Original Research Paper

Tilted Beta Binomial Linear Regression Model: A Bayesian Approach

[1]**Edilberto Cepeda-Cuervo** and [2]**Maria Victoria Cifuentes-Amado**

[1]*Departamento Estadistica, Universidad Nacional de Colombia, Colombia*
[2]*Departamento Matematicas, Universidad Nacional Colombia, Colombia*

Abstract: This paper proposes new linear regression models to deal with overdispersed binomial datasets. These new models, called tilted beta binomial regression models, are defined from the tilted beta binomial distribution, proposed assuming that the parameter of the binomial distribution follows a tilted beta distribution. As a particular case of this regression models, we propose the beta rectangular binomial regression models, defined from the binomial distribution assuming that their parameters follow a beta rectangular distribution. These new linear regression models, defined assuming that the parameters of these new distributions follow regression structures, are fitted applying Bayesian methods and using the OpenBUGS software. The proposed regression models are fitted to an overdispersed binomial dataset of the number of seeds that germinate depending on the type of chosen seed and root.

Keywords: Count Data, Overdispersion, Tilted Beta Distribution, Binomial Distribution, Tilted Beta Binomial Distribution, Bayesian Approach

Introduction

The binomial distribution is normally used to model the number of successes obtained in a finite number of experiments. However, in these cases, 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 (overdispersion), can lead to underestimation errors, lost efficiency of estimates and underestimation of the variance, wich that can in turn generate incorrect inferences about the regression parameters or the credible intervals (Collet, 1991; Cox, 1983; Williams, 1982).

There are several approaches to study overdispersed binomial datasets. Hinde and Demetrio (1998) categorized the majority of overdispersed binomial models in 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 of the distribution of the response variable is itself a random variable. In the first class, the double exponential family of distributions allows the researcher to obtain double

binomial models, which allow including a second parameter, which independently from the mean 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, results by assuming that the response variable follows a binomial distribution and the probability parameter of the binomial distribution follows a beta distribution. From the parameterization of the beta distribution, in terms of its mean and dispersion parameter (Jorgensen, 1997), a parameterization of the beta binomial beta distribution in terms of its mean and dispersion parameters is presented in Cepeda-Cuervo and Cifuentes-Amado (2017).

Despite the versatility of the beta distribution, Hahn (2008) proposed the rectangular beta distribution as a combination between the beta distribution and the uniform distribution, to admit heavier tails than that admitted by the beta distribution. After that, Hahn and Lopez Martın (2015) introduced tilted beta distribution, which has as particular cases the beta rectangular and the beta distributions.

In this paper, to obtain a good flexibility of the beta binomial distribution, we assume that p in the binomial $Bin(m, p)$ distribution follows a tilted beta distribution, assigning higher probability to tail values of p. Thus, we generalize the beta binomial regression models for fitting overdispersed binomial count dataset (Cepeda-Cuervo and Cifuentes-Amado, 2017) by introducing the tilted beta binomial linear regression model. For this, the tilted beta binomial probability is defined by assuming that the parameter of the binomial distribution follows the mean tilted beta distribution. In addition, the beta rectangular binomial models are presented as particular cases of the new proposed model, by assuming that the parameter of the binomial distribution has beta rectangular distribution. The proposed models are fitted using Bayesian methods. Finally, in order to illustrate of the tilted beta binomial model, we fit it to a seed germination count dataset and compare it with the rectangular beta binomial model and the binomial model, using their DIC values.

This paper is organized as follows. After the introduction, in Section 2, the tilted and the reparameterized tilted beta distributions are presented. In Section 3, the tilted beta binomial distribution is introduced and the rectangular beta binomial distribution is presented as a particular case. In Section 4, the tilted beta binomial linear regression model is defined. Finally, in Section 5, we analyze how the proportion of seeds that germinated on each of 21 dishes, is influenced by the type of seed and root, by fitting a tilted beta binomial linear regression, using the OpenBUGS software. The proposed model performance is compared with the binomial and beta binomial regression models.

The Tilted Beta Distribution

In different fields there is often a need to model continuous random variables that assume values in a bounded interval on a set of explanatory variables. Cepeda-Cuervo (2001) proposed the beta regression models, where mean and dispersion parameters follow regression structures (see also Cepeda and Gamerman (2005), Cepeda-Cuervo and Garrido (2015)). If the continuous variable Y assumes values in a bounded open interval (a, b), a beta regression models can be proposed, using the basic transformation $(y-a)/(b-a)$. However, in order to admit heavier tails than is possible in the beta distribution, Hahn (2008) proposed the rectangular beta distribution as a new distribution that, like the beta distribution, has as domain the open interval $(0, 1)$. The rectangular beta distribution consists of convex combination between the beta distribution and the uniform distribution $U(0, 1)$. Subsequently Hahn and Lopez Martın (2015), proposed the tilted beta distribution, consisting of a mixture of the beta distribution and the tilted distribution, which has as particular cases the beta rectangular distribution and the beta distribution. This section presents a reparameterization of the tilted beta distribution proposed by Hahn and Lopez Martın (2015), in terms of the mean and the dispersion parameters of the beta distribution μ_b and ϕ, respectively and the mean of the tilted beta distribution μt. The $(\mu_t, \mu_b, \phi, \theta)$-tilted beta binomial distribution results from the convex combination between the tilted reparameterized beta distribution and the binomial distribution.

