 2: Chains and histograms of the posterior samples.

6 Application

In this section we present the results of the analysis of a data set which consists of 38 households in a large U.S. city, taken from Griffiths et al. (1993, Table 15.4). The interest variable is the proportion of the income spent on food, and the explanatory variables are the level of income INC and the number of persons in the household NUM. Although this data set was analyzed using beta regression models assuming constant precision parameter, in this case, we assume a joint modeling of the mean and precision parameters. Thus, we initially assume the models given by equation (21).

$$\begin{aligned} \text{logit}(\mu) &= \beta_0 + \beta_1 INC + \beta_2 NUM \\ \log(\phi) &= \gamma_0 + \gamma_1 INC + \gamma_2 NUM \end{aligned} \tag{21}$$

Assuming normal prior distribution $N(a, 10^4 I)$, where I stands for the 3×3 identity matrix, $a = (0, 0, 0)$, the parameter estimates and the corresponding standard deviations, obtained using the usual MCMC algorithms, are given in Table 6. This estimate were obtained from a posterior sample obtained after a burning of 5000 steps with a sampling gap of size 10. Table (6) includes the parameters estimation of two models, where $\hat{\theta}$ denotes the sample posterior mean of the parameters. The first one with all explanatory variables, as in equation (21). The second one without the random variable INC in the precision model, given that the standard deviation associated with $\hat{\gamma}_1$ seems to show that this parameter is not statistically different from zero.

	Mean model			Precision model		
Parameters	β_0	β_1	β_2	γ_0	γ_1	γ_2
$\hat{\theta}$	-0.7404	-0.0093	0.0978	4.6529	0.0048	-0.3593
s.d.	0.2074	0.0031	0.0411	0.7840	0.0123	0.2113
$\hat{\theta}$	-0.7582	-0.0091	0.1021	4.8428	....	-0.3500
s.d.	0.2091	0.0030	0.0386	0.4936	.....	0.1298

Table 3: Parameters estimation.

The BIC value of the model where all the variables are included in the mean and precision models is $BIC = -118.2793$. When we consider the same model for the mean but only including number of persons in the household NUM in the precision model, the BIC value is given by $BIC = -118.8582$. If the joint beta regression model with mean model given by (21) and precision regression model only with INC as explanatory is considered, the BIC value is given by $BIC = -82.6356$. Thus, according to the BIC values, the best model to fit this households data is which include INC and NUM in the mean model and NUM in the precision models, given that it has the least BIC value.

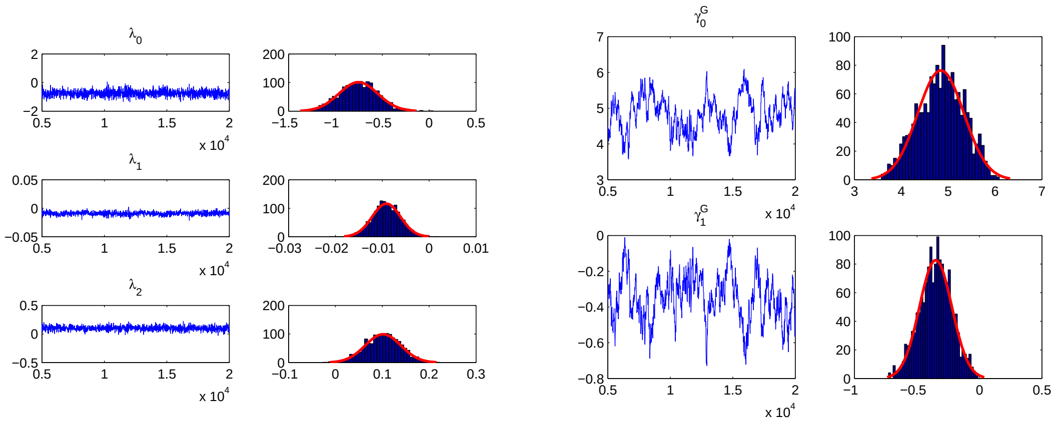

Figure 3: Estimates mean parameters.

