Communications for Statistical Applications and Methods
2024, Vol. 31, No. 3, 263–277
https://doi.org/10.29220/CSAM.2024.31.3.263
Print ISSN 2287-7843 / Online ISSN 2383-4757

Tilted beta regression and beta-binomial regression models: Mean and variance modeling

Edilberto Cepeda-Cuervo[1,a]

[a]Department of Statistics, Universidad Nacional de Colombia, Colombia

Abstract

This paper proposes new parameterizations of the tilted beta binomial distribution, obtained from the combination of the binomial distribution and the tilted beta distribution, where the beta component of the mixture is parameterized as a function of their mean and variance. These new parameterized distributions include as particular cases the beta rectangular binomial and the beta binomial distributions. After that, we propose new linear regression models to deal with overdispersed binomial datasets. These new models are defined from the proposed new parameterization of the tilted beta binomial distribution, and assume regression structures for the mean and variance parameters. These new linear regression models are fitted by applying Bayesian methods and using the OpenBUGS software. The proposed regression models are fitted to a school absenteeism dataset and to the seeds germination rate according to the type seed and root.

Keywords: count data, overdispersion, tilted beta distribution, binomial distribution, tilted beta binomial distribution, Bayesian approach

1. Introduction

The beta distribution is usually used to study continuous variables X that take values in an open interval (a, b), given that the random variable $Y = (X-a)/(b-a)$ takes values in the open interval $(0, 1)$ and can be assumed to have beta distribution, $B(p, q)$. Beta distribution appears in many applications, such as in the analysis of population growth, interest rates, disease incidence, and unemployment rates. In different fields, there is often a need to model continuous random variables that are assumed to follow the beta distribution as a function of a set of explanatory variables. For this type of analysis, Cepeda-Cuervo (2001) proposed the beta regression models, where the mean, $\mu = p/(p + q)$, and dispersion, $\nu = p + q$, parameters follow regression structures. These beta regression models were appropriately extended by Simas *et al.* (2010), assuming the regression structure to be nonlinear in the mean and in the dispersion parameters, $\nu = p + q$, who has provided valuable insights in the development of new research in recent years, especially using frequentist methods. Taking into account the conditional interpretation of ν, the mean and variance beta regression models are proposed by assuming that an appropriate function of the mean and variance parameters of the beta distribution follows a linear regression structure (Cepeda-Cuervo, 2015, 2023). These models improve the regression parameter inferences and interpretations.

In order to admit heavier tails in the beta distribution, Hahn (2008) proposed the beta rectangular distribution as a new distribution that, like the beta distribution, has the open interval $(0, 1)$ as domain.

[1]Corresponding author: Department of Statistics, Universidad Nacional de Colombia, Bogotá-Colombia, Carrera 45 No 26-85, Bogotá D.C. 111321, Colombia. E-mail: ecepedac@unal.edu.co

Published 31 May 2024 / journal homepage: http://csam.or.kr

The beta rectangular distribution consists of a convex combination between the beta distribution and the uniform distribution $U(0,1)$. Subsequently, Hahn and López Martín (2015) proposed the tilted beta distribution, that consists of a mixture of the beta distribution and the tilted distribution, which has as particular cases the beta rectangular distribution and the beta distribution. In this paper, from the mean and variance beta distribution, a new parameterization of the tilted beta distribution is proposed, where mean and variance beta distributions and beta rectangular distributions were applied to improve the above proposal by including the interpretation advantages of the mean and variance beta distributions.

In count datasets, it is often found that the variance of the response variable Y exceeds the theoretical variance of the binomial distribution. This phenomenon, known as extra-binomial variation, can lead to underestimation errors, lower efficiency of estimates and underestimation of the variance, which can in turn can generate incorrect inferences about the regression parameters or the credible intervals (Collet, 1991; Cox, 1983; Williams, 1982). Combining the beta distribution with the binomial distribution leads to the beta-binomial distribution. This distribution is normally used to model the number of successes obtained in a finite number of experiments, and to study overdispersed datasets.

There are several approaches to studying overdispersed binomial datasets. Hinde and Demetrio (1998) categorized the majority of overdispersed binomial models into two classes: (1) those in which a more general shape for the variance function is assumed by adding additional parameters; and (2) models in which it is assumed that the parameter p of the binomial distribution $\text{bin}(m, p)$ is itself a random variable. In the first class, the double exponential family of distributions allows researchers to obtain double binomial models. This enables the inclusion of a second parameter, which regardless of the mean,that controls for the variance of the response variable and can be modeled from a subset of some explanatory variables (Efron, 1986). In the second class, the beta binomial distribution assumes that the response variable follows a binomial distribution while the probability parameter follows a beta distribution, $B(a,b)$. When the beta distribution is parameterized in terms of its mean and the dispersion parameter (Cepeda-Cuervo, 2001), the beta binomial distribution is presented in terms of the mean and dispersion parameters as described by Cepeda-Cuervo and Cifuentes-Amado (2017).

To obtain a model with better flexibility from the beta binomial distribution, where the beta distribution assigns small probability tail values of p, Cepeda-Cuervo and Cifuentes-Amado (2017) proposed the tilted beta binomial distribution by assuming that p in the binomial distribution $\text{bin}(m, p)$ follows a tilted beta distribution, which, in turn, assumes that p follows a tilted mean and dispersion beta distribution. As a particular case of this distribution, these authors defined the beta rectangular binomial distribution by assuming that the parameter of the binomial distribution p has a beta rectangular distribution. This distribution allows for defining more general overdispersion regression models than that defined from the beta binomial regression model, which produces better estimates of the regression parameters, credibility or confidence intervals, and statistical inferences in the analysis of overdispersed binomial datasets. The tilted beta-binomial distribution was applied by Hahn (2022) in the analysis of overdispersed data, which assumes maximum likelihood and Bayesian methods. He applied this distribution to the analysis of the population dataset from the 2010 US Census. He found that the tilted beta-binomial distribution provided a better fit than the beta-binomial distribution. As he reported, the tilted beta-binomial distribution generalized the beta-binomial distribution, and it is capable of modeling datasets with greater overdispersion than the beta-binomial distribution.

We take into account the restricted interpretation of the dispersion parameter ν in which a constant mean increases when the variance decreases and decreases when the variance increases. In this paper, we propose to improve the binomial and the tilted binomial distributions, assuming that p in the binomial distribution follows the mean and variance beta distribution proposed in Cepeda-Cuervo

(2023). One advantage of this alternative versus the mean and dispersion regression models is that the interpretation of the dispersion parameter is only possible for fixed values of the mean (ν can be interpreted as a precision parameter in the sense that for the fixed values of μ, the variance decreases when ν increases), while in the proposed models, changes in the variance can be explained directly through changes in the explanatory variables corresponding to the variance regression structure. Thus, this paper proposes tilted beta binomial regression models where the mean and variance of the beta distribution, the mean of tilted distributions, and the mixture parameter follow regression structures.

