Revista Colombiana de Estadística - Theorical Statistics
January 2023, volume 46, issue 1, pp. 63–79
http://doi.org/10.15446/rce.v46n1.105335

$\mu\sigma^2$-Beta and $\mu\sigma^2$-Beta Binomial Regression Models

Modelos de regresión $\mu\sigma^2$-Beta y $\mu\sigma^2$-Beta binomial

Edilberto Cepeda-Cuervo[a]

Department of Statistics, Faculty of Sciences, Universidad Nacional de Colombia, Bogotá D.C., Colomb

Abstract

This paper proposes new parameterizations of the beta and beta binomial distributions as functions of the mean and variance parameters. From these new parameterizations, new beta and beta binomial linear regression models are formulated by assuming that appropriate real functions of the mean and variance follow linear regression structures. These models were fitted to real datasets by applying Bayesian methods, using the OpenBUGS software. The new beta regression models were fitted to the Dyslexia Reading Accuracy dataset and the new beta binomial regression models were applied to the School Absenteeism Dataset. In both cases, the results obtained by fitting these models were compared with those obtained by fitting the usual mean and dispersion beta regression models and the mean and dispersion beta binomial regression models, respectively.

***Key words*:** Mean and variance beta and beta binomial distributions; Beta and beta-binomial regression models; Bayesian methods.

Resumen

Este artículo propone nuevas parametrizaciones de las distribuciones beta y beta binomial como funciones de los parámetros de media y varianza. A partir de estas nuevas parametrizaciones, se formulan nuevos modelos de regresión lineal beta y beta binomial asumiendo que funciones reales apropiadas de la media y la varianza siguen estructuras de regresión lineal. Estos modelos se ajustaron a conjuntos de datos reales mediante la aplicación de métodos bayesianos, utilizando el software OpenBUGS. Los nuevos modelos de regresión beta se ajustaron al conjunto de datos de precisión de lectura de niños con dislexia y los nuevos modelos de regresión beta binomial se aplicaron al conjunto de datos de ausentismo escolar. En ambos casos, los

[a]Ph.D. Professor. E-mail: ecepedac@unal.edu.co

resultados obtenidos ajustando estos modelos se compararon con los obtenidos ajustando los modelos habituales de regresión beta de media y dispersión y los modelos de regresión beta binomial de media y dispersión, respectivamente.

***Palabras clave*:** Media y varianza; Distribución beta; Distribución beta binomial; Modelos de regresión beta y beta-binomial; Métodos bayesianos.

Introduction

This paper presents results of analyzing situations where the observations of the variables of interest are associated with the beta distribution. The beta distribution has applications in uncertainty of random variation of probability, fraction or prevalence, among others. Thus, this distribution has many applications in areas such as financial sciences, social sciences like education (Cepeda-Cuervo & Núñez Antón, 2013) and epidemiology (Hunger et al., 2012), where random variables are continuous in a bounded interval that is isomorphic to the interval $(0, 1)$. To mention an example, in studies of the quality of education, a number in a continuous scale from 0 to 5 (or any other positive integer bounds) is assigned as a measure of performance in school subjects like math or language (Cepeda & Gamerman, 2005). In these cases, the measure assigned to each student can be expressed as a number from zero to one. Thus, it can be assumed that the level of student performance is a random variable with a beta distribution. In education systems and other fields, a need often exist to model continuous random variables that assume values in a bounded interval on a set of explanatory variables. In these cases, if a continuous random variable W assumes values in a bounded open interval (a, b), a beta regression model can be proposed by defining a random variable $Y = (W - a)/(b - a)$, which can be assumed to follows the beta distribution.

Some variables can be assumed to follow beta distributions where their parameters are modeled as functions of explanatory variables. For example, students' performance can be explained by educational factors like educational level of mothers; land concentration can be explained by random variables associated with social and political factors; and the proportion of income spent monthly can be explained by the number of persons in the household. With these ideas, beta Bayesian regression models, with joint modeling of the mean and dispersion parameters, were initially proposed by Cepeda-Cuervo (2001), p. 63 in the framework of joint modeling in the biparametric exponential family. He described beta regression models where the mean, $\mu = p/(p+q)$, and the so called dispersion parameter of the beta distribution, $\nu = p + q$, follow linear regression structures:

$$h(\mu) = \boldsymbol{x}^t\boldsymbol{\beta} \text{ and} \quad (1)$$

$$g(\nu) = \boldsymbol{z}^t\boldsymbol{\gamma}, \quad (2)$$

where $h(\cdot)$ is the logit function and $g(\cdot)$ is the logarithmic function; $\boldsymbol{x}$ and $\boldsymbol{z}$ are vectors of explanatory variables and, $\boldsymbol{\beta} = (\beta_0, \ldots, \beta_k)$ and $\boldsymbol{\gamma} = (\gamma_0, \ldots, \gamma_p)$ are the respective vectors of regression parameters. Other appropriate link functions can be considered. These parameterization of the beta distribution can be appropriate in the beta regression models' definition, where ν can be interpreted as a "precision" parameter in the sense that, for fixed values of μ, larger values of ν correspond to smaller values of the variance (a conditional interpretation of ν). Independently of this proposal, in the same year, Paolino (2001) proposed beta regression models assuming that p and q, the parameters of the beta distribution $B(p, q)$, follow regression structures.

Since 2001, beta regression models have been extensively studied and applied in statistics. Using the μ-ν parameterizations of the beta distribution, Ferrari & Cribari-Neto (2004) proposed beta regression models with constant dispersion parameters, where the mean follows a regression structure given by $h(\mu) = \boldsymbol{x}^t\boldsymbol{\beta}$, in which h is an appropriate twice differentiable function. With this parameterization, they wrote the beta density function in terms of μ and ν, obtained the parameter estimates using maximum likelihood. Smithson & Verkuilen (2006) propose frecuentist aproach to the beta regression models, where both mean and precision are modeled with distinct sets of predictors. Further work has been published by Simas et al. (2010), proposing nonlinear beta regression models, where the mean parameter varies through a nonlinear regression structure and the precision parameters vary through a linear structure. Cepeda & Achcar (2010) proposed nonlinear beta regression in the context of double generalized nonlinear models. The beta regression models were extended in Cepeda-Cuervo & Núñez Antón (2013), assuming that the observations are spatially correlated.