The Tilted Distribution

A random variable Y follows an inclined distribution with a parameter v (Hahn and Lopez Martın, 2015) if its density is given by:

$$c(y \mid v) = \left[2v - 2(2v-1)y\right] I_{(0,1)}(y), \ 0 \le v \le 1 \qquad (1)$$

The mean of Y, denoted $\mu_t := E(Y|v)$, is equal to $\mu_t = (2-v)/3$. By reparameterizing (1) in terms of the mean, the density function is defined by:

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

where, $1/3 \le \mu_t \le 2/3$, given that the moments, $E_t(Y^n)$, of a random variable Y which follows the density function (2) are given by:

$$E_t\left(Y^n\right) = \int_0^1 y^n \left[3(2\mu_t - 1)(2y-1)+1\right] dy$$
$$= \frac{3n(2\mu_t - 1) + n + 2}{(n+1)(n+2)}, n = 1, 2, \ldots \qquad (3)$$

Their variance, $V_t(Y)$, is given by:

$$V_t\left(Y\right) = \frac{6(2\mu_t - 1) + 4}{12} - \mu_t^2$$
$$= \mu_t\left(1 - \mu_t\right) - \frac{1}{6}.$$

Reparameterized Tilted Beta Distribution

The tilted beta distribution was introduced by Hahn and Lopez Martın (2015), as the convex

combination between the tilted distribution and the beta distribution. If this distribution is obtained from the combination of the mean tilted distribution (2) and the mean and the dispersion beta distribution, $Beta(\mu_b, \phi)$, the density function of the tilted beta distribution is given by (4):

$$f\left(y\mid\mu_t,\mu_b,\phi,\theta\right)=\theta c\left(y\mid\mu_t\right)+(1-\theta)f_{Beta}\left(y\mid\mu_b,\phi\right) \qquad (4)$$

where, $0 < y < 1$ and $0 \le \theta \le 1$. The notation $Y \sim BI(\mu_t, \mu_b, \phi, \theta)$ is used to denote that Y follows a tilted beta distribution. Since the n-th-moment of Y is given by:

$$
\begin{aligned}
E\left(Y^n\right) &= \int_0^1 y^n \theta c\left(y\mid\mu_t\right)dy + \int_0^1 y^n(1-\theta)f_{Beta}\left(y\mid\mu_b\right)dy \\
&= \theta E_t\left(Y^n\right)+(1-\theta)E_b\left(Y^n\right) \qquad (5) \\
&= \theta\frac{3n(2\mu_t-1)+n+2}{(n+1)(n+2)}+(1-\theta)\frac{\Gamma(\mu_b\phi+n)\Gamma(\phi)}{\Gamma(\mu_b\phi)\Gamma(\phi+n)},
\end{aligned}
$$

the mean and the variance of the tilted beta distribution are:

$$
\begin{aligned}
E\left(Y\mid\mu_t,\mu_b,\phi,\theta\right) &= \theta\mu_t+(1-\theta)\mu_t \\
V\left(Y\mid\mu_t,\mu_b,\phi,\theta\right) &= E\left(Y^2\mid\mu_t,\mu_b,\phi,\theta\right)-E\left(Y\mid\mu_t,\mu_b,\phi,\theta\right)^2 \\
&= \left[\theta E\left(Y^2\right)+(1-\theta)E_b\left(Y^2\right)\right]-\left[\theta\mu_t+(1-\theta)\mu_b\right]^2 \qquad (7) \\
&= \theta V_t\left(Y\right)+(1-\theta)V_b\left(Y\right)+\theta\mu_t^2+(1-\theta)\mu_b^2-\left[\theta\mu_t+(1-\theta)\mu_b\right]^2 \\
&= \theta V_t\left(Y\right)+(1-\theta)V_b\left(Y\right)+\theta(1-\theta)\left(\mu_t+\mu_b\right)^2
\end{aligned}
$$

The rectangular beta distribution (Hahn, 2008) is a particular case of (4) when $\mu_t = 0.5$ (the slope of the tilted distribution is zero). By replacing this value of $\mu_t = 0$ in (4), the density function of the tilted beta distributions is defined by:

$$f\left(y\mid\mu,\phi,\theta\right)=\theta+(1-\theta)f_{Beta}\left(y\mid\mu,\phi\right), \qquad (8)$$

where $0 < y < 1$.

