Double Generalized Beta-Binomial and Negative Binomial Regression Models

Modelos de regresión beta-binomial y binomial negativa doblemente generalizados

Edilberto Cepeda-Cuervo[1,a], María Victoria Cifuentes-Amado[2,b]

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

[2]Departamento Matemáticas, Facultad de Ciencias, Universidad Nacional de Colombia, Bogotá, Colombia

---

Abstract

Overdispersion is a common phenomenon in count datasets, that can greatly affect inferences about the model. In this paper develop three joint mean and dispersion regression models in order to fit overdispersed data. These models are based on reparameterizations of the beta-binomial and negative binomial distributions. Finally, we propose a Bayesian approach to estimate the parameters of the overdispersion regression models and use it to fit a school absenteeism dataset.

Key words: Bayesian Approach, Beta-Binomial Distribution, Distribution, Gamma Distribution, Negative Binomial, Overdispersion, Poisson Distribution.

Resumen

La sobredispersión es un fenómeno común en conjuntos de datos de conteo, que puede afectar en alto grado las inferencias relacionadas con el modelo. En este artículo desarrollamos tres modelos de regresión conjunta de media y dispersión para ajustar datos sobredispersos. Estos modelos se basan en reparameterizaciones de las distribuciones beta-binomial y binomial negativa. Finalmente, proponemos un enfoque Bayesiano para la estimación de los parámetros de los modelos de regresión sobredispersos y lo utilizamos para ajustar un conjunto de datos de ausentismo escolar.

Palabras clave: distribución beta-binomial, distribución binomial negativa, distribución de Poisson, distribución gamma, enfoque bayesiano, sobredispersión.

---

[a]PhD. E-mail: ecepedac@unal.edu.co

[b]PhD (c). E-mail: mvcifuentesa@unal.edu.co

1. Introduction

The binomial and Poisson distributions are widely used to fit discrete count data. However, a serious complication arises when the variance of the response variable Y exceeds the nominal variance. This phenomenon, called overdispersion, can lead to underestimation of standard errors (Cox 1983) in addition to misleading inference of the regression parameters, such as confidence intervals (Breslow 1984). In practice, the mentioned phenomenon is very common and even the theoretical dispersion of a variable is sometimes considered an exception.

In order to fit such data, for both binomial and Poisson overdispersed data, many authors have proposed models and estimation methods. Regarding to binomial data, authors such as Williams (1982) and Collet (1991) have studied the extra-binomial variation, which is another name for overdispersion in binomial data, and have proposed methods such as the incorporation of an extra-binomial variation components in the maximum likelihood estimation of the loglinear model.

Moreover, Breslow (1984) analyzed extra-Poisson variation data (or overdispersed Poisson data) by extending the results in Williams (1982) to the Poisson case. Later, Lawless (1987) explored the robustness and efficiency of the methods used to deal with extra-Poisson variation in regression models. In turn, McCullagh & Nelder (1989) presented a more general discussion of overdispersion within the framework of generalized linear models .

Demétrio & Hinde (1998) stated that the different models for overdispersion can be categorized by two general approaches:

1. Those including additional parameters for the variance function.

2. Those assuming a two-stage model, meaning the distribution parameter is itself a random variable.

Demétrio & Hinde (1998) proposed the Hinde-Demétrio (HD) regression models in order to handle overdispersed count data and discussed the effect of the dispersion parameters on this fit. For this topic, it is worth mentioning the work of (Quintero-Sarmiento, Cepeda-Cuervo & Núñez-Antón 2012), who published a review about overdispersion and the different methods to model it, such as GLMs with random mean, the quasi-likelihood functions, and the double exponential families.

The aim of this paper is to introduce a new reparameterization of the beta binomial and the negative binomial distribution to propose reparameterized overdispersed regression models and develop Bayesian and classic methods to fit the proposed models. The paper is organized as follows: In Section 2, we define the (μ, ϕ)- beta binomial distribution and analyze its main statistical properties. Section 3 presents the (μ, α)- negative binomial distribution with its definition, mean, variance, graphics, density and characteristic functions. Section 4 develops the (μ, σ^2)- negative binomial distribution. In Section 5, the definitions of the proposed overdispersion models are established. In Section 6, a Bayesian approach

is proposed to fit the proposed models by applying the Metropolis Hastings algorithm. As an extension, in Section 7, we discuss the application of the Newton Rapson method to fit the proposed models. Finally, in Section 8 we present the motivating data to analyze the factors affecting school absenteeism in rural New South Wales.

2. (μ, ϕ)- Beta Binomial Distribution

There are many day-to-day phenomena which show binomial behavior. Therefore, improving the fit of binomial distribution models is important. However, most of these models in real experiments exhibit a significant amount of trials with estimated variance larger than the predicted by the theoretical binomial model (Demétrio & Hinde 1998). In this case, the beta-binomial model is an alternative to the binomial one since it captures the overdispersion and thereby results in a better fit of the observed data (Williams 1975). In order to correct estimation problems caused by overdispersion in binomial families, the p parameter (proportion of successes) is assumed to be a random variable following a beta distribution.

To build a reparameterization of the beta-binomial distribution, we first refer to the reparameterization of the beta distribution, which was proposed by Jørgensen (1997) and later in Cepeda-Cuervo (2001). If $X \sim Beta(\alpha, \beta)$, an appropriate reparameterization of the beta distribution in terms of the mean and dispersion parameter is given by:

$$f(x, \mu, \phi) = \frac{\Gamma(\phi)}{\Gamma(\mu\phi)\Gamma(\phi(1-\mu))} x^{\mu\phi-1}(1-x)^{\phi(1-\mu)-1} I_{(0,1)}(x) \qquad (1)$$

where $\mu = \frac{\alpha}{\alpha+\beta}$ and $\phi = \alpha + \beta$ (Ferrari & Cribari-Neto 2004). In (1), $\Gamma(.)$ denotes the gamma function. Regarding the mean and variance of the beta-binomial, we have:

$$E(X) = \mu \qquad (2)$$

$$Var(X) = \frac{\mu(1-\mu)}{\phi+1}$$

In this paper we use $p \sim Beta(\mu, \phi)$ to denote that p follows a beta distribution with mean μ and dispersion parameter ϕ. Thus, the (μ, ϕ)-beta binomial distribution is the distribution of a random variable Y such that, conditional to p, has a binomial distribution $Y \mid p \sim Bin(m, p)$, where p is a random variable with a beta distribution $p \sim Beta(\mu, \phi)$. We use the notation $Y \sim BB(\mu, \phi)$ to denote that Y has a beta-binomial distribution with mean μ and dispersion parameter ϕ.

The (μ, ϕ)-beta binomial density functions is determined by:

$$
\begin{aligned}
f(y) &= \binom{m}{y} \frac{\Gamma(\phi)}{\Gamma(\mu\phi)\Gamma(\phi(1-\mu))} \int p^{y+\mu\phi-1}(1-p)^{m-y+\phi(1-\mu)-1} I_{(0,1)}(p) \, dp \\
&= \binom{m}{y} \frac{\Gamma(\phi)}{\Gamma(\mu\phi)\Gamma(\phi(1-\mu))} \frac{\Gamma(y+\mu\phi)\Gamma(m-y+\phi(1-\mu))}{\Gamma(m+\phi)} \\
&= \binom{m}{y} \frac{\Gamma(\phi)}{\Gamma(\mu\phi)\Gamma(\phi(1-\mu))} B(y+\mu\phi, m-y+\phi(1-\mu)) \\
&= \binom{m}{y} \frac{B(y+\mu\phi, m-y+\phi(1-\mu))}{B(\mu\phi, \phi(1-\mu))}
\end{aligned}
\tag{3}
$$

where $B(x, y)$ is defined in terms of the gamma function as $B(x, y) := \Gamma(x)\Gamma(y)/\Gamma(x+y)$. The mean and variance of Y are:

$$
E(Y) = mE(p) = m\mu \tag{4}
$$

$$
\begin{aligned}
Var\left(\frac{Y}{m}\right) &= E\left[Var\left(\frac{Y}{m}\Big|p\right)\right] + Var\left[E\left(\frac{Y}{m}\Big|p\right)\right] \\
&= E\left[\left(\frac{1}{m}\right)p(1-p)|\mu, \phi\right] + Var(p|\mu, \phi) \\
&= \frac{1}{m}(\mu - E(p^2)) + \frac{\mu(1-\mu)}{\phi+1} \\
&= \frac{1}{m}\left(\mu - \mu^2 - \frac{\mu(1-\mu)}{\phi+1}\right) + \frac{\mu(1-\mu)}{\phi+1} \\
&= \frac{1}{m}\mu(1-\mu) + \left(\frac{m-1}{m}\right)\frac{\mu(1-\mu)}{\phi+1} \\
&= \frac{\mu(1-\mu)}{m}\left(1 + \frac{m-1}{\phi+1}\right)
\end{aligned}
$$