Taking into account the conditional and restricted interpretation of ν, the so called precision parameter, and following the proposal presented in Cepeda-Cuervo & Garrido (2015), in which a restricted mean and variance regression model was proposed, in this paper we propose the mean and variance parameterization of the beta distribution, with clear and straightforward parameter interpretation. With this parameterization of the beta distribution, new beta regression models are proposed:

$$h(\mu) = \boldsymbol{x}^t\boldsymbol{\beta} \text{ and} \tag{3}$$

$$g(\sigma^2) = \boldsymbol{z}^t\boldsymbol{\lambda}, \tag{4}$$

where $h(.)$ and $g(.)$ are appropriate real valued functions (one to one and two differentiable) from the open interval $(0, 1)$ to the real number set. There are many examples of the mean link function $h(.)$, like logit, probit and complementary log-log functions. Examples of the variance link functions are given by the mean link functions, defined over an appropriate transformation of the interval $(0, 1/4)$ to the interval $(0, 1)$. Here we apply the beta regression models to the reading accuracy data for dyslexic and non-dyslexic Australian children (Smithson & Verkuilen, 2006) and compare the results with that obtained by usual models presented in the literature.

From these new parameterizations of the beta distribution, many statistical extensions can be proposed. For example, using a $\mu\sigma^2$ parameterization of the beta distribution, a new parameterization of the beta binomial distribution is proposed, where this distribution is defined by assuming that $Y \mid \pi^*$ follows binomial distribution, $Y \mid \pi^* \sim Bin(n, \pi^*)$, where π^* follows $\mu\sigma^2$-beta distribution. From these parameterizations of the beta binomial distribution, new beta binomial regression models are proposed and applied to the School Absenteeism dataset analysis.

This paper is organized as follows. After this introduction, in Section 1, we define the $\mu\sigma^2$-beta distribution and the $\mu\sigma^2$-beta binomial distribution. In Section 2, we propose the $\mu\sigma^2$ beta regression models and the $\mu\sigma^2$-beta binomial regression models. In Section 3, we describe the results of applying these models to the analysis of dyslexia and the School Absenteeism datasets. Finally, some conclusions of the paper are included in Section 4.

1. New Parameterizations of the Beta and Beta Binomial Distributions

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

$$f(y|p,q) = \frac{\Gamma(p+q)}{\Gamma(p)\Gamma(q)} y^{p-1}(1-y)^{q-1} I_{(0,1)}(y), \tag{5}$$

where $p > 0$, $q > 0$ and $\Gamma(.)$ denotes the gamma function. The mean and variance of Y, $\mu = E(Y)$ and $\sigma^2 = Var(Y)$, are given by:

$$\mu = \frac{p}{p+q} \text{ and} \tag{6}$$

$$\sigma^2 = \frac{p\,q}{(p+q)^2(p+q+1)}. \tag{7}$$

Many random variables can be assumed to have a beta distribution. Some examples, are income inequality and land distribution when these are measured using the Gini index, as proposed by Atkinson (1970), or the performance of students in subjects such as mathematics and language. In last case, if the performance or some other continuous variable X takes values in an open interval (a, b), the random variable $Y = (X - a)/(b - a)$ can be assumed to have a beta distribution. Student performance can be explained, for example, by household socioeconomic variables, since these can have a large impact on students' cognitive achievement. The level of students' achievement is also closely related to the educational level of their parents, especially by the educational level of the mothers and the number of hours devoted to studying a subject. Thus, the beta regression model can be appropriate to explain the behavior of school performance as a function of associated factors in a mean and variance regression structure.

1.1. Mean and variance ($\mu\sigma^2$) beta distribution

From the beta density function defined in (5), the mean (μ) and dispersion (ν) beta distribution is defined by setting $\nu = p + q$. A random variable Y follows a $\mu\nu$-beta distribution if it density function is given by:

$$f(y|\mu,\nu)=\frac{\Gamma(\nu)}{\Gamma(\mu\nu)\Gamma(\nu(1-\mu))}y^{\mu\nu-1}(1-y)^{\nu(1-\mu)-1}I_{(0,1)}(y), \tag{8}$$

where $0<\mu<1$ and $\nu>0$. In this parameterization $E(Y)=\mu$ and $\text{Var}(Y)=\sigma^2=\frac{\mu(1-\mu)}{1+\nu}$. Hence, $0<\sigma^2<1/4$. This parameterization, proposed in Jorgensen (1997) and in Cepeda-Cuervo (2001), p. 63, was presented in (Ferrari & Cribari-Neto, 2004).

Assuming the mean and variance parameterization of the beta distribution proposed in Cepeda-Cuervo (2015), the mean and variance beta density function is given by (9), where $K=\frac{\Gamma(\frac{\mu(1-\mu)}{\sigma^2}-1)}{\Gamma(\frac{\mu^2(1-\mu)}{\sigma^2}-\mu)\Gamma(\frac{\mu(1-\mu)^2}{\sigma^2}-(1-\mu))}$.

$$f(y|\mu,\sigma^2)=Ky^{\frac{\mu^2(1-\mu)}{\sigma^2}-\mu-1}(1-y)^{\frac{\mu(1-\mu)^2}{\sigma^2}-(1-\mu)-1}I_{(0,1)}(y), \tag{9}$$

This parameterization of the beta density function can be obtained from (8), by setting $\nu=\frac{\mu(1-\mu)}{\sigma^2}-1$. Thus, the beta density function, written as a function of the precision parameter $\phi=1/\sigma^2$, is given by:

$$f(y|\mu,\sigma^2)=Ky^{\mu^2(1-\mu)\phi-\mu-1}(1-y)^{\mu(1-\mu)^2\phi-(1-\mu)-1}I_{(0,1)}(y), \tag{10}$$

where $K=\frac{\Gamma(\mu(1-\mu)\phi-1)}{\Gamma(\mu^2(1-\mu)\phi-\mu)\Gamma(\mu(1-\mu)^2\phi-(1-\mu))}$.