This paper is organized as follows. After the introduction, in Section 2, the mean tilted distribution is presented. In Section 3, three parameterizations of the beta distribution are considered. In Section 4, the $\mu\sigma^2$-tilted beta binomial distribution is introduced and the $\mu\sigma^2$-rectangular beta binomial distribution is presented as a particular case. In Section 5, the $\mu\sigma^2$-tilted beta binomial distributions are defined. Section 6 presents a summary of the mean and variance beta regression models proposed by Cepeda-Cuervo (2023). In Section 7, the tilted beta binomial regression models are defined. Finally, Section 8, reports the results of two applications. Section 8.1 contains the results of analyzing a school absenteeism dataset by applying $\mu\sigma^2$-beta binomial regression models, and Section 8.2 descrribes the influence of the type of seed and root on the proportion of germinated seeds in each of 21 dishes by fitting a tilted beta binomial linear regression using the OpenBUGS software. The proposed model's performance is compared with the binomial and beta binomial regression models.

2. Tilted distribution

The tilted distribution was proposed by Hahn and López Martín (2015), and the following alternative definition was proposed by Cepeda-Cuervo and Cifuentes-Amado (2017): a random variable Y follows a tilted distribution with parameter ν if its density function is given by:

$$c\,(y \mid \upsilon) = \left[2\upsilon - 2\,(2\upsilon - 1)\,y\right] I_{(0,1)}(y), \quad 0 \le \nu \le 1. \tag{2.1}$$

The mean of Y, denoted by $\mu_t := E(Y|\nu)$, is $\mu_t = (2 - \upsilon)/3$. Thus, by parameterizing the density function (2.1) in terms of μ, this density function is given by:

$$c\,(y \mid \mu_t) = \left[3\,(2\mu_t - 1)\,(2y - 1) + 1\right] I_{(0,1)(y)}\ , \tag{2.2}$$

where $1/3 \le \mu_t \le 2/3$. The variance of a random variable Y that follows the density function (2.2) is given by $V_t(Y) = \mu_t(1 - \mu_t) - 1/6$. According to (2.2), it is clear that this distribution is equal to the uniform distribution when $\mu_t = 0.5$, is leaning to the right when μ_t is smaller than 0.5, and leaning to the left when μ_t is bigger than 0.5.

3. $\mu\sigma^2$-beta distribution

In this section, three parameterizations of the beta distributions are presented. A random variable Y follows a beta distribution if its density function is given by:

$$f_B\,(y \mid p, q) = \frac{\Gamma(p+q)}{\Gamma(p)\Gamma(q)} y^{p-1}(1-y)^{q-1} I_{(0,1)}(y)\ , \tag{3.1}$$

where $p > 0$, $q > 0$ and $\Gamma(\cdot)$ denotes the gamma function. The mean and variance of Y, $\mu_b = E(Y)$ and $\sigma_b^2 = \mathrm{Var}(Y)$, are respectively given by $\mu_b = p/(p+q)$ and

$$\sigma_b^2 = \frac{p\,q}{(p+q)^2(p+q+1)}. \tag{3.2}$$

From the beta density function (3.1), the mean (μ_b) and dispersion (ν) beta distribution (3.3) is defined, where $\nu = p + q$. A random variable Y follows a $\mu_b\nu$-beta distribution if its density function is given by:

$$f_B(y \mid \mu_b, \nu) = \frac{\Gamma(\nu)}{\Gamma(\mu_b\nu)\Gamma(\nu(1-\mu_b))} y^{\mu_b\nu-1}(1-y)^{\nu(1-\mu_b)-1} I_{(0,1)}(y). \tag{3.3}$$

This parameterizations of the beta distribution, presented in Ferrari and Cribari-Neto (2004), was already proposed in the literature, for example by Jorgensen (1997) and Cepeda-Cuervo (2001), p. 63. In this parameterization of the beta distribution, the variance of Y is given by $\sigma^2 = \mu(1-\mu)/(1+\nu)$. Thus, $\nu = p + q$ has an interpretation that for a fixed mean, the variance of Y increases when ν decreases and the variance of Y decreases when ν increases.

Finally, assuming the mean and variance parameterizations of the beta distribution, proposed in Cepeda-Cuervo (2015) and Cepeda-Cuervo (2023), the beta density function is given by (3.4), where $\phi = 1/\sigma^2$ and $K = \Gamma(\mu_b(1-\mu_b)\phi - 1)/(\Gamma(\mu_b^2(1-\mu_b)\phi - \mu_b)\Gamma(\mu_b(1-\mu_b)^2\phi - (1-\mu_b)))$. This formulation of the beta density function is proposed in Cepeda-Cuervo (2023).

$$f_B\left(y \mid \mu_b, \sigma_b^2\right) = K y^{\mu_b^2(1-\mu_b)\phi-\mu_b-1}(1-y)^{\mu_b(1-\mu_b)^2\phi-(1-\mu_b)-1} I_{(0,1)}(y). \tag{3.4}$$

The advantage of the density function (3.4) arises from the limited interpretation of the dispersion parameter $\nu = p + q$ in (3.3), while in equation (3.4), the precision parameter is given by $\phi = 1/\sigma^2$, where $\sigma^2 = \mathrm{Var}(Y)$.

4. Tilted beta distributions

This section presents a new parameterization of the tilted beta distribution proposed by Hahn and López Martín (2015), in terms of the mean (μ_b) and the variance (σ^2) parameters of the beta distribution and the mean of the tilted distribution μ_t. This new parameterization of the tilted beta distribution is obtained from the convex combination of the μ_t-tilted distribution proposed by Cepeda-Cuervo and Cifuentes-Amado (2017) and the $\mu_b\sigma^2$-beta distribution proposed by Cepeda-Cuervo (2023).

1. **Tilted $\mu\nu$-beta distribution**.