Therefore, the variance of a (μ, ϕ)-beta-binomial variable is:

$$
Var(Y) \quad = \quad m\mu(1-\mu)\left(\frac{\phi+m}{\phi+1}\right) \tag{5}
$$

The behavior of the beta-binomial density function is illustrated in Figure 1. Each of the four graphs represents the behavior of the beta-binomial distribution for different mean and dispersion values, where m (number of trials) is assumed to be constant and equal to 20. The title of each graph contains two numbers, which represent, respectively, the mean and dispersion parameters. For instance, in the first graph (top left box), $\mu = 0.3$ and $\phi = 2$, thus the mean of the beta-binomial variable is $m\mu = 6$, and since the dispersion parameter is small, the graph does not show a noticeable accumulation around the mean 6. On the other hand, a distribution like the third (bottom left box), in spite of having the same mean as the first, shows greater accumulation around mean 6 because the dispersion parameter is bigger ($\phi = 20$). The other graphs can be similarly interpreted.
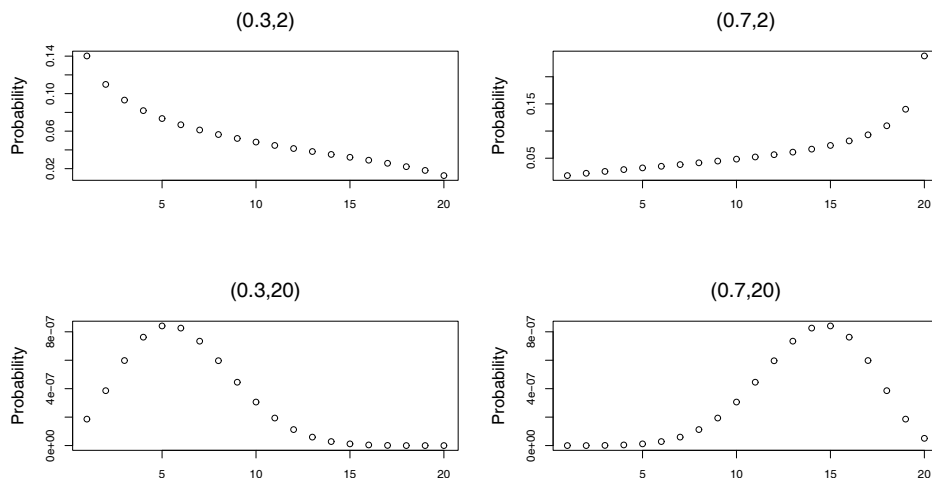
FIGURE 1: Reparameterized (μ,ϕ)-beta binomial probability function.

By definition, the characteristic function of the binomial distribution is given by:

$$
\begin{aligned}
E(e^{itY}) &= E(E(e^{itY} \mid p)) \\
&= E(((1-p) + pe^{it})^n) \\
&= E\left(\sum_{j=0}^{m} \binom{m}{j} (1-p)^{m-j} (pe^{it})^j \right) \\
&= E\left(\sum_{j=0}^{m} \binom{m}{j} \left(\sum_{k=0}^{m-j} \binom{m-j}{k} (-p)^k \right) p^j e^{itj} \right) \\
&= E\left(\sum_{j=0}^{m} \sum_{k=0}^{m-j} \binom{m}{j} \binom{m-j}{k} (-1)^k p^{k+j} e^{itj} \right) \\
&= \sum_{j=0}^{m} \sum_{k=0}^{m-j} \frac{m!}{j!k!} (-1)^k E(p^{k+j}) e^{itj}
\end{aligned}
$$

Thus, given that $p \sim Beta(\mu, \phi)$, the $E(p^{k+j})$ is the $(k+j)$-moment of the beta distribution. In fact, the characteristic function of the beta-binomial distribution can be seen as:

$$
E(e^{itY}) = \sum_{j=0}^{m} \sum_{k=0}^{m-j} \frac{m!}{j!k!} (-1)^k \left[\frac{d^{k+j}}{dt^{k+j}} \phi_p(t) \right]_{t=0} e^{itj} \tag{6}
$$

such that $\phi_p(t)$ is the characteristic beta distribution function.

3. (μ, α)- Negative Binomial Distribution

The negative binomial model, first proposed in Margolin, Kaplan & Zeiger (1981), assumes that the random variable X_i, conditioned on λ, follows a Poisson distribution $X \mid \lambda \sim \mathcal{P}(\lambda)$, where λ is itself a random variable with a gamma distribution $\lambda \sim G(\alpha, \beta)$, where $\mu = \alpha/\beta$, $\alpha > 0$ and $\beta > 0$.

Cepeda-Cuervo (2001) and Cepeda-Cuervo & Gamerman (2005) propose a new parameterization for the gamma distribution in terms of the mean μ and the shape parameter α. This reparameterization, the gamma density function is given by:

$$f(x; \mu, \alpha) = \frac{1}{x\Gamma(\alpha)} \left(\frac{\alpha x}{\mu} \right)^{\alpha} e^{-\frac{\alpha x}{\mu}}.$$

Hereafter, we use $\lambda \sim G(\mu, \alpha)$ to mean that λ follows a gamma distribution with mean μ and shape parameter α. With this reparameterization, the negative binomial density function, denoted $NB(\mu, \alpha)$, has a density function given by:

$$
\begin{aligned}
f(x; \alpha, \mu) &= \int_0^{\infty} \left(\frac{e^{-\lambda} \lambda^x}{x!} \right) \left(\frac{\alpha}{\mu} e^{-\frac{\alpha}{\mu}\lambda} \frac{\left(\frac{\alpha}{\mu}\lambda \right)^{\alpha-1}}{\Gamma(\alpha)} \right) d\lambda \\
&= \left(\frac{\alpha}{\mu} \right)^{\alpha} \frac{1}{x!\Gamma(\alpha)} \int_0^{\infty} e^{-\left(1+\frac{\alpha}{\mu}\right)\lambda} \lambda^{x+\alpha-1} d\lambda \\
&= \left(\frac{\alpha}{\mu} \right)^{\alpha} \frac{1}{x!\Gamma(\alpha) \left(1+\frac{\alpha}{\mu}\right)^{x+\alpha}} \int_0^{\infty} e^{-t} t^{x+\alpha-1} \, dt \\
&= \left(\frac{\alpha}{\mu} \right)^{\alpha} \frac{\Gamma(x+\alpha)}{x!\Gamma(\alpha) \left(1+\frac{\alpha}{\mu}\right)^{x+\alpha}}
\end{aligned}
\tag{7}
$$

The mean and variance of the random variable $X \sim NB(\mu, \alpha)$ are:

$$E(X) = E(\lambda) = \mu \tag{8}$$

$$
\begin{aligned}
Var(X) &= E(Var(X \mid \lambda)) + Var(E(X \mid \lambda)) \\
&= E(\lambda) + Var(\lambda) \\
&= \mu + \left(\frac{\mu^2}{\alpha} \right) \\
&= \frac{\mu(\mu + \alpha)}{\alpha}
\end{aligned}
\tag{9}
$$

It is possible to observe several phenomena which can be studied using negative binomial distributions. Some examples of the negative binomial distribution's applications are: the number of European red mites on apple leaves Demétrio, Kokonendji & Zocchi (2007); the number of coin flips necessary to get a determined value (whose domain are integers valued between 2 and infinity), or the

number of units for inspection until getting exactly a determined number of defective units from a production line. In sum, in the literature, this distribution is widely interpreted as the required number of independent Bernoulli experiments (success/failure events) to achieve k successes.