1.2. $\mu\sigma^2$-Beta Binomial Distribution

The beta binomial distribution assumes that the random variable Y, conditional on π^*, has binomial distribution $Bin(m,\pi^*)$ and that π^* follows the beta distribution defined by (5). The beta binomial probability is given by:

$$f(y|n,p,q)=\binom{n}{y}\frac{B(y+p,n-y+q)}{B(p,q)}I_A(y),\quad p>0,q>0, \tag{11}$$

where $A=\{0,1,2,\ldots,n\}$ and $B(.)$ is the beta function. If Y follows the beta binomial distribution defined by (11), $V(Y)=m\pi(1-\pi)[1+\rho(m-1)]$, where $\rho=1/(1+p+q)$. As a consequence, ρ can be interpreted as a dispersion parameter, in the sense that for constant mean, the variance of Y increases when ρ increases and the variance Y decreases when ρ decreases. In this parameterization of the beta-binomial distribution, the variance of Y is $1+\rho(m-1)$ times the variance of the binomial model. In this distribution, if the random variable Y, conditional on π^*, has binomial distribution $Bin(m_i,\pi^*)$ and π^* follows a $\mu\sigma^2$-beta distribution with $\text{E}(\pi^*)=\pi$ and $\text{Var}(\pi^*)=\sigma^2$, the unconditional distribution of Y is given by:

$$f(y|n,\pi,\sigma^2)=\binom{n}{y}\frac{B(y+\mu\nu,n-y+\nu(1-\mu))}{B(\mu\nu,\nu(1-\mu))}I_A(y), \tag{12}$$

where $\nu=\frac{\mu(1-\mu)}{\sigma^2}-1$ (or $\nu=\mu(1-\mu)\phi-1$). In this case, Y is said to follow a $\pi\sigma^2$-beta binomial distribution with mean and variance (precision) given by:

$$
\begin{aligned}
E(Y) &= E(E(Y|\pi^*)) = mE(\pi^*) = m\pi \quad \text{and} \\
V(Y) &= E[Var(Y|\pi^*)] + Var[E(Y|\pi^*)] \\
&= mE[\pi^*(1-\pi^*)] + m^2 Var(\pi^*) \\
&= m\pi(1-\pi) + \sigma^2 m(m-1) \qquad (13)
\end{aligned}
$$

Thus, in model (12), $\sigma^2 = Var(\pi^*)$ and $\sigma^2 m(m-1)$ is the overvariance of the beta binomial distribution. If $\sigma^2 > 0$, there is overvariance, while if $\sigma^2 = 0$, the $\pi\sigma^2$-beta binomial distribution is reduced to the binomial distribution. If $\sigma^2 = \rho\pi(1-\pi)$ in (13), the variance of the beta binomial distribution (11) is obtained. In this new parameterization of the beta binomial distribution σ^2 has clear interpretation, given that it is the variance of π^*.

2. Beta Regression Models

A beta regression model was proposed by Cepeda-Cuervo (2001), with joint modeling of the mean (μ) and dispersion ($\nu = p + q$) parameters of the beta distribution, assuming that the appropriate function of both follows a linear regression structure. Under a general framework, he assumed a random sample $Y_i \sim Beta(p_i, q_i)$, $i = 1, 2, \ldots, n$, with mean and dispersion regression structures given by:

$$
\begin{aligned}
\mathrm{h}(\mu_i) &= \mathbf{x}_i^t \boldsymbol{\beta} \qquad (14) \\
\mathrm{g}(\nu_i) &= \mathbf{z}_i^t \boldsymbol{\gamma}, \qquad (15)
\end{aligned}
$$

where h is the logit function and g is the logarithmic function; $\boldsymbol{\beta} = (\beta_0, \beta_1, \ldots, \beta_k)^t$ and $\boldsymbol{\gamma} = (\gamma_0, \gamma_1, \ldots, \gamma_p)^t$ are the vectors of the mean and dispersion regression parameters structures, respectively; $\mathbf{x}_i$ is the vector of the mean explanatory variables; and $\mathbf{z}_i$ is the vector of precision explanatory variables, at the i-th observation. This model implies a parameterizations of the beta density function (5) in terms of μ_i and ν_i.

In subsequent paper, Ferrari & Cribari-Neto (2004) presented a frequentist approach to the beta regression models, assuming that $h(\mu_i) = \mathbf{x}_i^t \boldsymbol{\beta}$, where h is an appropriate real valued function, strictly monotonic and twice differentiable, defined on $(0, 1)$, and ϕ, the dispersion parameter, is a constant in the range of values of the explanatory variables. The joint beta regression model proposed by Cepeda-Cuervo (2001), was later studied by (Smithson & Verkuilen, 2006), using Bayesian methods, and by Simas et al. (2010), from a frequentist perspective. Multiple studies and applications have been developed in recent years in the framework of beta regression models. Although many variations of mean and dispersion beta regression models have been developed since the work of Cepeda-Cuervo (2001), these proposals have at least two drawbacks. The first is the interpretability of ν in the $\mu\nu$-beta regression model (8), given that ν is considered to be a "precision" parameter in the sense that, for constant mean, the variance decreases when ν increases, $\sigma^2 = \frac{\mu(1-\mu)}{\nu+1}$. A second problem is the lack of an explicit regression structure for the variance, which impair the quality of the posterior regression parameter inferences.

2.1. $\mu\sigma^2$-Beta Regression Models

A first approximation of the mean and variance beta regression models was proposed in (Cepeda-Cuervo, 2015), from a Bayesian perspective. In that model, the mean regression structure is defined as a function of the explanatory variables as in equation (14), where h is, for example, the logit function. In that paper, the variance regression structure was defined as a function of the explanatory variables as $g(\sigma_i^2) = \mathbf{z}_i^t\boldsymbol{\lambda}$, where $\boldsymbol{\lambda} = (\lambda_0, \lambda_1, \ldots, \lambda_s)^t$ is the vector of variance regression parameter structure and g is assumed to be the logarithmic function. However, given that $0 < \sigma^2 < 1/4$, posterior samples of $\boldsymbol{\lambda}$ should be obtained from restricted parameter spaces.

In this paper, from the new parameterization of the beta distribution given in (9), we propose the mean and variance beta regression models by defining appropriate variance link functions, in order to obtain regression parameter estimates from unrestricted parameter spaces. Thus, the parameter estimates can be obtained by applying Bayesian or maximum likelihood methods without any parameter restrictions, if the link functions are once and twice differentiable, respectively.