   The tilted beta distribution was introduced by Hahn and López Martín (2015), as a convex combination of the tilted and the beta distributions. In the $(\mu_t, \mu_b, \nu, \theta)$ parameterized form, this distribution was proposed in Cepeda-Cuervo and Cifuentes-Amado (2020) from a combination of the mean tilted distribution (2.2) and the mean and dispersion beta distribution (3.3). Thus, the $(\mu_t, \mu_b, \nu, \theta)$ density function of this distribution is given by:

   $$f(y \mid \mu_t, \mu_b, \nu, \theta) = \theta c(y \mid \mu_t) + (1-\theta) f_B(y \mid \mu_b, \nu), \tag{4.1}$$

   where $0 < y < 1$ and $0 \leq \theta \leq 1$. The notation $Y \sim \mathrm{TB}(\mu_t, \mu_b, \nu, \theta)$ is used to denote that Y follows this tilted beta parameterized distribution, with the mean and variance given by:

$$E(Y \mid \mu_t, \mu_b, \nu, \theta) = \theta\mu_t + (1-\theta)\mu_b \tag{4.2}$$

$$\begin{aligned} V(Y \mid \mu_t, \mu_b, \nu, \theta) &= E\left(Y^2 \mid \mu_t, \mu_b, \nu, \theta\right) - E^2(Y \mid \mu_t, \mu_b, \nu, \theta) \\ &= \left[\theta E_t\left(Y^2\right) + (1-\theta)E_b\left(Y^2\right)\right] - [\theta\mu_t + (1-\theta)\mu_b]^2 \\ &= \theta V_t(Y) + (1-\theta)V_b(Y) + \theta(1-\theta)(\mu_t + \mu_b)^2, \end{aligned} \tag{4.3}$$

where $E_t(Y^2)$ and $V_t(Y)$ denote the expectation of Y^2 and the variance of Y, by assuming that Y follows the tilted distribution, and $E_b(Y^2)$ and $V_b(Y)$ denote the expectation of Y^2 and the variance of Y, by assuming that Y follows the beta distribution (3.3).

The rectangular beta distribution is a particular case of (4.1) when $\mu_t = 0.5$ (the slope of the tilted distribution is zero). Thus, the rectangular beta density function is given by:

$$f(y \mid \mu, \nu, \theta) = \theta + (1-\theta) f_B(y \mid \mu, \nu) \ , \qquad (4.4)$$

where $0 < y < 1$.

The tilted beta distributions are appropriate to analyze datasets with larger variance than beta distributions with larger values of their density at the ends of the $(0, 1)$ interval. For example, when $\mu_t = 0.5$, the beta rectangular distribution (4.4) is obtained. For other values of μ_t, the tilted component of the mixture allocates more density on one side of the open $(0, 1)$ interval and less to the other side. The tilted beta distribution has been studied and applied in the project management context by García Perez *et al.* (2016) and Udoumoh *et al.* (2017), among others.

2. **Tilted mean and variance beta distribution**. The mean and variance tilted beta distribution is introduced as the convex combination of the tilted distribution (2.2) and $\mu_b\sigma^2$-beta distribution (3.4). The density function of a random variable Y that follows this distribution is given by:

$$f\left(y \mid \mu_t, \mu_b, \sigma_b^2, \theta\right) = \theta c\,(y \mid \mu_t) + (1-\theta) f_b\left(y \mid \mu_b, \sigma_b^2\right), \qquad (4.5)$$

where $0 < y < 1$ and $0 \le \theta \le 1$. The notation $Y \sim \text{TB}(\mu_t, \mu_b, \sigma^2, \theta)$ is used to denote that Y follows this distribution. The mean and the variance of Y are $E(Y) = \theta\mu_t + (1-\theta)\mu_b$ and

$$V(Y) \quad = \quad \theta\sigma_t^2 + (1-\theta)\sigma_b^2 + \theta(1-\theta)(\mu_t + \mu_b)^2, \qquad (4.6)$$

where $\sigma_t^2 = V_t(Y)$ is the variance of the tilted distribution and $\sigma_b^2 = V_b(Y)$ is the variance of the beta distribution.

3. **Tilted mean and precision beta distribution**. Given that precision is the inverse of variance, the tilted mean and precision beta distributions can be defined from (4.5) and written σ_b^2 as $1/\phi_b$. The mean and (variance) precision beta rectangular distribution can be defined as a particular case.

The tilted mean and variance (or precision) beta distributions are appropriate to analyze datasets with larger variance than the $\mu\nu$-beta distributions, but the these have the advantage of clearer and simpler parameter interpretations.

5. Tilted beta binomial distributions

At the beginning of this section, in Subsection 5.1, the mean and the "dispersion" ($\nu = a+b$) tilted beta binomial distribution is presented, following its definition proposed by Cepeda-Cuervo and Cifuentes-Amado (2017). After that, the mean and variance parametrization of this distribution is proposed in Subsection 5.2. Finally, in Subsection 5.3, following Cepeda-Cuervo (2023), we present the mean and variance beta binomial density function.

5.1. Tilted $\mu\nu$-beta binomial distributions

Let $Y|p \sim \text{bin}(m, p)$ be a random variable that follows the binomial distribution, where p follows the tilted beta distribution, $p \sim \text{TB}(\mu_t, \mu_b, \nu, \theta)$. Then Y follows a tilted beta binomial distribution with parameters μ_t, μ_b, ν and θ, which are denoted by $Y \sim \text{TBB}(\mu_t, \mu_b, \nu, \theta)$, if their probability function is given by:

$$\begin{aligned} f\,(y \mid \mu_t, \mu_b, \phi, \theta) &= \int_0^1 f_{Bin}\,(y \mid m, p)\left[\theta c\,(p \mid \mu_t) + (1-\theta) f_{Beta}\,(p \mid \mu_b, \nu)\right] dp \\ &= 2\theta\binom{m}{y}\left[\frac{y(6\mu_t - 3) + m(2 - 3\mu_t) + 1}{m+2}\right] B(y+1, m-y+1) + \\ &\quad (1-\theta) f_{BB(\mu_b,\nu)}(y)\,, \end{aligned} \tag{5.1}$$

where $y = 0, 1, \ldots, m$; $B(\cdot,\cdot)$ denotes the beta function, and $f_{BB(\mu_b,\nu)}(\cdot)$ denotes the density function of the beta binomial distribution, which is parameterized in terms of the mean and the dispersion parameters (Cepeda-Cuervo and Cifuentes-Amado, 2020).

The mean and variance of a random variable Y that follows the $(\mu_t, \mu_b, \nu, \theta)$-tilted beta binomial probability function are given by: $E(Y) = E(E(Y|p)) = mE(p) = m\left[\theta\mu_t + (1-\theta)\mu_b\right]$ and

$$\begin{aligned} V(Y) &= V(E(Y \mid p)) + E(V(Y \mid p)) \\ &= m^2 V(p) + mE(p) - mE(p^2)) \\ &= m\left\{(m-1)V(p) + E(p)(1 - E(p))\right\} \\ &= m\left\{(m-1)\left[\theta V_t + (1-\theta)V_b + \theta(1-\theta)(\mu_t + \mu_b)^2\right] + \left[\theta\mu_t + (1-\theta)\mu_b\right]\left[1 - \theta\mu_t + (1-\theta)\mu_b\right]\right\}, \end{aligned}$$

where μ_b and V_b denote the mean and variance of the beta distribution, respectively, and μ_t and V_t denote the mean and variance of the tilted beta distribution. The behavior of the $(\mu_t, \mu_b, \nu, \theta)$-tilted beta binomial probability function is illustrated in Cepeda-Cuervo and Cifuentes-Amado (2020), for different vectors of parameter values.