To compare with the results obtained by Ferrari and Cribari-Neto (2004) applying classical methodology, we also fit the beta regression model with constant precision parameter ϕ. The Bayesian parameter estimates and the classical parameter estimates are given in Table 6. In the case of the Bayesian beta regresion the BIC value is given by $BIC = -82.2636$

We also notice that there is agreement between classical and bayesian parameter estimation. Thus, as in Ferrari and Cribari (2004) a negative re-

Parameters	β_0	β_1	β_2	ϕ
Classical Estimates	-0.62255	-0.01230	0.11846	35.60975
d.s	0.22385	0.00304	0.03534	8.07960
Bayesian Estimates	-0.6237	-0.0124	0.1190	32.8666
d.s	0.2357	0.0033	0.0379	7.1815

Table 4: Clasical and Bayesian parameters estimation.

lationship between the mean response (proportion of income spent on food) and the level of income, and a positive relationship between the mean response and the number of persons in the household.

7 Conclusion

This paper presents a join mean and dispersion models proposed by Cepeda (2001) and the Bayesian methodology proposed to develop posterior parameter inferences of the parameters. The studies with simulated and real data sets indicate good performance of the proposed Bayesian methodology, showing the agreement between true and estimates values, in the case of the simulated study, and between Bayesian and classical estimates, in the case of real data set analysis.

Some extension of the models and the methodologies included in this paper can be proposed. In the theoretical framework, join modeling of the mean and precision parameter, as regression models, can be proposed in a bivariate beta distribution, defined using copulas functions. In this case, the posterior samples of the regression can be obtained as in this paper and a proposal to get samples of the association parameter can be developed using, for example, a random walk.

Acknowledgements

Cepeda´s work was supported by a grant from The Research Division of the National University of Colombia (Universidad Nacional de Colombia). The authors thank Dr. Dani Gamerman for helpful comments and an anonymous reviewer for providing thoughtful comments and suggestions which led to substantial improvement in the presentation of the material in this paper.

References

[1] Cepeda-Cuervo, E. , Achcar J. A. Heteroscedastic Nonlinear Regression Models *Communications in Statistics - Simulation and Computation*, 39(2):405-419, 2010

[2] Cepeda C. E. and Gamerman D. (2001). Bayesian modeling of variance heterogeneity in normal regression models, *Brazilian Journal of Probability and Statistics* , **14**,207-221.

[3] Cepeda C. E. and Gamerman D.. (2003). Bayesian Methodology for Modeling Parameters in the two Parameters Exponential Family. Tecnical Relatory. DME/IM-UFRJ.

[4] McCullagh, P. and J. A. Nelder (1989). Generalized Linear Models, *Second ed. London: Chapman and Hall.*

[5] Cepeda, E.C. (2001). Variability Modeling in Generalized Linear Models, *Unpublished Ph.D. Thesis. Mathematics Institute, Universidade Federal do Rio de Janeiro.*

[6] Cepeda, E.C. and Gamerman, D. (2005). Bayesian methodology for modeling parameters in the two parameter exponential family, *Revista Estadística*, **57(168-169)**,93-105.

[7] Cepeda-cuervo, E., Achcar J. (2010). Heteroscedastic nonlinear regression models. Commun. Stat., Simulation Comput. 39, No. 2, 405-419.

[8] Ferrari, S., Cribari-Neto, F. (2004). Beta regression for modeling rates and proportions. Journal of Applied Statistics 31, 799-815.

[9] Griffiths W.E., Hill R.C., Judge G.G. Learning and Practicing Econometrics. John Wiley & Sons 1993.

[10] Jørgensen, B. (1997). Proper dispersion models (with discussion). *Brazilian Journal of Probability and Statistics*, 11, 89-140.

[11] Smithson M, Verkuilen J (2006). A Better Lemon Squeezer? Maximum-Likelihood Regression with Beta-Distributed Dependent Variables., *Psychological Methods*, **11**(1),54-71.