In the $\mu\sigma^2$ beta regression models, the mean regression structure is given by (14), while for the variance, the proposed the regression structure is given by:

$$g(4\sigma_i^2) = \boldsymbol{z}^t\boldsymbol{\lambda}, \tag{16}$$

where $h(.)$ and $g(.)$ are real functions defined in the open interval $(0, 1)$ of the real number set R, like the logit, probit, log-log or the complementary log-log function. If, for example, $\text{logit}(4\sigma^2) = \boldsymbol{z}^t\boldsymbol{\lambda}$, the variance of the beta regression model is given by: $\sigma^2 = \exp(\boldsymbol{z}^t\boldsymbol{\lambda})/(4 + 4\exp(\boldsymbol{z}^t\boldsymbol{\lambda}))$. Thus, the parameter estimates of the mean and variance regression structures are easily interpretable. For example, if the logit link function is assumed for the mean and variance regression structure and:

1. X_1 is an explanatory variable associated with a parameter β_1 where $\beta_1 > 0$, increasing behavior of X_1 is associated with an increasing mean, while if $\beta_1 < 0$, increasing behavior of X_1 is associated with a decreasing mean.

2. Z_1 is an explanatory variable associated with a parameter λ_1 where $\lambda_1 > 0$, increasing behavior of Z_1 is associated with increasing variance, while if $\lambda_1 < 0$, increasing behavior of Z_1 is associated with decreasing of variance.

The mean and precision ($\phi = 1/\sigma^2$) beta regression model is defined by the mean regression structure (14) and by $g(\phi - 4) = \boldsymbol{z}^t\boldsymbol{\lambda}$, where $g(.)$ is the logarithmic function or some other appropriate real function defined from the positive real numbers $\Re^+$ to the real numbers $\Re$, such as the logarithmic function.

2.2. $\mu\sigma^2$-Beta Binomial Regression Models

In this section, the $\mu\sigma^2$-beta binomial regression models are defined by assuming that $Y_i \sim BB(m_i, \pi_i, \sigma_i^2)$, $i = 1, \ldots, n$, are n independent random variables that follow the beta binomial distribution defined in (12). Thus, in the $\mu\sigma^2$-beta binomial regression models, it is assumed that the random variables Y_i have, conditional on π^*, binomial distribution $Bin(m_i, \pi_i^*)$, and that π_i^* follows the beta distribution defined by (9), with:

$$\mathrm{E}(\pi_i^*) = \mu_i, \quad \text{and} \qquad \mathrm{Var}(\pi_i^*) = \sigma_i^2$$

The systematic and link function components are defined by $h(\pi_i) = \boldsymbol{x}_i^t\boldsymbol{\beta}$ and $g(\sigma_i^2) = \boldsymbol{z}_i^t\boldsymbol{\lambda}$, assuming the probability (mean) regression structures given by (14) and variance regression structures defined in (16). The use of this beta-binomial regression model is recommended when the researcher believes that the data come from a population having different subpopulations, and also when there is correlation between the Bernoulli events within each binomial observation (Quintero-Sarmiento et al., 2012). This assumption allow us to have a larger variance than the onethat considered in the GLM with binomial response variable, but assuming that the overvariance follows regression structures as functions of the explanatory variables.

3. Applications

In this section, posterior inferences of the $\mu\sigma^2$-beta and $\pi\sigma^2$-beta binomial regression models are presented in the framework of two applications. Section 3.1 includes results of analyzing reading accuracy dataset of dyslexic and non dyslexic students with $\mu\sigma^2$-beta regression models. These results illustrate the good performance of the proposed models and the advantages of their interpretations against the usual beta mean and dispersion regression models. Section 3.2 presents results of applying $\pi\sigma^2$-beta binomial regression models to a postnatal mortality dataset and their interpretations.

3.1. Dyslexic versus Non-Dyslexic Reading Accuracy

This section presents results of applying $\mu\sigma^2$-beta regression models to the analysis of the reading accuracy data for dyslexic and non-dyslexic Australian children (Smithson & Verkuilen, 2006). The variable of interest is accuracy, measured by scores on a test of reading accuracy taken by 44 children, which is predicted by two regressors variables: dyslexia (a dichotomic variable which has value 1 for dyslexic children and -1 for non-dyslexic children) and nonverbal intelligence quotient (IQ), converted to standardized Z scores. The sample includes 19 dyslexic and 25 controls recruited from primary schools in the Australian Capital Territory. The children's ages ranged from eight years five months to twelve years three months. This dataset was analyzed by Smithson & Verkuilen (2006), using mean and "dispersion" beta regression models and applying the Fisher scoring algorithm to fit the models. In this paper, we initially analyze this data set by applying the Bayesian beta regression model proposed by Cepeda-Cuervo (2001) as in Smithson & Verkuilen (2006), assuming that:

$$\text{logit}(\mu_i) = \beta_0 + \beta_1 DIS + \beta_2 IQ + \beta_3 DISIQ \quad (17)$$
$$\log(\nu_i) = \gamma_0 + \gamma_1 DIS + \gamma_2 IQ. \quad (18)$$

After that, in order to apply the $\mu\sigma^2$-beta regression models, we assume a beta regression model where the mean and variance regression structures are given by (17) and (19), respectively.

$$\text{logit}(4\sigma^2) = \lambda_0 + \lambda_1 DIS + \lambda_2 IQ. \quad (19)$$

The posterior parameter estimates of the beta regression model defined by (17) and (18) are given in Table 1. According to these estimates, the mean of the score reading decreases with IQ for dyslexic children and increases for non dyslexic children, given that $\text{logit}(\hat{\mu}_i) = (\hat{\beta}_0 + \hat{\beta}_1) + (\hat{\beta}_2 + \hat{\beta}_3)IQ$ for dyslexic children and $\text{logit}(\hat{\mu}_i) = (\hat{\beta}_0 - \hat{\beta}_1) + (\hat{\beta}_2 - \hat{\beta}_3)IQ$ for their non-dyslexic counterparts. The so called "dispersion" parameter increases with IQ and is lower for non dyslexic children, but with the interpretation of the dispersion as in Section 2, for fixed values of μ. For this model, the deviance information criterion (DIC) value is -117.5. The error sum of squar $SS = 0.4403$. The error sum of square for the mean and precision beta regression model defined by (17) and (18), obtained from the parameter estimates reported by Cribari-Neto & Zeileis (2010), is SS=0.4618. Thus, the parameter estimates and the results obtained by fitting this model via Bayesian methods are consistent with that obtained by Smithson & Verkuilen (2006) for this dataset. In both cases, the null hypothesis $H: \beta_i = 0$, i=1,2,3, and $H: \gamma_i = 0$, j=1,2, are rejected at a level of 95%.