To observe the behavior of the gamma-Poisson density function, we change one parameter at a time and we get Figure 2: each graph refers to the negative binomial density function for different values of its parameters (μ and α). For instance, in the first (left-top box) there are two numbers, 5 and 8, meaning that for this case the mean of this distribution is 5 and the shape parameter is 8. The graph shows bigger values for the density function near 5, and in comparison with the cases $\mu = 5$, $\alpha = 4$ (top-right box), this first graph has greater accumulation around the mean. The other graphs can be interpreted in the same way.
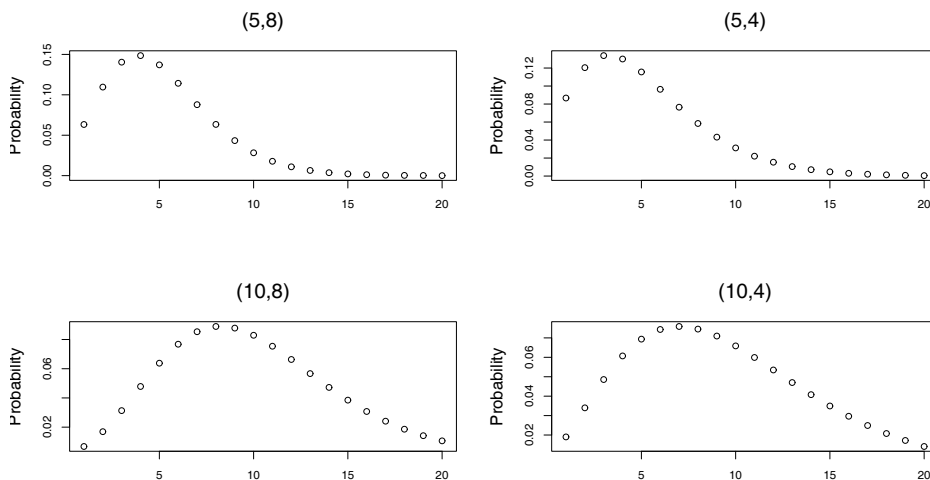

FIGURE 2: Reparameterized (μ,α)-negative binomial probability function.

From (7), the characteristic function of the reparameterized negative binomial distribution is given by:

$$\phi_X(t) = \left(\frac{1 - \frac{\mu}{\mu+\alpha}}{1 - \frac{\mu}{\mu+\alpha}e^{it}} \right)^{\alpha}$$

$$= \left(\frac{\alpha}{\mu(1 - e^{it}) + \alpha} \right)^{\alpha} \tag{10}$$

Moreover, by making $r = \alpha$ and $p = \mu/(\mu + \alpha)$ in equation (7), we get a usual parameterized density function for the negative binomial distribution:

$$
\begin{aligned}
f(x; r, p) &= \left(\frac{1-p}{p}\right)^r \frac{\Gamma(x+r)}{x!\Gamma(r)\left(1 + \frac{1-p}{p}\right)^{x+r}} \\
&= \left(\frac{1-p}{p}\right)^r \frac{\Gamma(x+r)}{x!\Gamma(r)\left(\frac{1}{p}\right)^{x+r}} \\
&= \frac{\Gamma(x+r)}{x!\Gamma(r)} p^x (1-p)^r
\end{aligned}
\tag{11}
$$

where $x = 0, 1, 2, \ldots$ It $X \sim NB(r, p)$ is used to denote a discrete random variable X follows a negative binomial distribution with parameters r and p defined in (11), where X can be intuitively considered as the number of successes in a sequence of independent and identically distributed Bernoulli trials before r failures occur and with a success probability equal to p in each trial.

4. $\left(\mu, \sigma^2\right)$ - Negative Binomial Distribution

The (μ, σ^2)-negative binomial model assumes that the random variable X, conditioned on λ, follows a Poisson distribution, $X \mid \lambda \sim \mathcal{P}(\lambda)$, where λ is itself a random variable that has gamma distribution with mean $\mu > 0$ and variance σ^2. Thus, $E(X) = \mu$ and $Var(X) = \sigma^2$. We use the notation $NB(\mu, \sigma^2)$ to denote that X follows a negative binomial distribution with mean μ and variance σ^2.

Thus, given that $\alpha = \mu^2/(\sigma^2 - \mu)$, from equations (7) and (10) the (μ, σ^2)-negative binomial distribution and its characteristic function are given respectively by:

$$
\begin{aligned}
f(x; \mu, \sigma^2) &= \left(\frac{\mu}{\sigma^2 - \mu}\right)^{\frac{\mu^2}{\sigma^2 - \mu}} \frac{\Gamma\left(x + \frac{\mu^2}{\sigma^2 - \mu}\right)}{x!\Gamma(\frac{\mu^2}{\sigma^2 - \mu})\left(\frac{\sigma^2}{\sigma^2 - \mu}\right)^{x + \frac{\mu^2}{\sigma^2 - \mu}}} \\
&= \left(\frac{\mu}{\sigma^2}\right)^{\frac{\mu^2}{\sigma^2 - \mu}} \frac{\Gamma\left(x + \frac{\mu^2}{\sigma^2 - \mu}\right)}{x!\Gamma(\frac{\mu^2}{\sigma^2 - \mu})\left(\frac{\sigma^2}{\sigma^2 - \mu}\right)^{x}}
\end{aligned}
\tag{12}
$$

$$
\begin{aligned}
\phi_X(t) &= \left(\frac{\frac{\mu^2}{\sigma^2 - \mu}}{\mu(1 - e^{it}) + \frac{\mu^2}{\sigma^2 - \mu}}\right)^{\frac{\mu^2}{\sigma^2 - \mu}} \\
&= \left(\frac{\mu}{\sigma^2 + (\mu - \sigma^2)e^{it}}\right)^{\frac{\mu^2}{\sigma^2 - \mu}}
\end{aligned}
\tag{13}
$$

In Figure (3), each graph represents the $NB(\mu, \sigma^2)$ distribution for different values of mean and variance. The first number which appears in the top of each box refers to the mean μ and the second one to the variance σ^2. For instance in the first (top-left box) the numbers 8 and 10 respectively, are the mean and variance of the $NB(\mu, \sigma^2)$ distribution.
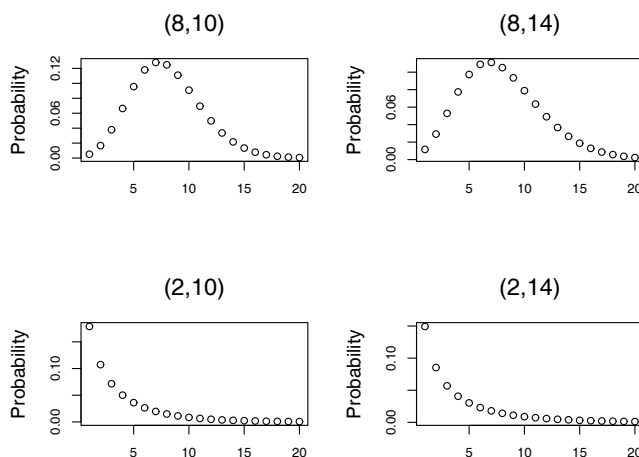

FIGURE 3: Reparameterized NB(μ,σ^2) probability function.

5. A New Class of Overdispersed Regression Models

Let Y_i, $i = 1 \ldots n$, a sample of a variable of interest; $\mathbf{x_i} = (x_{i1}, \ldots, x_{is})'$ and $\mathbf{z_i} = (z_{i1}, \ldots, z_{ik})'$, $i = 1, 2, \ldots, n$, vector values of covariates, and $\boldsymbol{\beta} = (\beta_1, \ldots, \beta_s)'$ and $\boldsymbol{\gamma} = (\gamma_1, \ldots, \gamma_k)'$, two vectors of the regression parameters. Using this notation, we define the following overdispersed regression models:

Definition 1. The $BB(\mu, \phi)$ regression model is defined by:

- A random component: Let $Y_i \sim BB(\mu_i, \phi_i)$, $i = 1 \ldots n$; a sample of n independent random variables.

- The linear predictors η_{1i} and η_{2i}, such that $\eta_{1i} = \mathbf{x_i'}\boldsymbol{\beta}$ and $\eta_{2i} = \mathbf{z_i'}\boldsymbol{\gamma}$.

- Link functions $h(.)$ and $g(.)$, such that $h(\mu_i) := \eta_{1i}$ and $g(\phi_i) = \eta_{2i}$.

  In this model, the usual links are the logit function for the mean and the logarithmic function for the dispersion parameter.

Definition 2. The $NB(\mu, \alpha)$ regression model is defined by:

- A random variable: Let $Y_i \sim NB(\mu_i, \alpha_i)$, $i = 1 \ldots n$, independent random variables.

- Linear predictors: η_{1i} and η_{2i} such that: $\eta_{1i} = \mathbf{x_i'}\boldsymbol{\beta}$ and $\eta_{2i} = \mathbf{z_i'}\boldsymbol{\gamma}$.