A particular case of this distribution is the Tilted (μ_b, ν, θ)-beta rectangular binomial distribution. Y follows this distribution if $Y|p$ follows a binomial distribution, $Y|p \sim \text{bin}(m, p)$, where p follows the beta rectangular distribution (4.4). This density function of Y can be obtained as a particular case of the tilted beta binomial distribution (5.1) by replacing μ_t with 0.5:

$$f\,(y \mid \mu_b, \phi, \theta) = \binom{m}{y}\theta B(y+1, m-y+1) + (1-\theta) f_{BB(\mu_b,\nu)}\,(y \mid \mu_b, \nu)\,, \tag{5.2}$$

where $y = 0, 1, \ldots, m$. From the equations of the mean (4.2) and variance (4.3) of the tilted beta binomial distribution with a setting $\mu_t = 0.5$, the mean and variance of the rectangular beta distribution are given by:

$$\begin{aligned} E(Y) &= m\left[\frac{\theta}{2} + (1-\theta)\mu\right] \\ V(Y) &= \left(m^2 - m\right)\left[\frac{\mu(1-\mu)}{1+\nu}(1-\theta)(1 + \theta(1+\phi)) + \frac{\theta}{12}(4 - 3\theta)\right] \\ &\quad + m\left[\frac{\theta}{2} + (1-\theta)\mu\right]\left[\frac{2-\theta}{2} - (1-\theta)\mu\right] \end{aligned}$$

5.2. Tilted $\mu\sigma^2$-beta binomial distributions

Let $Y|p \sim \text{bin}(m, p)$ be a random variable that follows a binomial distribution, where p follows the tilted beta distribution, $p \sim \text{TB}(\mu_t, \mu_b, \sigma^2, \theta)$. Then Y follows the tilted beta binomial distribution with parameters μ_t, μ_b, σ^2 and θ, which are denoted by $Y \sim \text{TBB}(\mu_t, \mu_b, \sigma^2, \theta)$. The probability of this distribution is given by:

$$f\,(y \mid \mu_t, \mu_b, \phi, \theta) = 2\theta\binom{m}{y}\left[\frac{y(6\mu_t - 3) + m(2 - 3\mu_t) + 1}{m + 2}\right]B(y + 1, m - y + 1) + (1 - \theta)f_{BB(\mu_b,\sigma^2)}(y)\,, \tag{5.3}$$

which for $\mu_t = 0.5$ is the beta rectangular binomial distribution.

A particular case of the distribution (5.3) is the $(\mu_b, \sigma^2, \theta)$- tilted beta rectangular binomial distribution. In this case, Y follows the (μ_b,σ^2,θ)-beta rectangular binomial distribution if $Y|p \sim \text{bin}(m, p)$ is a random variable that follows a binomial distribution, and where p follows the beta rectangular distribution. The density function of this distribution can be obtained as a particular case of the tilted beta binomial distribution (5.3), by replacing μ_t with $1/2$:

$$f\left(y \mid \mu_b, \sigma^2, \theta\right) = \binom{m}{y}\theta B(y + 1, m - y + 1) + (1 - \theta)f_{BB(\mu_b,\sigma^2)}(y)\,, \tag{5.4}$$

where $y = 0, 1, \ldots, m$.

5.3. $\mu\sigma^2$-beta binomial distribution

The $\mu\sigma^2$-beta binomial distribution, defined in Cepeda-Cuervo (2023), is obtained by assuming that a random variable Y follows a binomial distribution $B(m, p)$, where p follows the $\mu\sigma^2$-beta distribution (3.4). The $\mu\sigma^2$-beta binomial probability function is given by:

$$f\left(y \mid \mu_b, \sigma^2, \theta\right) = \binom{n}{r}\frac{B(y + \mu(\mu(1 - \mu)\phi - 1), m - y + (\mu(1 - \mu)\phi - 1)(1 - \mu)}{B(\mu(\mu(1 - \mu)\phi - 1), (\mu(1 - \mu)\phi - 1)(1 - \mu)}\,, \tag{5.5}$$

where $0 < \mu < 1$, $\phi = 1/\sigma^2$ and $0 < \sigma^2 < 1/4$. ϕ is the precision parameter of the beta distribution.

6. Mean and variance beta regression models

The beta regression model was proposed in Cepeda-Cuervo (2001), under a Bayesian framework by assuming that the mean (μ) and the dispersion ($\nu = a+b$) parameters follow linear regression structures given by:

$$\text{h}(\mu_i) = \mathbf{x}_i^t\beta, \tag{6.1}$$

$$\text{g}(\nu_i) = \mathbf{z}_i^t\gamma, \tag{6.2}$$

where h is the logit function; g is the logarithmic function; and $\beta = (\beta_0,\beta_1,\ldots,\beta_k)^t$ and $\gamma = (\gamma_0,\gamma_1,\ldots,\gamma_p)^t$ are the vectors of the mean and dispersion regression parameters, respectively; $\mathbf{x}_i = (x_{i1},\ldots,x_{ik})^t$ is the vector of the mean explanatory variables; and $\mathbf{z}_i = (z_{i1},\ldots,z_{ip})^t$ is the vector of the dispersion explanatory variables at the i^{th} observation. A frequentist approach to the beta regression models was presented by Ferrari and Cribari-Neto (2004), assuming that h is an appropriate real valued function, strictly monotonic and twice differentiable, defined on the interval $(0, 1)$, and ν is a constant dispersion parameter. These authors presented a wide range of applications where the practitioner needs to assume regression structures to explain the behavior of the variables of interest. Although many variations of mean and dispersion beta regression models have been developed in recent years, these proposals have at least two drawbacks. The first is the interpretability of the dispersion parameter ν, given that ν is considered to be a precision parameter for a constant mean, the variance decreases when ν increases. A second problem is the lack of an explicit regression structure for the variance, which impairs the quality of the posterior regression parameter inferences.