$(\mu_t, \mu_b, \phi, \theta)$ - Tilted Beta Binomial Distribution

Let $Y|p \sim Bin(m, p)$ be a random variable that follows the binomial distribution, where p follows the tilted beta distribution, $p \sim BI(\mu_t, \mu_b, \phi, \theta)$. Then Y follows a tilted beta binomial distribution with parameters μ_t, μ_b, ϕ and θ, denoted by $Y \sim BIB(\mu_t, \mu_b, \phi, \theta)$. The probability of this distribution is given by:

$$
\begin{aligned}
&f\left(y\mid\mu_t,\mu_t,\phi,\theta\right) \\
&= \int_0^1 f_{Bin}\left(y\mid m,p\right)\left[\theta c\left(p\mid\mu_t\right)+(1-\theta)f_{Beta}\left(p\mid\mu_b,\phi\right)\right]dp \\
&= \binom{m}{y}\Bigg[\theta\int_0^1\left[3(2\mu_t-1)(2p-1)+1\right]p^y(1-p)^{m-y}dp \\
&\quad +\frac{(1-\theta)\Gamma(\phi)}{\Gamma(\mu_b\phi)\Gamma(\phi(1-\mu_b))}*\int_0^1 p^{y+\mu_b\phi-1}(1-p)^{m-y+\phi(1-\mu_b)-1}dp\Bigg] \\
&= \theta\binom{m}{y}\left[\begin{array}{l}(12\mu_t-6)\int_0^1 p^{y+1}(1-p)^{m-y}dp \\ +(-6\mu_t+4)\int_0^1 p^y(1-p)^{m-y}dp\end{array}\right] \\
&\quad +(1-\theta)f_{BB_{(\mu_b,\phi)}}\left(y\mid\mu_b,\phi\right) \\
&= 2\theta\binom{m}{y}\left[\begin{array}{l}(6\mu_t-3)\frac{y+1}{y+1+m-y+1} \\ +(-3\mu_t+2)\end{array}\right]B\left(y+1,m-y+1\right) \\
&\quad +(1-\theta)f_{BB_{(\mu_b,\phi)}}\left(y\mid\mu_b,\phi\right) \\
&= 2\theta\binom{m}{y}\left[\frac{y(6\mu_t-3)+m(2-3\mu_t)+1}{m+2}\right]B\left(y+1,m-y+1\right) \\
&\quad +(1-\theta)f_{BB_{(\mu_b,\phi)}}\left(y\right), \qquad (9)
\end{aligned}
$$

where, $y = 0, 1,..., m$; $B(\cdot,\cdot)$ denotes the beta function and $f_{BB_{(\mu_b,\phi)}}(\cdot)$ denotes the probability function of the beta binomial distribution, parameterized in terms of the mean and the dispersion parameters.

The behavior of the $(\mu_t, \mu_b, \phi, \theta)$-tilted beta binomial probability function is illustrated in Fig. 1, for different vectors of parameter values.

The mean and variance of a random variable Y that follows the $(\mu_t, \mu_b, \phi, \theta)$-tilted beta binomial probability function are given by:

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

where, μ_b, V_b denote the mean and variance of the beta distribution, respectively and μ_t, V_t denote the mean and variance of the tilted beta distribution.
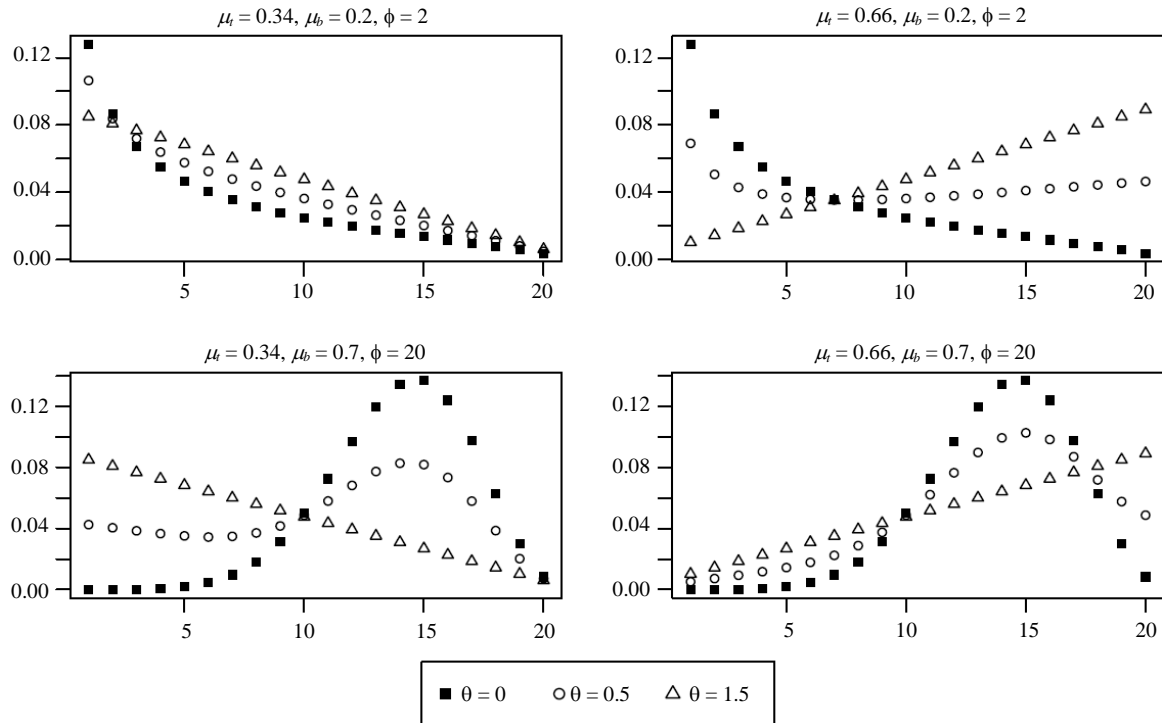
Fig. 1: Density function of the $(\mu_t, \mu_b, \phi, \theta)$-tilted beta binomial distribution