- Link functions: $h(.)$ and $g(.)$, such that $h(\mu_i) := \eta_{1i}$ and $g(\alpha_i) = \eta_{2i}$.

  For this model, the usual link functions are the logarithmic functions for both h and g.

The $NB(\mu, \sigma^2)$ regression model is defined as the $NB(\mu, \alpha)$ regression model.

6. Bayesian Estimation

In following subsections we develop the Bayesian method for the three proposed two-stage models: the (μ, ϕ)-beta-binomial, the (μ, α)-negative binomial and the (μ, σ^2)-negative binomial models. In these models, we have two vector parameters to estimate: the mean regression parameters $\boldsymbol{\beta}$, and the shape (or variance) regression parameters $\boldsymbol{\gamma}$. The Bayesian method used to fit the proposed models is defined in the next nine points following Cepeda-Cuervo (2001).

1. Let Y_i, $i = 1, 2, \ldots, n$, be n independent observed values obtained from one of the two-parameter distributions.

2. The regression structures are defined as follows:

$$
\begin{aligned}
h(\mu_i) &= \mathbf{x_i'}\boldsymbol{\beta} = \eta_{1i} \\
g(\tau_i) &= \mathbf{z_i'}\boldsymbol{\gamma} = \eta_{2i},
\end{aligned} \tag{14}
$$

   where $\boldsymbol{\beta} = (\beta_1, \ldots, \beta_s)'$, $\boldsymbol{\gamma} = (\gamma_1, \ldots, \gamma_k)'$ are vectors of unknown regression parameters related to the mean and shape (or variance) parameters, respectively, such that $s + k < m$. The vectors $\mathbf{x_i} = (x_{i1}, \ldots, x_{is})'$, $\mathbf{z_i} = (z_{i1}, \ldots, z_{ik})'$ correspond to the i-th vector values of covariates and η_{1i}, η_{2i} are the linear predictors. We assume $x_{i1} = 1, z_{i1} = 1, \forall i = 1, \ldots, n$.

3. The link function should be strictly monotonic, twice differentiable in classic regression, and once in the Bayesian approach.

4. Without loss of generality, the mean and shape (or variance) regression parameters are assumed to have independent normal prior distributions:

$$
\begin{aligned}
\boldsymbol{\beta} &\sim N(\mathbf{b}, \mathbf{B}) \\
\boldsymbol{\gamma} &\sim N(\mathbf{g}, \mathbf{G})
\end{aligned}
$$

5. Let $L(\boldsymbol{\beta}, \boldsymbol{\gamma} \mid \mathbf{Y}, \mathbf{X}, \mathbf{Z})$ be the likelihood function and $p(\boldsymbol{\beta}, \boldsymbol{\gamma})$ the joint prior distribution. The likelihood function is $L(\boldsymbol{\theta} \mid Y) = \Pi f(y_i \mid \boldsymbol{\theta})$, where $\boldsymbol{\theta} = (\boldsymbol{\beta}, \boldsymbol{\gamma})'$ is the vector of the regression parameters.

6. As the posterior distribution $\pi(\boldsymbol{\beta}, \boldsymbol{\gamma}) \propto L(\boldsymbol{\beta}, \boldsymbol{\gamma})p(\boldsymbol{\beta}, \boldsymbol{\gamma})$ is analytically intractable, we propose using the Metropolis Hastings algorithm to get samples of the posterior parameters from the conditional distributions $\pi_{\boldsymbol{\beta}}(\boldsymbol{\beta} \mid \boldsymbol{\gamma}, \mathbf{Y}, \mathbf{X}, \mathbf{Z})$ and $\pi_{\boldsymbol{\gamma}}(\boldsymbol{\gamma} \mid \boldsymbol{\beta}, \mathbf{Y}, \mathbf{X}, \mathbf{Z})$.

7. Since the posterior conditional distribution $\pi(\boldsymbol{\beta} \mid \boldsymbol{\gamma})$ is analytically intractable, we propose to use working variables to build a kernel transition function to propose posterior samples of the $\boldsymbol{\beta}$ parameter vector. This working variable is determined by the first order Taylor approximation of $h(.)$ around the current parameter value of μ_i:

$$h(y_i) \simeq h(\mu_i) + h'(\mu_i)(y_i - \mu_i) = \tilde{y}_{1i},$$

where $E(y_i) = \mu_i$, and

$$E(\tilde{y}_{1i}) = \mathbf{x}_i'\boldsymbol{\beta}$$
$$Var(\tilde{y}_{1i}) = [h'(\mu_i)]^2 \, Var(y_i).$$

Therefore, if $\boldsymbol{\beta}^{(c)}$ and $\boldsymbol{\gamma}^{(c)}$ are the current values of the parameters, then:

$$\tilde{y}_{1i} = \mathbf{x}_i'\boldsymbol{\beta}^{(c)} + h'\left(h^{-1}\left(\mathbf{x}_i'\boldsymbol{\beta}^{(c)}\right)\right)\left[y_i - h^{-1}\left(\mathbf{x}_i'\boldsymbol{\beta}^{(c)}\right)\right], \forall i = 1 \ldots n \quad (15)$$

8. For this proposal, we assume the kernel transition function given by:

$$q_1(\boldsymbol{\beta} \mid \boldsymbol{\gamma}) = N(\mathbf{b}^*, \mathbf{B}^*) \quad (16)$$

where

$$\mathbf{b}^* = \mathbf{B}^*(\mathbf{B}^{-1}\mathbf{b} + \mathbf{X}'\boldsymbol{\Sigma}^{-1}\tilde{\mathbf{Y}}_1)$$
$$\mathbf{B}^* = (\mathbf{B}^{-1} + \mathbf{X}'\boldsymbol{\Sigma}^{-1}\mathbf{X})^{-1}$$

$\tilde{\mathbf{Y}}_2 = (\tilde{y}_{11}, \ldots, \tilde{y}_{2n})'$ and $\boldsymbol{\Sigma} = Diag\left(Var\left(\tilde{y}_{1i}\right)\right)$. With the kernel function (16), the values of $\boldsymbol{\beta}$, which will appear in the sample of the posterior distribution $\pi(\boldsymbol{\beta}, \boldsymbol{\gamma})$, will be generated.

9. Since the posterior conditional distribution $\pi(\boldsymbol{\gamma} \mid \boldsymbol{\beta})$ is analytically intractable, it is necessary to use working variables and a second kernel function to generate posterior samples of the $\boldsymbol{\gamma}$ parameters. To do this, Cepeda-Cuervo (2001) assumes there are variables t_i, such that $E(t_i) = \tau_i$, where τ_i is the second parameter to be modeled, and that $\tau_i = g^{-1}(\mathbf{z}_i'\boldsymbol{\gamma})$. So, the working variables related to the second parameter, denoted by $\tilde{y}_{2i}$ are given by:

$$g(t) \simeq g(\tau) + g'(\tau)(t - \tau) = \tilde{y}_{2i} \quad (17)$$

The mean and variance of these working variables are given by:

$$E(\tilde{y}_{2i}) = \mathbf{z}_i'\boldsymbol{\gamma}$$
$$Var(\tilde{y}_{2i}) = [g'(\tau_i)]^2 \, Var(t_i)$$

Then, the working variables are:

$$\tilde{y}_{2i} = \mathbf{z}_i'\boldsymbol{\gamma}^{(c)} + g'\left(g^{-1}\left(\mathbf{z}_i'\boldsymbol{\gamma}^{(c)}\right)\right)\left[t_i - g^{-1}\left(\mathbf{z}_i'\boldsymbol{\gamma}^{(c)}\right)\right], \forall i = 1 \ldots m \quad (18)$$

10. The second kernel transition function, q_2, is

$$q_2\left(\boldsymbol{\gamma} \mid \boldsymbol{\beta}\right) = N(\mathbf{g}^*, \mathbf{G}^*) \tag{19}$$

where

$$\mathbf{g}^* = \mathbf{G}^*\left(\mathbf{G}^{-1}\mathbf{g} + \mathbf{Z}'\boldsymbol{\Psi}^{-1}\tilde{\mathbf{Y}}_2\right)$$
$$\mathbf{G}^* = \left(\mathbf{G}^{-1} + \mathbf{Z}'\boldsymbol{\Psi}^{-1}\mathbf{Z}\right)^{-1},$$

$\tilde{\mathbf{Y}}_2 = (\tilde{y}_{21}, \ldots, \tilde{y}_{2n})'$ and $\boldsymbol{\Psi} = Diag(Var\left(\tilde{y_{2i}}\right))$.