A first approach to the mean and variance beta regression models was proposed in Cepeda-Cuervo (2015) and a general definition was formulated in Cepeda-Cuervo (2023). In Cepeda-Cuervo (2023), the mean regression structure is given by (6.3) and the variance (or precision) regression structure is given by (6.4), where $h(\cdot)$ and $g(\cdot)$ are real functions defined in the open interval $(0, 1)$, like the logit, probit, log-log and complementary log-log functions.

$$\mathrm{h}(\mu_i) = \mathbf{x}_i^t\beta, \tag{6.3}$$

$$\mathrm{g}\left(4{\sigma_i}^2\right) = \mathbf{z}^t\gamma. \tag{6.4}$$

If, as in Cepeda-Cuervo (2023), for example, the mean and variance of the beta regression model are given by $\mathrm{logit}(\mu) = \mathbf{x}^t\beta$ and $\mathrm{logit}(4\sigma^2) = \mathbf{z}^t\gamma$, then the parameter estimates of the mean and variance regression structures are easily interpretable.

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

2. If Z_1 is an explanatory variable associated with parameter γ_1 where $\gamma_1 > 0$, increasing behavior of Z_1 is associated with increasing variance, and where $\gamma_1 < 0$, increasing behavior of Z_1 is associated with decreasing variance.

The mean and precision ($\phi = 1/\sigma^2$) beta regression model can be defined by the mean regression structure (6.3) and by $g(\phi - 4) = \mathbf{z}^t\gamma$, where $g(\cdot)$ is the logarithmic function or some other appropriate real function defined from the positive real number set to the real numbers, such as the logarithmic function.

The results of the statistical analysis of the dyslexic dataset presented in Cepeda-Cuervo (2023), and obtained by applying $\mu\sigma^2$-beta regression models, reveal the good performance of this model and the easy interpretation of the posterior parameter inferences compared with that obtained from fitting the $\mu\nu$-beta regression model to this dataset. In the $\mu\sigma^2$-beta regression model, the variance of the variable of interest is interpreted according to items 1 and 2 of this section, which is unconditional to the mean values. Thus, the mean and variance beta regression models, defined in (6.3) and (6.4), have a substantial interpretative advantage compared with the mean and "dispersion" models, defined by (6.1) and (6.2).

Additionally, in the results of simulation processes, the mean and variance models outperform the mean and dispersion models, which can be established by statistical methods. In these simulations, the explanatory variables can be generated from uniform distributions, the mean and dispersion parameters are obtained from their respective mean and dispersion structures, and the observations of the variable of interest are generated from the beta distributions. Finally, the beta regression models were fitted to the resulting dataset, and the model with the best fit was the mean and variance beta regression model, which had the smallest residuals.

With the new parameterization of the beta distributions proposed by Cepeda-Cuervo (2023), the tilted mean and variance beta regression model is defined from the mixture distribution (4.5), a convex combination of the tilted distribution (2.2) and the $\mu_b\sigma^2$-beta density function (3.4), where an appropriate function of their parameters follows linear the regression structures:

Let $Y_i \sim \text{TB}(\mu_{ti}, \mu_{bi}, \sigma^2_{bi}, \theta_i)$, $i = 1, 2, \ldots, n$, be independent random variables with tilted mean and variance beta distribution. Let $\mathbf{x}_i = (x_{i1}, \ldots, x_{ip})^t$, $\mathbf{z}_i = (z_{i1}, \ldots, z_{ik})^t$, $\mathbf{w}_i = (w_{i1}, \ldots, w_{il})^t$ and $\tilde{\mathbf{x}}_i = (\tilde{x}_{i1}, \ldots, \tilde{x}_{is})^t$ be the covariate vectors of μ_{bi}, σ^2_{bi}, θ_i and μ_{ti} regression structures, and $\beta = (\beta_1, \ldots, \beta_p)^t$, $\gamma = (\gamma_1, \ldots, \gamma_k)^t$, $\boldsymbol{\delta} = (\delta_1, \ldots, \delta_l)^t$ and $\alpha = (\alpha_1, \ldots, \alpha_s)^t$, be the respective regression parameter vectors. Thus, the tilted mean and variance regression models are defined from the mean and variance tilted beta distribution (5.3) by assuming the following regression structures.

$$\text{logit}(\mu_{bi}) = \mathbf{x}_i^t\beta, \tag{6.5}$$

$$\log\left(4\sigma^2_{bi}\right) = \mathbf{z}_i^t\gamma, \tag{6.6}$$

$$\text{logit}(\theta_i) = \mathbf{w}_i^t\boldsymbol{\delta}, \tag{6.7}$$

$$\text{logit}\,(3\mu_{t_i} - 1) = \tilde{\mathbf{x}}_i^t\boldsymbol{\alpha}. \tag{6.8}$$

This parameterization of the tilted beta distribution has some interpretive advantages, that is related to other parameterization of this distribution. In this parameterization, if Y follows a tilted beta distribution, then

$$E(Y_i) = \theta\left(\frac{\exp(\mathbf{x}_i^t\beta)}{1 + \exp(\mathbf{x}_i^t\beta)}\right) + (1 - \theta)\left(\frac{\exp(\tilde{\mathbf{x}}_i^t\alpha)}{3 + 3\exp(\tilde{\mathbf{x}}_i^t\alpha)} + \frac{1}{3}\right).$$

Thus, the contribution of an explanatory variable to the mean behavior of the variable of interest Y can be easily established. A similar argument can be established to explain the contribution of the explanatory variables to the behavior of the variance.

Many extensions of the mean and variance beta regression models can be proposed, which provide valuable insights Simas *et al.* (2010), by assuming nonlinear regression structures for the mean and variance of the beta regression models.

7. Tilted beta binomial regression models

In Item 1 of this section, the tilted beta binomial regression models defined in Cepeda-Cuervo and Cifuentes-Amado (2020), where μ_b, ν and θ_i follow regression structures, are extended to include a mean regression structure of the tilted mixture parameter components. Additionally, in Item 2, considering the reduced interpretation of the "dispersion" parameter in the beta density function (3.3) and in the tilted binomial distribution (5.1), we propose the $\mu_t\mu_b\sigma^2\theta$-tilted beta binomial regression models, where μ_t and σ^2 also follow regression structures. Finally, in item 3, as a particular case of item 2, the $\mu_b\sigma^2$-beta binomial regression models, proposed in Cepeda-Cuervo (2023), are presented.