(μ_b, ϕ, θ)-Beta Rectangular Binomial Distribution

Let $Y|p \sim Bin(m, p)$ be a random variable that follows the binomial distribution, where p follows the beta rectangular distribution (8). Thus, Y follows the (μ_b, ϕ, θ)-beta rectangular binomial distribution. This density function can be obtained as a particular case of the tilted beta binomial distribution (9), by replacing μt by 0.5:

$$
f\left(y \mid \mu_b, \phi, \theta\right) = \binom{m}{y} \theta B\left(y+1, m-y+1\right) \\
+ \left(1-\theta\right) f_{BB\left(\mu_b, \phi\right)}\left(y \mid \mu_b, \phi\right)
\tag{10}
$$

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

From the equations of the mean (6) and variance (7) of the tilted beta binomial distribution, setting $\mu_t = 0.5$, the mean and variance of the rectangular beta distribution are obtained as:

$$
E(Y) = m\left[\frac{\theta}{2} + (1-\theta)\mu\right]
$$

$$
V(Y) = \left(m^2 - m\right)\left[\frac{\mu(1-\mu)}{1+\phi}(1-\theta)(1+\theta(1+\phi)) + \frac{\theta}{12}(4-3\theta)\right] \\
+ m\left[\frac{\theta}{2} + (1-\theta)\mu\right]\left[\frac{2-\theta}{2} - (1-\theta)\mu\right]
$$

Tilted Beta Binomial Regression Model

Let $Y \sim BIB(\mu_t, \mu_b, \phi, \theta)$, $i = 1, 2,..., n$, be independent random variables with tilted beta binomial distribution. Let $\mathbf{x}_i = (x_{i1},..., x_{is})^T$, $\mathbf{z}_i = (z_{i1},..., z_{ik})^T$ and $\mathbf{w}_i = (w_{i1},..., w_{il})^T$ the covariate vectors of μ_b, ϕ and θ regression structures and $\beta = (\beta_1,..., \beta_s)^T$, $= (\gamma_1,..., \gamma k)^T$ and $\delta = (\delta_1,..., \delta_l)^T$ the respective regression parameter vectors, such that:

$$
\mu_{bi} - \exp\left(x_i^T \beta\right) / \left(1 + \exp\left(x_i^T \beta\right)\right)
$$

$$
\phi_i = \exp\left(z_i^T \gamma\right)
$$

$$
\theta_i = \exp\left(w_i \delta\right) / \left(1 + \exp\left(w_i \delta\right)\right)
$$

Thus, if μ_t is assumed to be constant, the likelihood function of the $BIB(\mu_t, \mu_{bi}, \phi_i, \theta_i)$- regression model is:

$$
l\left(\mu_t, \left(\beta^T, \gamma^T, \delta^T\right)^T\right) = \frac{\exp\left(w_i \delta\right)}{1 + \exp\left(w_i \delta\right)} \sum_{\{i: y_i \geq 0\}} l_t\left(\mu_t \mid y_i\right) \\
+ \frac{1}{1 + \exp\left(w_i \delta\right)} \sum_{\{i: y_i \geq 0\}} l_{BB}\left(x_i^T \beta, z_i^T \gamma \mid y_i\right)
$$

with:

$$l_t\left(\mu_t \mid y_i\right) = \log\left\{2\binom{m}{y_i}\right\}$$

$$+\log\left\{\frac{y\left(6\mu_t - 3\right) + m\left(2 - 3\mu_t\right) + 1}{m + 2}\right\} \quad (11)$$

$$+\log\left\{B\left(y_i + 1, m - y_i + 1\right)\right\}$$

and:

$$l_{BB}\left(x_i^T\beta, z_i^T\gamma \mid y_i\right)$$

$$= \log\left\{f_{BB(\mu,\phi)}\left(y_i \left| \frac{\exp\left(x_i^T\beta\right)}{1 + \exp\left(x_i^T\beta\right)}, \exp\left(z_i^T\gamma\right)\right.\right)\right\},$$

where $B(\cdot,\cdot)$ represents the beta function.