Once the kernel transition functions and working variables are established, the Metropolis Hastings algorithm is defined by the following steps:

1. Begin the chain iteration counter at $j = 1$.

2. Set initial chain values $\boldsymbol{\beta}^{(0)}$ and $\boldsymbol{\gamma}^{(0)}$ for $\boldsymbol{\beta}$ and $\boldsymbol{\gamma}$, respectively.

3. Propose a new value δ for $\boldsymbol{\beta}$, generated from the kernel transition function $q_1(\boldsymbol{\beta}^{(j-1)}, \cdot)$.

4. Calculate the acceptance probability for the new value δ: $\alpha\left(\boldsymbol{\beta}, \delta\right)$. For large values of this probability, the new value is accepted. If the movement is accepted, then $\boldsymbol{\beta}^{(j)} = \delta$, otherwise $\boldsymbol{\beta}^{(j)} = \boldsymbol{\beta}^{(j-1)}$.

5. Propose a new value δ for $\boldsymbol{\gamma}$, generated from the kernel transition function $q_2(\boldsymbol{\gamma}^{(j-1)}, \cdot)$.

6. Calculate the acceptance probability for the new value δ: $\alpha\left(\boldsymbol{\gamma}, \delta\right)$. For large values of this probability, the new value is accepted. If the movement is accepted, then $\boldsymbol{\gamma}^{(j)} = \delta$, otherwise $\boldsymbol{\gamma}^{(j)} = \boldsymbol{\gamma}^{(j-1)}$.

7. Return to 3 until convergence.

6.1. Working Variables in the BB(μ, ϕ) Regression Model

We assume the (μ, ϕ)-beta-binomial regression model defined in Section 5 with the logit link function for the mean and logarithm link function for the dispersion parameter. Thus it follows from equation (15) that the working variables to define the kernel transition function to propose samples of the mean regression parameters are:

$$\tilde{y}_{1i} = \mathbf{x}'_i\boldsymbol{\beta}^{(c)} + \frac{y_i/n_i - \mu_i^{(c)}}{\mu_i^{(c)}(1 - \mu_i^{(c)})}, \forall\ i = 1 \ldots m$$

To propose a kernel transition function to obtain samples of the dispersion regression parameter vector $\boldsymbol{\gamma}$, we use the working variables established in equation (18). These working variables were proposed in Cepeda-Cuervo, Migon, Garrido

& Achcar (2014) in their framework of the generalized linear models with random effects. In order to get these working variables, we have to ascertain the expression for t_i such that $E(t_i) = \phi$. Then, with $t_i = \phi_i^{(c)} y_i / \mu_i^{(c)} n_i$, using equation (18), this working variable is defined by:

$$\tilde{y}_{2i} = \mathbf{z}_i' \boldsymbol{\gamma}^{(c)} + \frac{\phi_i^{(c)} y_i / \mu_i^{(c)} n_i - \phi_i^{(c)}}{\phi_i^{(c)}} = \mathbf{z}_i' \boldsymbol{\gamma}^{(c)} + \frac{y_i}{\mu_i^{(c)} n_i} - 1, \quad \forall i = 1 \ldots m,$$

where $\mu_i^{(c)} = e^{\mathbf{x}_i' \boldsymbol{\beta}^{(c)}} / \left(1 + e^{\mathbf{x}_i' \boldsymbol{\beta}^{(c)}}\right)$ and $\phi_i^{(c)} = e^{\mathbf{z}_i' \boldsymbol{\gamma}^{(c)}}$.

The variances of these working variables are, respectively:

$$\begin{aligned}
Var(\tilde{y}_{1i}) &= Var\left(\frac{y_i}{n_i}\right) \left[\mu_i^{(c)}(1 - \mu_i^{(c)})\right]^{-2} \\
&= \frac{\mu_i^{(c)}(1 - \mu_i^{(c)})}{n_i} \left(\frac{\phi_i^{(c)} + n_i}{\phi_i^{(c)} + 1}\right) \left[\mu_i^{(c)}(1 - \mu_i^{(c)})\right]^{-2} \\
&= \left(\frac{\phi_i^{(c)} + n_i}{\phi_i^{(c)} + 1}\right) \left[n_i \mu_i^{(c)}(1 - \mu_i^{(c)})\right]^{-1}, \quad \forall i = 1 \ldots m \\
Var(\tilde{y}_{2i}) &= Var(y_i) \left[\mu_i^{(c)} n_i\right]^{-2} \\
&= n_i \mu_i^{(c)} \left(1 - \mu_i^{(c)}\right) \left(\frac{\phi_i^{(c)} + n_i}{\phi_i^{(c)} + 1}\right) \left[\mu_i^{(c)} n_i\right]^{-2} \\
&= \left(\frac{\phi_i^{(c)} + n_i}{\phi_i^{(c)} + 1}\right) \left[\frac{1 - \mu_i^{(c)}}{n_i \mu_i^{(c)}}\right], \quad \forall i = 1 \ldots m.
\end{aligned}$$

6.2. Working Variables in the $NB(\mu, \alpha)$ Regression Model

In the $NB(\mu, \alpha)$ regression model defined in Section 5, assuming the logarithmic function for both the mean and the shape parameters from the first-order Taylor approximation with $t_i = y_i$, the working variables used to define the kernel transition function to propose samples of the mean regression parameters are given by:

$$\tilde{y}_{1i} = \mathbf{x}_i' \boldsymbol{\beta}^{(c)} + \frac{1}{\mu_i^{(c)}} \left(y_i - \mu_i^{(c)}\right) = \mathbf{x}_i' \boldsymbol{\beta}^{(c)} + \frac{y_i}{\mu_i^{(c)}} - 1, \quad \forall i = 1 \ldots m \qquad (20)$$

As developed in Cepeda-Cuervo (2001), Cepeda-Cuervo & Achcar (2009) Cepeda-Cuervo & Gamerman (2005) and Cepeda-Cuervo et al. (2014), the working variables used to propose samples of the shape regression parameter require a variable t such that $E(t) = \alpha$, thus $t = \alpha y / \mu$. Then, from the first order Taylor approximation of the logarithmic function around α, the working variables are:

$$\tilde{y}_{2i} = \mathbf{z}_i' \boldsymbol{\gamma}^{(c)} + \frac{1}{\alpha_i^{(c)}} \left(\frac{\alpha_i^{(c)} y_i}{\mu_i^{(c)}} - \alpha_i^{(c)}\right) = \mathbf{z}_i' \boldsymbol{\gamma}^{(c)} + \frac{y_i}{\mu_i^{(c)}} - 1, \quad \forall i = 1 \ldots m \qquad (21)$$

The variances of these working variables are, respectively:

$$Var(\tilde{y}_{1i}) = Var(y_i) \left[\mu_i^{(c)} \right]^{-2} = \frac{\mu_i^{(c)} \left(\mu_i^{(c)} + \alpha_i^{(c)} \right)}{\alpha_i^{(c)}} \left[\mu_i^{(c)} \right]^{-2}, \quad \forall i = 1 \ldots m$$

$$Var(\tilde{y}_{2i}) = Var(y_i) \left[\mu_i^{(c)} \right]^{-2} = Var(\tilde{y}_{1i}), \quad \forall i = 1 \ldots m.$$

6.3. Working Variables in the $NB(\mu, \sigma^2)$ Regression Model

Assuming the logarithm link function for the mean and variance parameters, the working variables $\tilde{y}_{1i}$, obtained using $t = Y$; and $\tilde{y}_{2i}$, obtained using $t = \frac{\sigma^2 Y}{\mu}$, remain as in equations (20) and (21), respectively. Therefore, the variances of the working variables are:

$$Var(\tilde{y}_{1i}) = Var(y_i) \left[\mu_i^{(c)} \right]^{-2} = \frac{\sigma_i^{2(c)}}{\mu_i^{(c)2}}, \quad \forall i = 1 \ldots m$$

$$Var(\tilde{y}_{2i}) = Var(y_i) \left[\mu_i^{(c)} \right]^{-2} = \frac{\sigma_i^{2(c)}}{\mu_i^{(c)2}}, \quad \forall i = 1 \ldots m$$

7. Proposed Regression Models: A Classic Approach

In order to obtain the parameter estimates of the proposed models using the Newton Raphson algorithm, in this section, we develop the first-order and second-order partial derivatives of the logarithm of the likelihood functions for each of the proposed models. For each of the regression models defined in Section 5, the logarithm of the likelihood function is given by:

$$L(\boldsymbol{\theta}) = \sum_{i=1}^{n} l_i(\boldsymbol{\theta}), \tag{22}$$

where $l_i(\boldsymbol{\theta}) = \log f(x_i \mid \boldsymbol{\theta})$ and $\boldsymbol{\theta} = (\boldsymbol{\beta}, \boldsymbol{\gamma})$ is the vector of the regression parameters.