1. **Tilted $\mu_t\mu_b\nu\theta$-beta binomial regression models:** Let $Y_i \sim \text{TBB}(\mu_{ti}, \mu_{bi}, \nu_i, \theta_i)$, $i = 1, 2, \ldots, n$, be independent random variables with tilted beta binomial distribution. Let $\mathbf{x}_i = (x_{i1}, \ldots, x_{ip})^t$, $\mathbf{z}_i = (z_{i1}, \ldots, z_{ik})^t$, $\mathbf{w}_i = (w_{i1}, \ldots, w_{il})^t$ and $\tilde{\mathbf{x}}_i = (\tilde{x}_{i1}, \ldots, \tilde{x}_{is})^t$ be the covariate vectors of the μ_{bi}, ν_i, θ_i and μ_{ti} regression structures, and $\beta = (\beta_1, \ldots, \beta_p)^t$, $\gamma = (\gamma_1, \ldots, \gamma_k)^t$, $\boldsymbol{\delta} = (\delta_1, \ldots, \delta_l)^t$ and

$\alpha = (\alpha_1, \ldots, \alpha_s)^t$, be the respective regression parameter vectors such that:

$$\text{logit}(\mu_{bi}) = \mathbf{x}_i^t\beta, \tag{7.1}$$
$$\log(\nu_i) = \mathbf{z}_i^t\gamma, \tag{7.2}$$
$$\text{logit}(\theta_i) = \mathbf{w}_i^t\boldsymbol{\delta}, \tag{7.3}$$
$$\text{logit}(3\mu_{t_i} - 1) = \tilde{\mathbf{x}}_i^t\boldsymbol{\alpha}. \tag{7.4}$$

Thus, the likelihood function of the TBB$(\mu_{ti}, \mu_{bi}, \nu_i, \theta_i)$-regression model is given by: $L(\mu_{ti}, \mu_{bi}, \nu_i, \theta_i)$ $= \prod_i^n f(y|\mu_{ti}, \mu_{bi}, \phi_1, \theta_i)$, where $f(\cdot|\mu_{ti}, \mu_{bi}, \phi_1, \theta_i)$ is given by (7.5).

$$f\,(y_i \mid \mu_{ti}, \mu_{bi}, \phi_1, \theta_i) = 2\theta_i \binom{m_i}{y_i} \left[\frac{y_i(6\mu_{ti} - 3) + m_i(2 - 3\mu_{ti}) + 1}{m_i + 2}\right] \times$$
$$B\,(y_i + 1, m_i - y_i + 1) + (1 - \theta_i)\, f_{BB(\mu_{bi}, \nu_i)}(y_i), \tag{7.5}$$

2. $\mu_t\mu_b\sigma^2\theta$**-tilted beta binomial regression models**. These models are defined from the $\mu_t\mu_b\sigma^2\theta$-tilted beta binomial distribution (5.3) by assuming the following regression structures: (7.1) for μ_b, (7.3) for θ_i, (7.4) for μ_t, and $\text{logit}(4\sigma_i^2) = \mathbf{z}_i^t\boldsymbol{\gamma}$ for σ_i^2.

3. $\mu_b\sigma^2$**-beta binomial regression models**. These models are defining from the $\mu_b\sigma^2$-beta binomial distribution given in 5.3 by assuming the following regression structures: (7.1) for μ_b and $\text{logit}(4\sigma_i^2) = \mathbf{z}_i^t\boldsymbol{\gamma}$ for σ_i^2, as proposed in Cepeda-Cuervo (2023).

Hahn (2022), in his applications and simulations, established that the performance of the tilted beta-binomial distribution is better than the beta-binomial model, including a first application to big data. The author found evidence for the existence of the beta-binomial and tilted binomial components in applications of demographic datasets.

8. Applications

This section includes posterior parameter inferences that are obtained by applying the $\mu_t\mu_b\sigma^2\theta$-tilted beta binomial regression models to analyze the school absenteeism dataset in Section 8.1 and seed germination dataset in Section 8.2. In both cases, in order to define the Bayesian tilted beta binomial regression model, the following a priori distributions are assumed for the regression parameters: $\beta \sim N(0, \boldsymbol{B})$, $\boldsymbol{\gamma} \sim N(0, \boldsymbol{G})$, $\boldsymbol{\delta} \sim N(0, \boldsymbol{D})$ and $\boldsymbol{\alpha} \sim N(0, \boldsymbol{D})$. If there are no explanatory variables for μ_t, then (7.4), as given by $\text{logit}(3\mu_{t_i} - 1) = \alpha_0$ and $\alpha_0 \sim N(0, 10^k)$, where k is a positive real number, can be assumed to be the prior distribution. Also, given that $\mu_t \sim U(1/3, 2/3)$, the uniform distribution $\mu_t \sim U(1/3, 2/3)$ can be assumed as the prior distribution of μ_t.

8.1. School absenteeism dataset

The first 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 between the two sexes and equal numbers from between 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. In this dataset, the response variable of interest is the number of days that a child was absent during the school year (days absent: Y). The explanatory variables are the following factors with two levels:

Table 1: Parameter estimates of the $\mu_t\mu_b\sigma^2\theta$-tilted beta binomial model (Model 1) and the $\mu\sigma^2$-beta binomial model (Model 2) in the analysis of the school absenteeism dataset

	Model 1		Model 2	
Param.	Mean (S.D.)	95% Cred. Int.	Mean (S.D.)	95% Cred. Int
β_0	−1.985(0.149)	(−2.25,−1.683)	−1.976(0.142)	(−2.250,−1.693)
β_3	−0.531(0.111)	(−0.736,−0.335)	−0.533(0.107)	(−0.743,−0.323)
β_4	−0.490(0.198)	(−0.821,−0.117)	−0.501(0.192)	(−0.869,−0.121)
γ_0	−3.363(0.261)	(−3.836,−2.76)	−3.344(0.264)	(−3.833,−2.811)
γ_2	−0.984(0.402)	(−1.654,-0.325)	−1.007(0.396)	(−1.771,−0.207)
g_0	−11.450(5.430)	(−23.840,-4.152)	- - -	- - -
d_0	−0.951(10.480)	(−21.460, 19.720)	- - -	- - -
DIC(Dhat)	−372.7 (−383.2)		−372.3 (−383.1)	
SSE	0.6264		0.6259	

- *Cultural or ethnic background (CB)*: Aboriginal (0) and White (1).
- *Learning ability (LA)*: Slow learner (0) Average learner (1).