Table 1: Parameter estimates, standard deviations, 95% credible intervals and DIC values for of mean and "dispersion" beta regression parameters as defined by (17) and (18).

logit-mean and log-dispersion $log(\nu)$			
Parameter	Estimate	S.D.	95% C.I.
β_0	1.129	0.156	(0.813, 1.425)
β_1	-0.747	0.155	(-1.036, -0.429)
β_2	0.428	0.174	(0.079, 0.768)
β_3	-0.515	0.181	(-0.866, -0.146)
γ_0	3.104	0.250	(2.575, 3.555)
γ_1	1.583	0.305	(0.963, 2.156)
γ_2	0.997	0.444	(0.099, 1.839)
	Dhat: -130.8	DIC: -117.5	

Table 2 includes the parameter estimates of the $\mu\sigma^2$-beta regression model, assuming the logit-mean and logit-variance regression structures given by (17) and (19), respectively. The mean parameter estimates agree for these two beta regression models, as seen by comparing Table 1 and Table 2, in the sense that β_0 and β_2 are positive and, β_1 and β_3 are negative. However, the posterior parameter inferences are very different. In the mean and variance model, the null hypothesis

$\beta_2 = 0$ and $\lambda_2 = 0$ are not rejected at a level of 0.05, given that zero belongs to the 95% credible intervals.

TABLE 2: Parameter estimates, standard deviations, 95% credible intervals for the parameter estimates and DIC values for the $\mu\sigma^2$-beta regression model defined by (17) and (19).

logit-mean and logit-variance			
Parameter	Estimates	S.D.	95% C.I.
β_0	1.208	0.141	(0.926, 1.490)
β_1	-0.830	0.146	(-1.133,-0.537)
β_2	0.215	0.128	(-0.025, 0.452)
β_3	-0.298	0.135	(-0.546, -0.018)
λ_0	-3.596	0.276	(-4.091,-3.023)
λ_1	-0.928	0.344	(-1.632,-0.205)
λ_2	-0.801	0.428	(-1.560, 0.056)
	$Dhat = -128.7$	$DIC = -114.5$	

Figure 1 is the plot of variance versus Dyslexia and IQ, obtained by fitting the $\mu\sigma^2$-beta regression model to the dyslexia dataset, where the variance follows the regression structure given by (19). This plot shows that variance is bigger for non-dyslexic children than for dyslexic ones, and that it decreases when IQ increases. Figure 2 is the plot of the residuals of the logit-mean and logit-variance beta regression models. The plots of the residuals are very similar to those of the mean and "dispersion" beta regression models. From these plots, it is not possible to find differences between residuals of these two regression models. However, the sum of squared residuals for the logit-mean and logit-variance regression models ($SS = 0.4044$) is smaller than the sum of squared residuals for the logit-mean and logarithmic dispersion regression models ($SS = 0.4052$).

From these results, we applied a variable elimination process in the mean and variance regression structures, obtaining the beta regression models given by (20) and (21), where IQ and DISIQ were eliminated from the mean regression structure and IQ from the variance models.

$$\text{logit}(\mu_i) = \beta_0 + \beta_1 DIS \tag{20}$$

$$\text{logit}(4\sigma^2) = \lambda_0 + \lambda_1 DIS \tag{21}$$

For this model, given by equations (20) and (21), the parameter estimates and their respective standard deviations are: $\hat{\beta}_0 = 1.300(0.137)$, $\hat{\beta}_1 = -0.872(0.136)$, $\hat{\lambda}_0 = -3.38(0.289)$, $\hat{\lambda}_1 = -0.642(0.292)$; the hypotheses $\beta_i = 0$ and $\lambda_i = 0$, $i = 0, 1$, are rejected at a level of 95%; their DIC value is equals -114.4 and the sum of square residuals equal to 0.4466. Finaly, under the DIC criterion, taking into account the 95% credibility intervals, this model is selected as the best.

Thus, the results obtained by applying $\mu\sigma^2$-beta regression models shows a big difference with that obtained from analysis of the dyslexic dataset by applying mean and dispersion regression models, as developed at the beginning of this section, or by Smithson & Verkuilen (2006) by applying ML estimation.

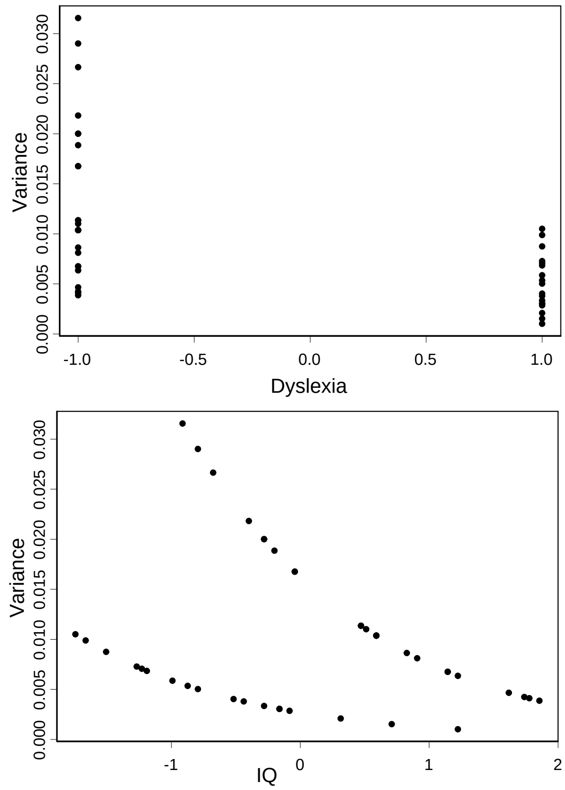

FIGURE 1: Variance behavior in the logit-mean and logit-variance beta regression models defined by (17) and (19).