Therefore, the first-order partial derivatives of the logarithm of the likelihood function are given by:

$$\frac{\partial l_i}{\partial \beta_j} = \frac{\partial l_i}{\partial \mu_i} \frac{\partial \mu_i}{\partial \eta_{1i}} \frac{\partial \eta_{1i}}{\partial \beta_j} = \frac{\partial l_i}{\partial \mu_i} \frac{\partial \mu_i}{\partial \eta_{1i}} x_{ji} \quad j = 1, \ldots, s \tag{23}$$

$$\frac{\partial l_i}{\partial \gamma_j} = \frac{\partial l_i}{\partial \alpha_i} \frac{\partial \alpha_i}{\partial \eta_{2i}} \frac{\partial \eta_{2i}}{\partial \gamma_j} = \frac{\partial l_i}{\partial \alpha_i} \frac{\partial \alpha_i}{\partial \eta_{2i}} z_{ji} \quad j = 1, \ldots, k \tag{24}$$

And the second-order partial derivatives are given by:

$$\frac{\partial^2 l_i}{\partial \beta_s \partial \beta_j} = \left[\frac{\partial^2 l_i}{\partial \mu_i^2} \left(\frac{\partial \mu_i}{\partial \eta_{1i}} \right)^2 + \frac{\partial l_i}{\partial \mu_i} \left(\frac{\partial^2 \mu_i}{\partial \eta_{1i}^2} \right) \right] x_{ji} x_{si}, \quad j, s = 1, \ldots, p \tag{25}$$

$$\frac{\partial^2 l_i}{\partial \gamma_s \partial \beta_j} = \left(\frac{\partial^2 l_i}{\partial \alpha_i \partial \mu_i} \frac{\partial \alpha_i}{\partial \eta_{2i}} \frac{\partial \mu_i}{\partial \eta_{1i}} \right) x_{ji} z_{si}, \qquad j = 1, \ldots, p, s = 1, \ldots, r \qquad (26)$$

$$\frac{\partial^2 l_i}{\partial \gamma_s \partial \gamma_j} = \left[\frac{\partial^2 l_i}{\partial \alpha_i^2} \left(\frac{\partial \alpha_i}{\partial \eta_{2i}} \right)^2 + \frac{\partial l_i}{\partial \alpha_i} \left(\frac{\partial^2 \alpha_i}{\partial \eta_{2i}^2} \right) \right] z_{ji} z_{si}, \quad j, s = 1, \ldots, r \qquad (27)$$

Thus, denoting by H the Hessian matrix and by $\mathbf{q}$ the vector of first derivatives of the logarithm of the likelihood function, the equation of the Newton-Rapson algorithm,

$$\boldsymbol{\theta}^{(k+1)} = \boldsymbol{\theta}^{(k)} - (\boldsymbol{H}^k)^{-1} \boldsymbol{q}^k, \qquad (28)$$

is well defined. The parameter estimates are obtained by setting initial parameter values and applying the Newton-Rapson equation until a convergence criterion is satisfied.

7.1. Fitting the BB(μ, ϕ) Regression Model

In order to fit the BB(μ, ϕ) regression model, if $Y_i \sim BB(\mu_i, \phi_i)$, $i = 1, \ldots, n$; is a sample of the variable of interest, the logarithm of the likelihood function is given by:

$$l_i(\theta) = \log \binom{m_i}{y_i} + \log(B(y_i + \mu_i \phi_i, m_i - y_i + \phi_i(1 - \mu_i))) - \log(B(\mu_i \phi_i, \phi_i(1 - \mu_i))),$$

where l_i is the i-th component of the likelihood function defined in (22).

Thus, using the development in equations (23) and (24), the first order derivatives of the logarithm of the likelihood function are given by:

$$\frac{\partial l_i}{\partial \beta_j} = [\Psi(y_i + \mu_i \phi_i) - \Psi(m_i - y_i + \phi_i(1 - \mu_i)) - \Psi(\mu_i \phi_i) + \Psi(\phi_i(1 - \mu_i))]$$
$$\times \left(\frac{\phi_i}{g'(\mu_i)} \right) x_{ji}$$
$$\frac{\partial l_i}{\partial \gamma_j} = [\Psi(y_i + \mu_i \phi_i) - \Psi(m_i - y_i + \phi_i(1 - \mu_i)) - \Psi(\mu_i \phi_i) + \Psi(\phi_i(1 - \mu_i))]$$
$$\times \left(\frac{\mu_i}{h'(\phi_i)} \right) z_{ji},$$

where $\Psi(z)$ represents the digamma function, $\Psi(z) := \frac{d}{dz} \ln \Gamma(z)$.

The second-order partial derivatives are given by equations (25), (26), and (27), where:

$$\frac{\partial l_i}{\partial \mu_i} = \phi_i \left[\Psi(y_i + \mu_i \phi_i) - \Psi(m_i - y_i + \phi_i(1 - \mu_i)) - \Psi(\mu_i \phi_i) + \Psi(\phi_i(1 - \mu_i)) \right]$$

$$\frac{\partial^2 l_i}{\partial \mu_i^2} = \phi_i^2 \left[\Psi'(y_i + \mu_i \phi_i) + \Psi'(m_i - y_i + \phi_i(1 - \mu_i)) - \Psi'(\mu_i \phi_i) - \Psi'(\phi_i(1 - \mu_i)) \right]$$

$$\frac{\partial \mu_i}{\partial \eta_{1i}} = \frac{1}{g'(\mu_i)}$$

$$\frac{\partial^2 \mu_i}{\partial \eta_{1i}^2} = \frac{-g''(\mu_i)}{[g'(\mu_i)]^2} \frac{\partial \mu_i}{\partial \eta_{1i}} = \frac{-g''(\mu_i)}{[g'(\mu_i)]^3}$$

$$\frac{\partial^2 l_i}{\partial \phi_i \partial \mu_i} = \left[\Psi(y_i + \mu_i \phi_i) - \Psi(m_i - y_i + \phi_i(1 - \mu_i)) - \Psi(\mu_i \phi_i) + \Psi(\phi_i(1 - \mu_i)) \right]$$
$$+ \mu_i \phi_i \left[\Psi'(y_i + \mu_i \phi_i) + \Psi'(m_i - y_i + \phi_i(1 - \mu_i)) - \Psi'(\mu_i \phi_i) - \Psi'(\phi_i(1 - \mu_i)) \right]$$

$$\frac{\partial l_i}{\partial \phi_i} = \mu_i \left[\Psi(y_i + \mu_i \phi_i) - \Psi(m_i - y_i + \phi_i(1 - \mu_i)) - \Psi(\mu_i \phi_i) + \Psi(\phi_i(1 - \mu_i)) \right]$$

$$\frac{\partial^2 l_i}{\partial \phi_i^2} = \mu_i^2 \left[\Psi'(y_i + \mu_i \phi_i) + \Psi'(m_i - y_i + \phi_i(1 - \mu_i)) - \Psi'(\mu_i \phi_i) - \Psi'(\phi_i(1 - \mu_i)) \right]$$

$$\frac{\partial \phi_i}{\partial \eta_{2i}} = \frac{1}{h'(\phi_i)}$$

$$\frac{\partial^2 \phi_i}{\partial \eta_{2i}^2} = \frac{-h''(\phi_i)}{[h'(\phi_i)]^3}$$

With these derivatives, the vector of the first derivatives **q** and the Hessian matrix **H** of equation (28) are well defined, and thus, the Newton-Rapson algorithm can be applied in order to obtain the maximum likelihood estimation of the model's regression parameters.