Since the variable Days Absent, Y, counts the number of events that occurred during a year, this dataset was analyzed by Cepeda-Cuervo and Cifuentes-Amado (2017) assuming a negative binomial model NB(μ, α), where the mean and the shape parameters follow linear regression structures . In this paper, assuming that a school year has 200 days, we analyze this dataset by applying the $\mu_t\mu_b\sigma^2\theta$-tilted beta binomial regression model, and assume the following linear regression structures:

$$\text{logit}(\mu_{bi}) = \beta_0 + \beta_3 CB_i + \beta_4 LA_i \tag{8.1}$$

$$\text{logit}(4\sigma_i^2) = \gamma_0 + \gamma_2 LA_i \tag{8.2}$$

$$\text{logit}(\theta) = g_0 \tag{8.3}$$

$$\text{logit}(3(\mu_t - 1/3)) = d_0. \tag{8.4}$$

This tilted beta binomial regression model was fitted to the dataset by applying Bayesian methods and using the OpenBugs software. Thus, assuming the mean and variance regression structures given by (8.1) to (8.4), respectively; the posterior parameter estimates, standard deviations and 95% credible intervals are given in Table 1 (Model 1). Thus, from the posterior samples of g_0 and d_0, and assuming the mean and variance regression structures given by (8.3) to (8.4), a posterior sample of θ and μ_t were obtained, respectively, with the posterior parameter estimates and the respective standard deviations (between parentheses) being $\hat{\theta} = 0, 0013368(0.0005343)$, $\hat{\mu}_t = 0.4789(0.1520)$. From these estimates, it is clear that the parameter estimate of θ is close to zero. For this reason, a $\mu\sigma^2$-beta binomial regression model, defined by (8.1) and (8.2) was fitted. Their parameter estimates are reported in the same table (Model 2).

The parameter estimates of the mean and variance regression structures of Model 2 agree with those of Model 1. The deviance information criterion (DIC) values and the sum of square errors are similar, but the DIC value of Model 1 is a little smaller than the DIC value of Model 2. In both models, the estimates of γ_2 are negative, which shows that decreasing values of LA are associated with increasing variance behavior.

The posterior credibility interval for a regression parameter is given by the real numbers L_I and L_S, $L_I < L_S$ such that the posterior probability, for which the parameter estimates lie between L_I and L_S, is 95%. These real numbers were obtained from the posterior sample assuming extreme tail samples of 2.5%.

Table 2: Parameter estimates of the $\mu_t\mu_b\sigma^2\theta$-tilted beta binomial regression models in the analysis of the seed germination dataset

	Model 3		Model 4	
1-5 Param.	Mean (S.D.)	95% Cred. Int.	Mean (S.D.)	95% Cred. Int
β_0	−0.822(0.266)	(−1.404,−0.336)	−0.774(0.264)	(−8.003,−0.960)
β_1	0.466(0.238)	(−0.001,0.935)	0.374(0.264)	(−0.102,0.922)
β_2	1.040(0.247)	(0.559,1.540)	1.021(0.248)	(0.528,1.515)
γ_0	−3.436(1.014)	(−5.891,−1.715)	−3.864(0.969)	(−6.219,−2.422)
γ_1	−1.832(1.847)	(−5.930,1.358)	—	—
c_0	−3.552(1.979)	(−7.780,−0.058)	−3.903(1.841)	(−8.003,−0.960)
c_1	−1.600(2.596)	(−7.243,−1.358)	—	—
μ_t	0.498(0.092)	(0.005,0.651)	0.4922 (0.092)	(0.347,0.652)
DIC	122.4		122.3	
SSE	420.205		405.124	

8.2. Seed germination dataset

The dataset analyzed in this section is available in Spiegelhalter *et al.* (2003) and corresponds to the number of seeds that germinated from an initial quantity arranged in each of 21 dishes organized according to a 2 by 2 factorial design (2 seed types and 2 root types). This data was initially reported by Crowder (1978). The variables involved in the experiment are described as follows:

- **Y**: number of seeds germinated in each dish.
- **n**: number of seeds initially arranged in each dish.
- $\mathbf{X_1}$: seed type (0) if it is *O. aegyptiaca 73* and (1) if it is *O. aegyptica 75*.
- $\mathbf{X_2}$: root type (0) if it is a bean and (1) if it is a cucumber.

In this experiment, there are 21 observations (21 dishes). Since the variable Y represents the number of germinated seeds in each dish, this variable can be assumed to follow the TBB(μ_t,μ_b,σ^2,θ) distribution, and thus, the seed germination dataset can be analyzed by applying the TBB linear regression model defined by the regression structures given in equations (7.1) to (7.4), which include all the explanatory variables in each of the regression structures. After the process of eliminating the explanatory variables, the best model (smallest DIC value) has the following regression structures:

$$\text{logit}(\mu_{ib}) = \beta_0 + \beta_1 x_{1i} + \beta_2 x_{2i} \tag{8.5}$$

$$\text{logit}\left(4\sigma_i^2\right) = \gamma_0 + \gamma_1 x_{2i} \tag{8.6}$$

$$\text{logit}(\theta_i) = c_0 + c_1 x_{2i} \tag{8.7}$$

with constant tilted mean μ_t. Thus, assuming normal prior distribution $N(0, 10^k)$ with k = 5, for the regression parameters (β_i, γ_i and c_i, $i = 1, 2, 3$) and uniform distribution $\mu_t \sim U(1/3, 2/3)$ for the mean of tilted distribution, the TBB(μ_t,μ_b, σ^2, θ) model was fitted to this dataset using OpenBUGS, which isa free program used to fit Bayesian models that apply Gibbs algorithms (Spiegelhalter *et al.*, 2003). The posterior parameter inferences obtained from a sample of size 100000 with a burn-in of 10000 and taking one sample every 10 to reduce autocorrelation, are summarized in Table 2 (Model 3).

Given that 0 belongs to the 95% credible intervals of γ_1 and c_1, a tilted $\mu\sigma^2$-beta binomial model with regression structures from (8.5) to (8.7) and without x_2 in the variance and mixture regression

Table 3: Parameter estimates of the tilted $\mu\nu$-beta binomial regression models in the analysis of the seed germination dataset

	Model 5		Model 6	
Param.	Mean (S.D.)	95% Cred. Int.	Mean (S.D.)	95% Cred. Int
β_0	−0.814(0.392)	(−1.429, −0.268)	−0.759(0.252)	(−1.292,−0.307)
β_1	0.444 (0.253)	(−0.053, 0.934)	0.384 (0.258)	(−0.097,0.923)
β_2	1.057 (0.3632)	(0.5149, 1.61)	1.016 (0.2484)	(0.5147, 1.49)
γ_0	3.388 (1.112)	(1.561, 6.10)	3.690 (0.894)	(2.331, 5.865)
γ_1	1.742(2.041)	(−1.745,6.507,)	—	—
c_0	−3.399 (2.180)	(−8.018, 0.547)	−3.803 (1.818)	(−7.811, −0.9436)
c_1	−1.663 (2.533)	(−6.800, 3.169)	—	—
μ_t	0.485 (0.092)	(0.345, 0.482)	0.491 (0.094)	(0.346, 0.651)
DIC	123.2		123.1	
SSE	411.233		408.215	

structures, was fitted to this dataset and their posterior parameter inferences reported in Table 2 (Model 4). From this table, it is possible to conclude that Model 4 is the best (smallest DIC value, smallest SSE and all the null hypotheses of the regression parameters rejected).