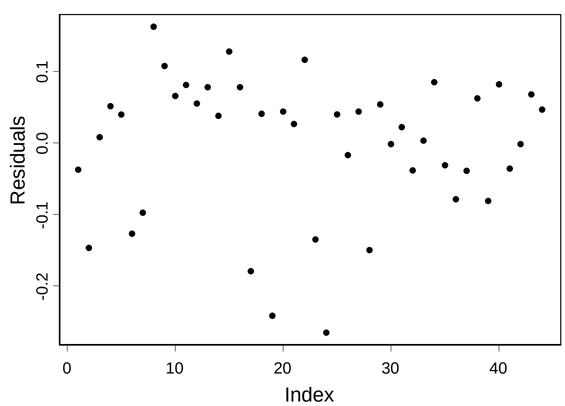

FIGURE 2: Residuals of logit-mean and logit-variance beta regression models.

3.2. School Absenteeism Dataset

The second dataset analyzed in this paper was originally presented in Quine (1975) and comes from a sociological study of Australian Aboriginal and White children from Walgett, New South Wales, with nearly equal numbers of the two sexes and equal numbers from the two cultural groups. Children were classified by culture, age, sex and learner status, and the number of days absent from school in a particular school year was recorded. The response variable of interest is the number of days that a child was absent from school during the schol year. Children who had suffered a serious illness during the year were excluded from this analysis.

The values for each observed variable were:

- **Age** (A). Four groups: Primary (0), First form (1), Second form (2) and Third form (3), where "Primary" includes children under 12 years of age; "First form" refers to ages between 12 and 13 years; "Second form" denotes ages between 13 and 14 years, and "Third form", represents ages between 14 and 15 years.
- **Gender** (G). Factor with two levels: Female (0) or Male (1).
- **Cultural Background**(CB). Ethnic background: Aboriginal (0) and White (1).
- **Learning Ability**(LA). Factor with two levels of learning capacity: Slow learner (0) and Average learner (1).
- **Days Absent**(Y): days absent from school in the year.

Since the variable of interest Y (days absent) is the number of events that occurred during a year, this dataset was analyzed by applying overdispersed models associated with counts (Cepeda-Cuervo & Cifuentes-Amado, 2017), assuming that it follows a negative binomial distribution $NB(\mu, \alpha)$, where the mean and the shape parameters follow linear regression structures. In this paper, assuming a schol year of 200 days, this dataset is analyzed by applying the $\mu\sigma^2$-beta binomial regression model, where:

$$\text{logit}(\mu_i) = \beta_0 + \beta_1 A_i + \beta_2 G_i + \beta_3 CB_i + \beta_4 LA_i \quad (22)$$

$$\text{logit}(4\sigma_i^2) = \lambda_0 + \lambda_1 CB_i + \lambda_2 LA_i. \quad (23)$$

This model was fitted to this dataset by applying Bayesian methods using the Open Bugs software. The posterior parameter estimates, standard deviations and credible intervals are given in Table 3, Model 1a. Assuming a mean structure (22) without Age (A) and Gender (G) as explanatory variables, eliminated from the model at a level of 95%, and the shape structure given by (23), the parameter estimates, standard deviations and 95% credible intervals are reported in the same Table, Model 1b. The DIC values for the $\mu\sigma^2$-beta binomial regression model defined by (22) and (23) is $DIC = 649.272$ and the DIC value of the reduced model is $DIC = 649.004$. In this case, Model 1b is assumed to be the best, given

TABLE 3: Parameter estimates of the (mean and variance) beta-binomial regression models. For Model 1a, $DIC = 649.273$. For Model 1b, $DIC = 649.004$.

	Model 1a			Model 1b		
Parameter	Estimate	S.D.	Cred.Interval	Estimate	S.D.	Cred.Interval
β_0	-2.212	0.237	(-2.707,-1.773)	-1.929	0.146	(-2.211,-1.64)
β_1	0.123	0.092	(-0.049,0.312)	–	–	–
β_2	-0.009	0.127	(-0.261,0.240)	–	–	–
β_3	-0.789	0.181	(-1.137,-0.429)	-0.777	0.188	(-1.154,-0.419)
β_4	-0.401	0.189	(-0.765,-0.020)	-0.444	0.170	(-0.785,-0.112)
λ_0	-3.261	0.293	(-3.805,-2.666)	-3.201	0.292	(-3.732,-2.588)
λ_1	-0.825	0.404	(-1.610, -0.009)	-0.866	0.401	(-1.670,-0.093)
λ_2	-0.997	0.422	(-1.774,-0.131)	-1.001	0.383	(-1.758,-0.238)

that it has the lower DIC value and all the null hypotheses of its parameters are rejected.

To compare the performance of the proposed model with the mean and dispersion beta binomial regression models, the beta binomial model obtained by assuming that $Y \sim Bin(m, \pi^*)$, where π^* follows the beta distribution defined by (8), was fitted to this dataset, assuming the mean regression structure (22) and dispersion regression structure given by:

$$\log(\nu_i) \quad = \quad \gamma_0 + \gamma_1 CB_i + \gamma_2 LA_i \tag{24}$$

Their parameter estimates, standard deviations and 95% credible intervals obtained from the application of Open Bugs are given in Table 4, Model 2a. Assuming the mean regression structure (22) without the explanatory variables Age and Gender, eliminated from the model at a level of 95%, and the "dispersion" structure given by (24), the parameter estimates, standard deviations and 95% credible intervals are reported in the same Table, Model 2b. The DIC values for the beta binomial model (11) with regression structures defined by (22) and (23) is $DIC = 650.4$. The DIC value of the reduced model is $DIC = 649.7$.

TABLE 4: Parameter estimates of the beta-binomial (mean and dispersion) regression model. For Model 2a, $DIC = 650.4$. For Model 2b, $DIC = 649.7$.