In order to define the Bayesian tilted beta binomial regression model, the following a priori distributions are assumed for β, γ, δ and μ_t:

$$\beta \sim N(0, B)$$
$$\gamma \sim N(0, G)$$
$$\delta \sim N(0, D)$$
$$\mu_t \sim U(1/3, 2/3)$$

Seeds Germination Regression Models

The dataset analyzed in this section is available in Spiegelhalter *et al.* (2014) 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). These data were initially reported by Crowder (1978). The variables involved in the experiment are described below:

- **y**: Number of seeds germinated in each dish.
- **n**: Number of seeds initially arranged in each dish.
- **x₁**: Seed type (1) if it is *O. aegyptiaca* 75 and (2) if it is *O. aegyptica* 73.
- **x₂**: Root type (1) if it is bean and (2) if it is cucumber.

In this experiment, there are 21 observations (21 dishes). Since the variable Y counts the number of germinated seeds in each dish, this variable can be modeled by a linear regression TBB(μ_t, μ_b, ϕ, θ) model, which includes all the explanatory variables in each of the regression structures. After the process of eliminating the explanatory variables, the best model (the model with smallest Deviance Information Criteria (DIC) value) has the following regression structures:

$$logit\left(\mu_{ib}\right) = c1 + c2 * x_1 + c3 * x_2$$
$$\log\left(\phi_i\right) = a1 + a2 * x_2$$
$$logit\left(\theta_i\right) = b1 + b2 * x_2$$
$$\mu_t \sim U\left(1/3, 2/3\right)$$

where, $i = 1,\dots, 21$. The TBB(μ_t, μ_b, ϕ, θ) model was fitted to the data using OpenBUGS, a free program used for Bayesian regression based on the Gibbs algorithm (Spiegelhalter *et al.*, 2003). The posterior parameter inferences obtained from a sample of size 100000, burn-in of the first 10000 and taking one sample every 10 iterations to reduce autocorrelation, are summarized in Table 1. The DIC value of this model is 118.9. The posterior credibility interval for a regression parameter is given by real numbers a and b such that the posterior probability that the parameter estimates lie between a and b is 95%. These real numbers were obtained from the posterior sample assuming extreme tail samples of 2.5%.

In Table 1 the M.C. error denotes an estimation of the standard Monte Carlo error (*ErrorMC*), wich measures the distance between the posterior estimation of the mean and the mean of the posterior distribution, which is expected to converge to zero when the number of iterations goes to infinity. The Monte Carlo error estimates obtained using the OpenBugs software, close to zero for all the regression parameters, is given by $ErrorMC \approx SD/\sqrt{Iterations}$ (Flegal, 2008). The DIC value of this model is 118.9. According to Figure 2, Pearson's residuals are close to zero, taking values between -0.4 and 0.2 and have no tendency through the iterations.

Table 1: Posterior parameter estimates of *TBB*(μ_t, μ_b, α, θ) model

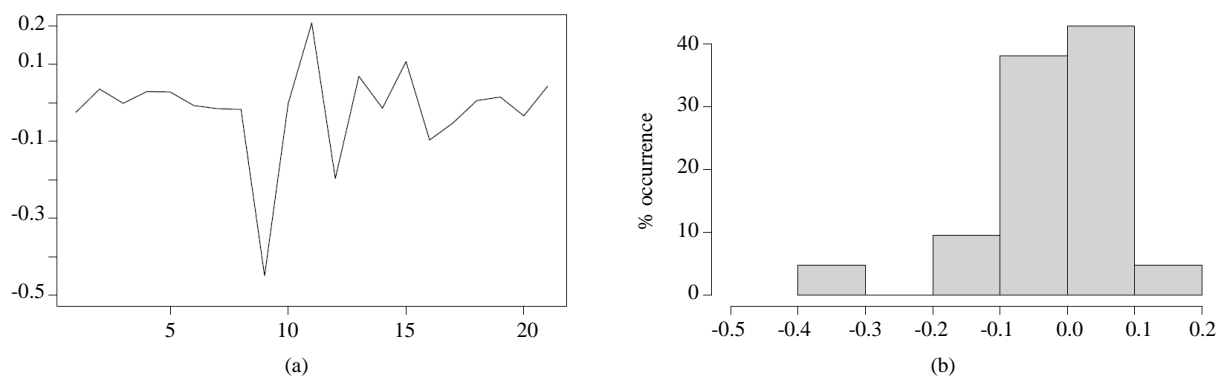| Parameter	Mean	S.D.	95% Cred. Interval	M.C. Error
$a1$	0.2967	1.2900	(-2.31, 2.763)	0.0239
$a2$	3.0260	1.5680	(0.5406, 6.552)	0.0363
$b1$	-0.6048	3.0300	(-6.604, 5.36)	0.0702
$b2$	-8.8400	6.0770	(-23.14, 0.1228)	0.0122
$c1$	-0.9004	0.4904	(-1.801, 0.1655)	0.0107
$c2$	-0.4582	0.2143	(-0.8892, -0.0497)	0.0047
$c3$	1.0240	0.2260	(0.5676, 1.469)	0.0049
μ_c	0.4958	0.0956	(0.3416, 0.657)	0.0020
deviance	115.6000	2.9800	(111.8, 123.1)	0.0642