To compare the performance of the $\mu_t\mu_b\sigma^2\theta$- and the $\mu_t\mu_b\nu\theta$-tilted beta binomial regression models in the analysis of the seed germination dataset, Table 3 presents the posterior parameter estimates that are obtained when the $\mu_t\mu_b\nu\theta$-tilted beta binomial regression models were fitted to this dataset, and assuming the regression structures given by:

$$\text{logit}(\mu_{ib}) = \beta_0 + \beta_1 x_{1i} + \beta_2 x_{2i} \tag{8.8}$$

$$\log(\nu_i) = \gamma_0 + \gamma_1 x_{2i} \tag{8.9}$$

$$\text{logit}(\theta_i) = c_0 + c_1 x_{2i} \,, \tag{8.10}$$

with constant mean μ_t of the tilted distribution, which assume the same prior distributions, like in the first application. The posterior parameter estimates of the tilted beta binomial model with regression structures given by (8.8), (8.9) and (8.10) are reported in Table 3, Model 5. The parameter estimates of the reduced tilted $\mu\nu$-beta binomial regression models are given in the same table (Model 6).

From the results reported in these tables, it is possible to conclude that the estimates of the means and mixture regression parameter structures agree with the $\mu_t\mu_b\nu\theta$ and $\mu_t\mu_b\sigma^2\theta$ tilted beta binomial regression models. The parameter estimates of the ν and σ^2 regression structures are congruent with their parameter definitions. The DIC and SSE values are the smallest for Model 4 among the $\mu_t\mu_b\sigma^2\theta$-tilted beta binomial regression models. Thus, in this application, the Model 4 is assumed to be the best.

9. Conclusions

This paper proposes the mean and variance parameterizations of the tilted beta binomial distribution, which include two particular cases: Mean and variance beta rectangular distributions and mean and variance beta binomial distribution, that improves the parameter interpretation of these distributions defined from the mean and "dispersion" ($\nu = p + q$) parameterizations of the beta distribution. From the new parameterized distributions, new linear regression models that deal with overdispersed binomial datasets are proposed, where the mean and variance of the beta distribution, the mean of the tilted distribution and the mixture parameter follow regression structures. These new linear regression models were fitted to the school absenteeism dataset and to the seed germination rate, which depended depending on the type of seed chosen, by applying Bayesian methods and using the OpenBUGS software. The models show good performance and a clear interpretation of their regression parameters.

Many extensions of these models can be proposed. One possibility is to use maximum likelihood methods to fit the proposed models. Additionally, following Simas *et al.* (2010), the tilted beta and beta-binomial nonlinear regression models can be formulated by assuming nonlinear regression structures for the mean and variance of the beta distribution component in the mixture.

According to Hahn (2022), the tilted beta-binomial distribution is clearly more appropriate than the beta-binomial distribution to analyze datasets with overdispersion. Thus, taking into account the better interpretability of the mean and variance, the tilted beta regression models can be proposed as good options for analyzing analyze these types of overdispersed count exit/failure datasets.

In many areas of knowledge there is a wide range of applications with random variables of interest of the exit/failure type, where the tilted mean and variance beta regression models can be applied. These types of data analysis are also possible extensions that can be used by researchers and students of statistics, and in the addition to studies and development of statistical packages for fitting the proposed models.

References

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

Cepeda-Cuervo E (2015). Beta regression models: Joint mean and variance modeling, *Journal of Statistical Theory and Practice*, **9**, 134–145.

Cepeda-Cuervo E (2023). $\mu\sigma^2$-beta and $\mu\sigma^2$-beta binomial regression models, *Revista Colombiana de Estadística*, **46**, 63–79.

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

Cepeda-Cuervo E and Cifuentes-Amado MV (2020). Tilted beta binomial linear regression model: A bayesian approach, *Journal of Mathematics and Statistics*, **16**, 1–8.

Collet D (1991). *Modeling Binary Data*, Chapman Hall, London.

Cox D (1983). Some remarks on overdispersion, *Biometrika*, **70**, 269–274.

Crowder MJ (1978). Beta-binomial ANOVA for proportions, *Applied Statistics*, **27**, 34–37.

Efron B (1986). Double exponential families and their use in generalized linear regression, *Journal of the American Statistical Association*, **81**, 709–721.

Ferrari S and Cribari-Neto F (2004). Beta Regression for Modelling Rates and Proportions, *Journal of Applied Statistics*, **31**, 799–815.

García Perez J, López MM, García García C, and MA SG (2016). Project management under uncertainty beyond beta: The generalized bicubic distribution, *Operations Research Perspectives*, 67–76.

Pérez JG, Martín MD, M. L., García, C. G., and Granero MÁS (2016). Project management under uncertainty beyond beta: The generalized bicubic distribution, *Operations Research Perspectives*, **3**, 67–76.

Hahn E (2008). Mixture densities for project management activity times: A robust approach to pert, *European Journal of Operational Research*, **188**, 450–459.

Hahn E (2022). The tilted beta-binomial distribution in overdispersed data: Maximum likelihood and bayesian estimation, *Journal of Statistical Theory and Practice*, **16**, 43.

Hahn E and López Martín M (2015). Robust project management with the tilted beta distribution, *SORT*, **39**, 253–272.

Hinde J and Demetrio C (1998). Overdispersion: Models and estimation, *Computational Statistics & Data Analysis*, **27**, 151–170.

Jorgensen B (1997). Proper dispersion models, *Brazilian Journal of Probability and Statistics*, **11**, 89–128.

Quine S (1975). *Achievement orientation of aboriginal and white australian adolescents* (Ph.d. thesis), Australian National University, Australia.

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

Spiegelhalter DJ, Thomas A, Best N, and Lunn D (2003). WinBUGS Version 1.4., *MRC Biostatistics Unit, Cambridge, UK.*

Udoumoh EF, Ebong DW, and Iwok IA (2017). Simulation of project completion time with Burr XII activity distribution, *Asian Research Journal of Mathematics*, **6**, 1–14.

Williams D (1982). Extra-binomial variation in logistic linear models, *Journal of the Royal Statistical Society: Series C (Applied Statistics)*, **31**, 144–148.

Received February 19, 2023; Revised July 04, 2023; Accepted December 16, 2023