	Model 2a			Model 2b		
Parameter	Estimate	S.D.	Cred.Interval	Estimate	S.D.	Cred.Interval
β_0	-2.325	0.307	(-2.927,-1.723)	-1.939	0.151	(-2.232,-1.640)
β_1	0.185	0.115	(-0.040, 0.407)	–	–	–
β_2	-0.063	0.163	(0.382, 0.261)	–	–	–
β_3	-0.747	0.191	(-1.119,-0.367)	-0.759	0.186	(-1.121,-0.394)
β_4	-0.415	0.190	(-0.788,-0.039)	-0.446	0.187	(-0.813,-0.074)
γ_0	2.419	0.252	(1.901, 2.889)	2.346	0.248	(1.842, 2.825)
γ_1	0.055	0.345	(-0.640, 0.711)	0.209	0.321	(-0.416, 0.825)
γ_2	0.680	0.329	(0.029, 1.322)	0.653	0.327	(0.002, 1.292)

Finally, given that in Model 2b, zero belongs to the credible interval of γ_1, we assumed a model with the mean regression structure (22), without age and gender variables, and a overdispersion regression structure given by (24), without cultural background ($\text{logit}(\pi_i) = \beta_0 + \beta_3 CB_i + \beta_4 LA_i$ and $\log(\nu_i) = \gamma_0 + \gamma_2 LA_i$). For this model, the posterior parameter estimates, standard deviations and 95%

credible intervals are reported in Table 5, model 3. For this model, the DIC value is $DIC = 648.0$. Of these mean and "dispersion" beta binomial regression models, Model 3 in the best. This model has the lowest DIC value and all the null hypotheses for the regression parameters are rejected at a level of 95%.

Table 5: Parameter estimates of the beta-binomial (mean and dispersion) regression model. For Model 3, $DIC = 648.0$.

	Model 3		
Parameter	Estimate	S.D.	Cred.Interval
β_0	-1.951	0.145	(-2.237,-1.662)
β_3	-0.709	0.163	(-1.027, -0.389)
β_4	-0.460	0.184	(-0.823,-0.098)
γ_0	2.421	0.215	(1.988, 2.822)
γ_2	0.683	0.325	(0.044, 1.327)

The mean parameter estimations, although slightly different, agree in the sense that the estimates belong to the respective credible intervals. However, Cultural Background is an explanatory variable of the overdispersion variance regression structure (Model 1b) that does not belongs to the overdispersion regression structure (Model 3). Of these two models, Model 1b is more appropriate given that it has smaller sum of square errors (the sum of square error is $SS_v = 245.138$ for Model 1b, and $SS_p = 245.885$ for model 3) and it has better interpretation of the regression parameters.

4. Conclusion

In this paper new parameterizations of the beta and beta-binomial distributions are proposed in terms of the mean and variance parameters. From these new parameterizations new beta and beta-binomial regression models are proposed by assuming that appropriate functions of the mean and variance (precision) parameters follow regression structures. The new regression models, improve the parameter interpretation and the posterior parameter inferences.

From the new parameterizations of the beta distributions, a new parameterization of the beta-binomial distribution is proposed by assuming that $Y \mid n, \pi^*$ follows a binomial distribution, where π^* follows the(μ, σ^2) beta distribution. From this distribution, the beta binomial regression model is proposed by assuming the mean and variance (precision) beta regression structures. The proposed $\mu\sigma^2$ beta and beta binomial regression models are fitted respectively to the School Absenteeism dataset and the Dyslexia Reading dataset. These applications illustrate the good performance of the proposed models, in addition to the advantage of the interpretation of the σ^2 related to the interpretation of $\nu = p + q$ in beta regression models. In both models, the regression parameter estimates are obtained by applying Bayesian methods using the OpenBugs software.

Many extensions of this paper can be developed. Appendix A proposes mean and variance working variables to define a Bayesian algorithm, like that proposed in Cepeda-Cuervo (2001), to obtain the posterior regression parameter estimates.

The respective frequentist parameter estimates can be obtained by applying maximum likelihood methods to define and develop a Fisher scoring algorithm. Based on the advantages of the mean and variance (precision) parameterization of the beta distribution, a new parameterization of the tilted beta and tilted beta-binomial distributions can also be proposed.

[Received: October 2022 — Accepted: December 2022]

References

Atkinson, A. B. (1970), 'On the measurement of inequality', *Journal of Economic Theory* **3**(32), 244–263.

Cepeda-Cuervo, E. (2001), *Modelagem da Variabilidade em Modelos Lineares Generalizados*, Unpublished Ph.D. thesis, Mathematics Institute, Universidade Federal do Rio de Janeiro.
URL: http://www.docentes.unal.edu.co/ecepedac/docs/TESISFINAL.pdf.

Cepeda-Cuervo, E. (2015), 'Beta regression models: Joint mean and variance modeling', *Revista Colombiana de Estadística* **9**(1), 134–145.

Cepeda-Cuervo, E. & Cifuentes-Amado, M. V. (2017), 'Double generalized beta-binomial and negative binomial regression models', *Revista Colombiana de Estadística* (40), 141–163.

Cepeda-Cuervo, E. & Garrido, L. (2015), 'Bayesian beta regression models with joint mean and dispersion modeling', *Monte Carlo Methods and Applications* **21**(1), 49–58.

Cepeda-Cuervo, E. & Núñez Antón, V. (2013), 'Spatial double generalized beta regression models: extensions and application to study quality of education in Colombia', *Journal of Educational and Behavioral Statistics* **6**(38), 604–628.

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

Cepeda, E. & Gamerman, D. (2005), 'Bayesian methodology for modeling parameters in the two parameter exponential family', *Revista Estadistica* **57**(168-169), 93–105.

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

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

Hunger, M., Döring, A. & Holle, R. (2012), 'Longitudinal beta regression models for analyzing health-related quality of life scores over time', *BMC medical research methodology* **1**(12), 1–12.

Jorgensen, B. (1997), 'Proper dispersion models', *Brazilian Journal of Probability and Statistics* **11**(2), 89–128. http://www.jstor.org/stable/43600934

Paolino, P. (2001), 'Maximum likelihood estimation of models with beta-distributed dependent variables', *Political Analysis* **4**(9), 325–346.

Quine, S. (1975), Achievement orientation of aboriginal and white Australian adolescents, Ph.D. thesis, Australia.

Quintero-Sarmiento, A., Cepeda-Cuervo, E. & Nunez-Anton, V. (2012), 'Estimating infant mortality in Colombia: some overdispersion modelling approaches', *Journal of Applied Statistics* **39**(5), 1011–1036.