Fig. 2: Pearson residuals – Seed germination; (a) Residuals versus iterations; (b) Residual histogram
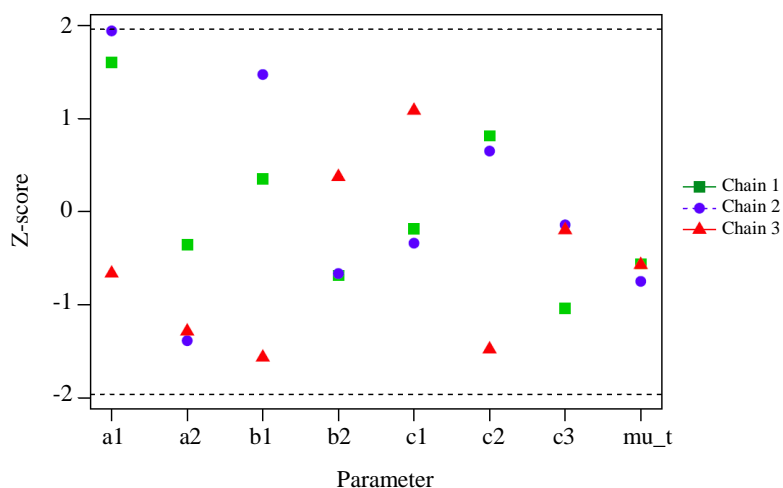

Fig. 3: Geweke convergence diagnostic - Seed germination

Chain Convergence in the Tilted Beta Binomial Model

In the parameter estimation process, three posterior samples were generated beginning from different starting values. In all chains, the autocorrelation is close to zero for a lag greater than or equal than 10 and a burn-in bigger than 10000.

To check the convergence of the chains, two convergence diagnoses were applied: The Geweke diagnostic (Geweke, 1992) and the Brooks-Gelman-Rubin convergence diagnostic (Brooks and Gelman, 1998). The Geweke-Brooks plot for the chains of the regression parameters can be observed in Fig. 3, where the value of the Z statistic versus the number of iterations is plotted to determine the burn-in of the chains. This figure shows that the statistic remains within the acceptance zone for a period of burn-in equal to zero. The second method applied is known as the Brooks-Gelman and Rubin convergence diagnostic. It was proposed by Brooks and Gelman (1998) and compares within-chain and between-chain variances through the estimation of the statistic of scale reduction R. Values of R well above 1 indicate that the chains have not converged. Figure 4 shows that for the regression parameters of this example, the R factor is very close to 1 after the 1000 iterations.