7.2. Fitting NB(μ, α) Regression Model

In the negative binomial regression model, if $Y_i \sim NB(\mu_i, \alpha_i)$, $i = 1, \ldots, n$; is a sample of the negative binomial distribution, the logarithm of the likelihood function is given by:

$$l_i(\theta) = \alpha_i \log \left(\frac{\alpha_i}{\mu_i} \right) + \log(\Gamma(y_i + \alpha_i)) - \log(y_i!) - \log(\Gamma(\alpha_i))$$
$$- (y_i + \alpha_i) \log \left(1 + \frac{\alpha_i}{\mu_i} \right),$$

where l_i is the i-th component of the likelihood function.

Thus, from equations (23) and (24), the first order derivatives of the logarithm of the likelihood function are given by:

$$\frac{\partial l_i}{\partial \beta_j} = \left[\frac{\mu_i - y_i}{\mu_i + \alpha_i}\right]\left(\frac{-\alpha_i}{\mu_i g'(\mu_i)}\right) x_{ji}$$

$$\frac{\partial l_i}{\partial \gamma_j} = [1 + \log(\alpha_i) + \Psi(y_i + \alpha_i) - \Psi(\alpha_i) - \left(\frac{y_i + \alpha_i}{\mu_i + \alpha_i}\right)$$

$$- \log(\mu_i + \alpha_i)]\left(\frac{1}{h'(\alpha_i)}\right) z_{ji}.$$

Finally, from equations (25), (26) and (27), the second order derivatives of the logarithm of the likelihood function are given by:

$$\frac{\partial l_i}{\partial \mu_i} = -\frac{\alpha_i}{\mu_i}\left(1 - \frac{y_i + \alpha_i}{\mu_i + \alpha_i}\right) = -\frac{\alpha_i(\mu_i - y_i)}{\mu_i(\mu_i + \alpha_i)}$$

$$\frac{\partial^2 l_i}{\partial \mu_i^2} = \frac{\alpha_i}{\mu_i^2}\left(1 - \frac{y_i + \alpha_i}{\mu_i + \alpha_i}\right) - \frac{\alpha_i}{\mu_i}\left(\frac{y_i + \alpha_i}{(\mu_i + \alpha_i)^2}\right)$$

$$= \frac{\alpha_i}{\mu_i(\mu_i + \alpha_i)}\left(\frac{\mu_i - y_i}{\mu_i} - \frac{y_i + \alpha_i}{\mu_i + \alpha_i}\right)$$

$$= \frac{\alpha_i(\mu_i^2 - 2y_i\mu_i - y_i\alpha_i)}{\mu_i^2(\mu_i + \alpha_i)^2}$$

$$\frac{\partial \mu_i}{\partial \eta_{1i}} = \frac{1}{g'(\mu_i)}$$

$$\frac{\partial^2 \mu_i}{\partial \eta_{1i}^2} = \frac{-g''(\mu_i)}{[g'(\mu_i)]^2}\frac{\partial \mu_i}{\partial \eta_{1i}} = \frac{-g''(\mu_i)}{[g'(\mu_i)]^3}$$

$$\frac{\partial^2 l_i}{\partial \alpha_i \partial \mu_i} = -\frac{1}{\mu_i}\left(1 - \frac{y_i + \alpha_i}{\mu_i + \alpha_i}\right) - \frac{\alpha_i}{\mu_i}\left(\frac{y_i + \alpha_i}{(\mu_i + \alpha_i)^2} - \frac{1}{\mu_i + \alpha_i}\right)$$

$$\frac{\partial l_i}{\partial \alpha_i} = 1 + \log(\alpha_i) + \Psi(y_i + \alpha_i) - \Psi(\alpha_i) - \left(\frac{y_i + \alpha_i}{\mu_i + \alpha_i}\right) - \log(\mu_i + \alpha_i)$$

$$= \frac{\mu_i - y_i}{\mu_i + \alpha_i} + \Psi(y_i + \alpha_i) - \Psi(\alpha_i) + \log\left(\frac{\alpha_i}{\mu_i + \alpha_i}\right)$$

$$\frac{\partial^2 l_i}{\partial \alpha_i^2} = \frac{1}{\alpha_i} + \Psi'(y_i + \alpha_i) - \Psi'(\alpha_i) - \left(\frac{1}{\mu_i + \alpha_i} - \frac{y_i + \alpha_i}{(\mu_i + \alpha_i)^2}\right) - \frac{1}{\mu_i + \alpha_i}$$

$$= \frac{\mu_i}{\alpha_i(\mu_i + \alpha_i)} + \Psi'(y_i + \alpha_i) - \Psi'(\alpha_i) - \frac{\mu_i - y_i}{(\mu_i + \alpha_i)^2}$$

$$\frac{\partial \alpha_i}{\partial \eta_{2i}} = \frac{1}{h'(\alpha_i)}$$

$$\frac{\partial^2 \alpha_i}{\partial \eta_{2i}^2} = \frac{-h''(\alpha_i)}{[h'(\alpha_i)]^3}.$$

With these derivatives, the Newton Rapson equation (28) can be proposed, and the maximum likelihood estimation of the regression parameters can be obtained using the Newton Rapson algorithm.

7.3. Fitting the NB(μ, σ^2) Regression Model

If $Y_i \sim NB(\mu_i, \sigma_i^2)$, $i = 1, \ldots, n$; is a sample of the variable of interest, then the logarithm of the likelihood function is given by:

$$l_i(\theta) = \frac{\mu_i^2}{\sigma_i^2 - \mu_i} \left(\log(\mu_i) - \log(\sigma_i^2)\right) + \log\left(\Gamma\left(y_i + \frac{\mu_i^2}{\sigma_i^2 - \mu_i}\right)\right) - \log(y_i!)$$
$$- \log\left(\Gamma\left(\frac{\mu_i^2}{\sigma_i^2 - \mu_i}\right)\right) - y_i \left(\log(\sigma_i^2) - \log(\sigma_i^2 - \mu_i)\right)$$

The first derivatives are:

$$\frac{\partial l_i}{\partial \beta_j} = \frac{\partial l_i}{\partial \mu_i} \left(\frac{1}{g'(\mu_i)}\right) x_{ji}$$

$$\frac{\partial l_i}{\partial \gamma_j} = \frac{\partial l_i}{\partial \sigma_i^2} \left(\frac{1}{h'(\mu_i)}\right) z_{ji},$$

where:

$$\frac{\partial l_i}{\partial \mu_i} = \frac{\mu_i}{\sigma_i^2 - \mu_i} \left[\frac{2\sigma_i^2 - \mu_i}{\sigma_i^2 - \mu_i} \left(\log(\mu_i) - \log(\sigma_i^2) + \frac{\Psi\left(y_i + \frac{\mu_i^2}{\sigma_i^2 - \mu_i}\right)}{\Gamma\left(y_i + \frac{\mu_i^2}{\sigma_i^2 - \mu_i}\right)} - \frac{\Psi\left(\frac{\mu_i^2}{\sigma_i^2 - \mu_i}\right)}{\Gamma\left(\frac{\mu_i^2}{\sigma_i^2 - \mu_i}\right)}\right) - \frac{y_i}{\mu_i} + 1\right]$$

$$\frac{\partial l_i}{\partial \sigma_i^2} = \frac{-\mu_i^2}{(\sigma_i^2 - \mu_i)^2} \left[\log(\mu_i) - \log(\sigma_i^2) + \frac{\sigma_i^2 - \mu_i}{\sigma_i^2} + \frac{\Psi\left(y_i + \frac{\mu_i^2}{\sigma_i^2 - \mu_i}\right) - 1}{\Gamma\left(y_i + \frac{\mu_i^2}{\sigma_i^2 - \mu_i}\right)} - \frac{y_i(\sigma_i^2 - \mu_i)}{\mu_i \sigma_i^2}\right].$$

The second derivatives can be obtained from equations (25), (26) and (27) as in sections 7.1, and 7.2, and the maximum likelihood estimation of the regression parameters can be obtained by applying the Newton Rapson algorithm.

8. School Absenteeism Data

The data analyzed in this paper are presented in (Quine 1975) and come from a sociological study of Australian Aboriginal and white children from Walgett, New South Wales. There are nearly equal numbers of the two sexes and an equal number 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 year; children who had suffered a serious illness during the year were excluded from this analysis.

The possible values for each observed variable are:

- **Age:** age group: Primary (0), First form (1), Second form (2), and Third form (3).

- **Gender:** sex: factor with levels Female or Male. The number (0) represents "female" and (1) corresponds to "male".

- **Cultural Background:** ethnic background: Aboriginal or Not: number 0 for Aboriginal and 1 for White.

- **Learning Ability:** learner status: factor with levels Average (1) or Slow learner (0).