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

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

Appendix A. Appendix

Appendix A.1. Bayesian Method

To propose a Bayesian method to fit the $\mu\sigma^2$-beta regression models, a prior distribution for $\boldsymbol{\theta} = (\boldsymbol{\beta}, \boldsymbol{\lambda})'$ should be specified. Following the Bayesian method proposed in Cepeda-Cuervo (2001), normal prior distribution should be assumed for the mean and variance regression parameters. Thus, given that the posterior distribution $\pi(\boldsymbol{\beta}, \boldsymbol{\lambda} \mid \text{data})$ is analytically intractable, samples of $\boldsymbol{\beta}$ and $\boldsymbol{\lambda}$ can be obtained from the full conditional posterior distributions, denoted by π_β and π_λ, respectively, updated by applying the Metropolis-Hastings algorithm.

This methods involves setting working variables to build kernel transition functions, established from the posterior distribution obtained from the combination of a normal working model and the normal prior distributions. When the mean link is assumed to be the logit function, the mean working variable (A.1), presented in Cepeda-Cuervo (2001), Cepeda & Gamerman (2005) and Cepeda-Cuervo & Garrido (2015), is used to build the mean kernel transition function.

$$\tilde{Y}_i = \boldsymbol{x_i^t}\boldsymbol{\beta^{(c)}} + \frac{Y_i - \mu_i^{(c)}}{\mu_i^{(c)}(1-\mu_i^{(c)})}, \quad i = 1, \dots n. \tag{A.1}$$

The kernel transition function is obtained from a combination of the prior distribution and the working observational models obtained by assuming that the working variable (A.1) follows a normal prior distribution. To obtain samples of the posterior conditional distribution π_λ, when $g(\sigma_i^2) = \text{logit}(4\sigma_i^2)$ is assumed as link function, the variance working variable is built from $t_i = (Y_i - \mu_i)^2$, a random variable t_i such that $E(t_i) = \sigma_i^2$, from the first order Taylor approximation of $\text{logit}(4t)$ around the current value $\sigma_i^{2(c)}$ of σ_i^2. Thus, the working variable is given by:

$$\tilde{Y}_i = \boldsymbol{z_i'}\boldsymbol{\lambda^{(c)}} + \frac{(Y_i - \mu_i^{(c)})^2) - \sigma_i^2}{4\sigma_i^2(1-4\sigma_i^2)} - 1, \quad i = 1, \ldots n, \tag{A.2}$$

and the kernel transition functions, as for the mean regression parameters, is obtained from the combining the conditional prior distribution and the working observational model obtained by assuming that the working observational variable (A.2) follows a normal distribution.

Appendix A.2. Maximum Likelihood Parameter Estimates

Given that in the $\mu\sigma^2$-beta regression models, the likelihood function is given by $L(\beta, \lambda) = \prod_{i=1}^n L(\mu_i, \sigma_i^2)$, where $L(\mu_i, \sigma_i^2)$ is given by (9), $0 < \mu_i < 1$ and $0 < \sigma_i^2 < \frac{1}{4}$, $h(\mu_i) = \eta_{1i}$, $g(4{\sigma_i}^2) = \eta_{2i}$, $\eta_{1i} = \boldsymbol{x}^t\boldsymbol{\beta}$ and $\eta_{2i} = \boldsymbol{z}^t\boldsymbol{\lambda}$. Thus, the first-order derivatives of the i-th components of $\ell(\beta, \gamma) = \log(L(\beta, \gamma))$ are given by:

$$\frac{\partial \ell_i}{\partial \beta_r} = \frac{\partial \ell_i}{\partial \mu_i}\frac{\partial \mu_i}{\partial \eta_{1i}}\frac{\partial \eta_{1i}}{\partial \beta_r} \tag{A.3}$$

$$\frac{\partial \ell_i}{\partial \lambda_s} = \frac{\partial \ell_i}{\partial \sigma_i^2}\frac{\partial \sigma_i^2}{\partial \eta_{2i}}\frac{\partial \eta_{2i}}{\partial \lambda_s}, \tag{A.4}$$

The second-order derivatives of the logarithm of the likelihood function should be obtained:

$$\frac{\partial^2 \ell_i}{\partial \beta_s \partial \beta_r} = \left(\frac{\partial^2 \ell_i}{\partial \beta_s \partial \mu_i}\right)\frac{\partial \mu_i}{\partial \eta_{1i}}\frac{\partial \eta_{1i}}{\partial \beta_r} + \frac{\partial \ell_i}{\partial \mu_i}\frac{\partial}{\partial \beta_s}\left(\frac{\partial \mu_i}{\partial \eta_{1i}}\frac{\partial \eta_{1i}}{\partial \beta_r}\right) \tag{A.5}$$

$$\frac{\partial^2 \ell_i}{\partial \lambda_s \partial \beta_r} = \left(\frac{\partial^2 \ell_i}{\partial \lambda_s \partial \mu_i}\right)\frac{\partial \mu_i}{\partial \eta_{1i}}\frac{\partial \eta_{1i}}{\partial \beta_r} + \frac{\partial \ell_i}{\partial \mu_i}\frac{\partial}{\partial \lambda_s}\left(\frac{\partial \mu_i}{\partial \eta_{1i}}\frac{\partial \eta_{1i}}{\partial \beta_r}\right) \tag{A.6}$$

$$\frac{\partial^2 \ell_i}{\partial \lambda_k \partial \lambda_s} = \left(\frac{\partial \ell_i}{\partial \lambda_k \partial \sigma_i^2}\right)\frac{\partial \sigma_i^2}{\partial \eta_{2i}}\frac{\partial \eta_{2i}}{\partial \lambda_s} + \frac{\partial \ell_i}{\partial \sigma_i^2}\frac{\partial}{\partial \lambda_k}\left(\frac{\partial \sigma_i^2}{\partial \eta_{2i}}\frac{\partial \eta_{2i}}{\partial \lambda_s}\right) \tag{A.7}$$

Thus, the maximum likelihood estimates of $\boldsymbol{\beta}$ and $\boldsymbol{\gamma}$ can be obtained, for example, by applying the Newton-Raphson algorithm.