Models Comparison

In order to determine the performance of the proposed model, the following models also were fitted to the seed germination dataset: Binomial $Bin(n, p)$, beta binomial $BB(\mu, \phi)$ and beta rectangular binomial $BRB(\mu, \phi, \theta)$. The Residual Sum of Squares (RSS), the deviance and the deviance information Criterion (DIC) for each of these models are given in Table 2, which shows that the lowest average of the deviance and the lowest DIC value correspond to the tilted beta binomial and the beta rectangular binomial models, where the first one presents the lowest DIC value and therefore is the best model.
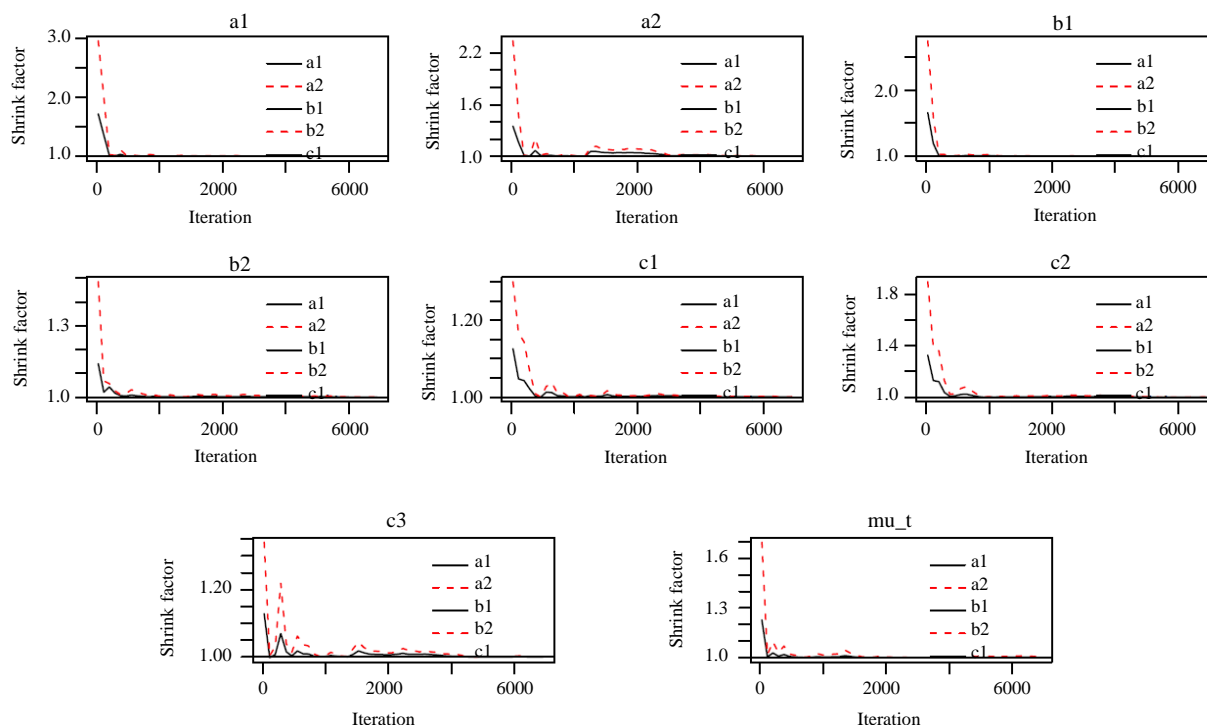
Fig. 4: Brooks-Gelman and Rubin convergence diagnostic - Seeds germination

Table 2: Statistics for model comparison - Seeds Germination

Model	RSS	DIC	Deviance			
			Mean	S.D.	Cred. Interval 95%	Median
Bin(n, p)	38.31100	122.5	119.3	2.449	(116.5,125.6)	118.6
BB(mu, ϕ)	2.49300	120.7	116.5	3.027	(112.7,124.2)	115.9
BRB(μ, ϕ, θ)	0.32892	120.0	115.9	3.080	(111.9,123.5)	115.5
TBB$(\mu_t, \mu_b, \phi, \theta)$	0.32637	118.9	115.6	2.980	(111.8,123.1)	115.0

Conclusion

In this paper two new distributions are proposed: The tilted beta binomial distribution and the beta rectangular binomial distribution. From these distributions, assuming that their parameters follow regression structures, new overdispersion regression models for count data are proposed: The tilted beta binomial regression model and the beta rectangular binomial regression model. These models are fitted using Bayesian methods and in the application, show better performance than the beta binomial regression models for statistical analysis of the seed germination dataset.

Given that the tilted beta distribution is flexible and allows considering varying amounts with greater likelihoods than the beta distribution in the extreme tail-area events, it permits accommodating different relative likelihoods of high versus low extreme tail-area events. Thus, the proposed tilted beta binomial regression model:

- Defines a more general overdispersion regression model than the beta binomial regression model.
- Allows considering count events with high or low likelihood of occurrence and better estimation of the regression parameters, credibility (or confidence) intervals and statistical inferences in the analysis of binomial-type overdispersion data.

In the analysis of seed germination dataset, the proposed tilted beta binomial regression model has better performance than the binomial and beta binomial regression models, in the sense that it has smaller DIC value and smaller RSS. Thus, the tilted beta binomial distributions can be more robust compared with the usual beta binomial distribution, in order to obtain a good flexibility of the proportion parameter p in the beta-binomial distribution.

Author's Contributions

All authors equally contributed in this work.

Ethics

This article is original and contains unpublished material. The corresponding author confirms that all of the other authors have read and approved the manuscript and no ethical issues involved.

References

Brooks, S. and A. Gelman, 1998. General methods for monitoring convergence of iterative simulations. J. Comput. Graphi. Stat., 7: 434-455.
DOI: 10.1080/10618600.1998.10474787

Cepeda, E. and D. Gamerman, 2005. Bayesian methodology for modeling parameters in the two parameter exponential family. Revista Estadistica, 57: 93-105.

Cepeda-Cuervo, E., 2001. Variability modeling in generalized linear models. Unpublished Ph.D. Thesis, Mathematics Institute, Universidade Federal do Rio de Janeiro.

Cepeda-Cuervo, E. and M.V. Cifuentes-Amado, 2017. Double generalized beta-binomial and negative binomial regression models. Revista Colombiana de Estadstica, 40: 141-141.
DOI: 10.15446/rce.v40n1.61779

Cepeda-Cuervo, E. and L. Garrido, 2015. Bayesian beta regression models with joint mean and dispersion modeling. Monte Carlo Meth. Applic., 21: 49-58.
DOI: 10.1515/mcma-2014-0007

Collet, D., 1991. Modeling Binary Data. 2nd Edn., Chapman Hall, London, ISBN-10: 0412388006, pp: 384.

Cox, D., 1983. Some remarks on overdispersion. Biometrika, 70: 269-274.

Crowder, M.J., 1978. Beta-binomial anova for proportions. Applied Stat., 27: 34-37.
DOI: 10.1093/biomet/70.1.269

Efron, B., 1986. Double exponential families and their use in generalized linear regression. J. Am. Stat. Assoc., 81: 709-721. DOI: 10.2307/2346223

Flegal, J.M., 2008. Monte Carlo standard errors for Markov chain Monte Carlo. Doctoral Dissertation, University of Minnesota Major, Statistics.

Geweke, J., 1992. Evaluating the Accuracy of Sampling-Based Approaches to Calculating Posterior Moments. In: Bayesian Statistics, Bernardo, J.M., J. Berger, A.P. Dawid and J.F.M. Smith (Eds.), Oxford University Press, Oxford, pp: 169-193.

Hahn, E., 2008. Mixture densities for project management activity times: A robust approach to pert. Eur. J. Operat. Res., 188: 450-459.
DOI: 10.1016/j.ejor.2007.04.032

Hahn, E. and M. Lopez Martın, 2015. Robust project management with the tilted beta distribution. SORT, 39: 253-272.

Hinde, J. and C. Demetrio, 1998. Overdispersion: Models and estimation. Comput. Stat. Data Anal., 27: 151-170. DOI: 10.1016/S0167-9473(98)00007-3

Jorgensen, B., 1997. Proper dispersion models. Brazilian J. Probability Stat., 11: 89-128.

Spiegelhalter, D., A. Thomas, N. Best and D. Lunn, 2003. WinBUGS user manual. MRC Biostatistics Unit, Institute of Public Health, Cambridge, UK.

Spiegelhalter, D., A. Thomas, N. Best, D. Lunn and K. Rice, 2014. BUGS: Bayesian inference using Gibbs sampling. MRC Biostatistics Unit, Cambridge, England.

Williams, D., 1982. Extra-binomial variation in logistic linear models. Applied Stat., 2: 144-188.
DOI: 10.2307/2347977

Appendix A OpenBUGS code for the TBB regression model

```
model {
#Likelihood:
for (i in 1: N) {
    zeros [i]<-0
    zeros [i] ~ dloglik (loglike [i])
    loglike [i]<-log (theta [i] * fd_b [i]+(1-theta [i] *
    fd_BB [i])
#fd_b: beta function part
#fd_BB: loglikelihood for beta binomial part
Fd_b [i]<-exp (log (2)+ logfact (n [i])- logfact (n [i]-y
[i])- logfact (y [i])
+loggam(y [i]+1) + loggam(n [i]-y [i]+1)-loggam(y
[i]+1+n [i]
-y [i]+1)+log (y [i] * (6 * mu_c-3)+n [i] * (2-
3*mu_c)+1)-log (n [i]+2))
Fd_BB [i]<-exp (logfact (n [i])- logfact (n [i]-y [i])-
logfact (y [i])
+loggam(phi   [i])+loggam(y  [i]  +  mu[i]*  phi
[i])+loggam(n [i]
-y [i]+phi [i]*(1-mu[i]))) -loggam(mu[i] * phi [i])-
loggam(phi [i]*(1-mu[i]))
-loggam(n [i]+phi [i])) log (phi [i])<-a [1]+a [2]*(x2
[i]+1)
logit (mu[i])<-c [1]+c [2]*(x1 [i]+1)+c [3]*(x2 [i]+1)
logit (the ta [i])<-b[1]+b [2]*(x2 [i]+1)
    }
#Priors:
  mu_t~ dunif (0.333333, 0.666666)
  for (i in 1: 2) {a [i] ~ dnorm(0, 0.1)}
  for (i in 1: 2) {b [i] ~ dnorm(0,0.1)}
  for (i in 1: 2) {c [i] ~ dnorm(0,0.1)}
}
```