- **Days Absent:** days absent from school in the year.

Since the variable "days absent" counts the number of events that occurred during a year, this variable can be modeled by a model associated with the Poisson distribution. Thus, if X is the days of absenteeism of a student during a year, we assume the negative binomial model $NB(\mu, \alpha)$, where the mean and the shape parameters follow linear regression structures given by:

$$\log(\mu_i) = \eta_{1i} = \mathbf{x}_i'\boldsymbol{\beta} \tag{29}$$

$$\log(\alpha_i) = \eta_{2i} = \mathbf{z}_i'\boldsymbol{\gamma}, \tag{30}$$

where $\boldsymbol{\beta} = (\beta_0, \beta_1, \beta_2, \beta_3)$, $\mathbf{x}_i' = (intercept_i, age_i, cultural.back_i, learning.ability_i)$, $\boldsymbol{\gamma} = (\gamma_0, \gamma_1)$ and $\mathbf{z}_i' = (intercept_i, cultural.back_i)$, $i = 1, \ldots, 113$.

We fitted this model to the data by applying two Bayesian methods, one using Open Bugs and the other using R software. In the second, the algorithm was built based on the Metropolis Hastings algorithm, defined in Section 6. The posterior parameter estimates, standard deviations and credibility intervals are given in Table 1.

TABLE 1: Parameter estimates of the beta-binomial regression model.

Parameter	Description	OpenBugs			RStudio		
		Estimate	S.D.	Cred. Interval	Estimate	S.D.	Cred. Interval
β_0	Intercept	2.823	0.303	(2.24,3.42)	2.7844	0.3	(2.22,3.39)
β_1	Age	0.1734	0.115	$(-0.054,0.4)$	0.1866	0.113	$(-0.04,0.41)$
β_2	Cultur.backg.	-0.6564	0.1884	$(-1.024,-0.2807)$	-0.6269	0.1886	$(-1.012,-0.268)$
β_3	Learn.ability	-0.3632	0.1762	$(-0.711,-0.021)$	-0.3613	0.1805	$(-0.706,-0.004)$
γ_0	Intercept	0.593	0.2123	(0.171,0.999)	0.5643	0.1712	(0.205,0.863)
γ_1	Cultur.backg.	-0.7276	0.3048	$(-1.32,-0.128)$	-0.74	0.2496	$(-1.21,-0.22)$

We calculated the Bayesian information criterion (BIC) for the model in both programs: the model in OpenBugs obtained a BIC equal to 826.97 while in R software the BIC was 827.24. The parameter estimates, their standard deviations, and the 95% credible intervals obtained from the two programs are given in Table 1. From this table, we can observe that Age can be removed from the mean structure of the model, since zero (0) belongs to the 95% credible interval.

Assuming, a mean structure (29) without the age variable, and if the shape structure given by (30), the parameter estimates, standard deviations, and 95% credible intervals are reported in Table 2.

Table 2: Parameter estimates of the beta-binomial regression model.

Parameter	Description	OpenBugs			RStudio		
		Estimate	S.D.	Cred. Interval	Estimate	S.D.	Cred. Interval
β_0	Intercept	3.235	0.136	(2.975,3.506)	3.1571	0.137	(2.89,3.43)
β_2	Cultural.backg.	-0.7253	0.1828	$(-1.083,-0.3647)$	-0.6644	0.181	$(-1.02,-0.31)$
β_3	Learning.ability	-0.3975	-0.3975	$(-0.741,-0.0432)$	-0.34	0.18	$(-0.68,-0.02)$
γ_0	Intercept	0.506	0.2014	(0.107,0.892)	0.43	0.204	(0.05,0.68)
γ_1	Cultural.backg.	-0.575	0.2895	$(-1.14,-0.003)$	-0.47	0.2524	$(-0.89,0.02)$

Again we calculated the BIC with both programs: R software yielded BIC equal to 829.83 while the BIC for the OpenBugs estimation was 833.92.

Therefore, from the parameter estimates that are given in Table 2, it is possible to conclude that the variable *Days.abseent* decreases while the *Cultural.background* or the *Learning.ability* increases. For Slow learner status (0), the mean of Days Absent is 2.4927 for Aboriginal background and 3.1571 for White cultural background. For average learner status (1), the mean of Days Absent is 8.61 for White cultural background and 16.73 for Aboriginal cultural background.

Conclusions

In this paper, new parameterizations of the beta-binomial and negative binomial distributions have been proposed. These formulations, in terms of the mean and shape parameters, are good options to explain the behavior of overdispersed count data. From these reparameterizations, overdispersed regression models have been defined assuming that the two parameters of these models follow regression structures.

We also propose Bayesian and classic methods to fit these models, which we apply to analyze school absenteeism data. In this application, the proposed methods show suitable performance and posterior inferences; those obtained by applying the Bayesian method in RStudio and in OpenBugs agree. The observed results showed that the mean depends on the cultural background and the learning ability, and the shape parameter depends on the cultural background.

References

Breslow, N. (1984), 'Extra-Poisson variation in Log-Linear models', *Journal of Applied Statistics* **31**, 38–44.

Cepeda-Cuervo, E. (2001), Modelagem da Variabilidade em Modelos Lineares Generalizados, Unpublished Math Ph.D. thesis, Mathematics Institute, Universidade Federal do Río de Janeiro.

Cepeda-Cuervo, E. & Achcar, J. (2009), 'Modelos de regresión heterocedásticos usando aproximación bayesiana', *Revista Colombiana de Estadística* **32**(2), 267–287.

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

Cepeda-Cuervo, E., Migon, H., Garrido, L. & Achcar, J. (2014), 'Generalized linear models with random effects in the two-parameter exponential family', *Journal of Statistical Computation and Simulation* **84**(3), 513–525.

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

Cox, D. (1983), 'Some remarks on overdispersion', *Biometrika* **70**(1), 269–274.

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

Demétrio, C., Kokonendji, C. & Zocchi, S. (2007), 'On Hinde-Demétrio regression models for overdispersed count data', *Statistical Methodology* **4**, 277–291.

Ferrari, S. & Cribari-Neto, F. (2004), 'Beta Regression for Modelling Rates and Proportions', *Journal of Applied Statistics* (31-7), 799–815.

Jørgensen, B. (1997), *The Theory of Dispersion Models*, Chapman & Hall, London.

Lawless, J. (1987), 'Negative binomial regression model', *Canadian Journal of Statistics* **15**(3), 209–225.

Margolin, B., Kaplan, N. & Zeiger, E. (1981), 'Statistical analysis of the Ames Salmonella/ microsome test', *Proceedings of the National Academy of Sciences* **76**, 3779–3783.

McCullagh, P. & Nelder, J. (1989), *Generalized Linear Models*, Chapman Hall, London.

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

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

Williams, D. (1975), 'The analysis of binary responses from toxicological experiments involving reproduction and teratogenicity', *Biometrics* **31**(4), 949–952.

Williams, D. (1982), 'Extra-binomial Variation in Logistic linear Models', *Journal of Applied Statistics* **31**, 144–148.

Appendix

In the classic approach, there were some calculations related to the first and second derivatives of the different likelihood functions that were dropped. In this section, some of them appear.

- Some details of the $BB(\mu, \phi)$ and the negative binomial maximum likelihood functions' derivatives that were used in the Newton Rapson algorithm (Section 7.1):

$$
\begin{aligned}
\frac{\partial l_i}{\partial \beta_j} &= \left[\frac{\frac{\partial B(y_i + \mu_i \phi_i, m_i - y_i + \phi_i(1 - \mu_i))}{\partial x_1} \phi_i - \frac{\partial B(y_i + \mu_i \phi_i, m_i - y_i + \phi_i(1 - \mu_i))}{\partial x_2} \phi_i}{B(y_i + \mu_i \phi_i, m_i - y_i + \phi_i(1 - \mu_i))} \right] \left(\frac{1}{g'(\mu_i)} \right) x_{ji} \\
&\quad - \left[\frac{\frac{\partial B(\mu_i \phi_i, \phi_i(1 - \mu_i))}{\partial x_1} \phi_i - \frac{\partial B(\mu_i \phi_i, \phi_i(1 - \mu_i))}{\partial x_2} \phi_i}{B(\mu_i \phi_i, \phi_i(1 - \mu_i))} \right] \left(\frac{1}{g'(\mu_i)} \right) x_{ji} \\
&= [\Psi(y_i + \mu_i \phi_i) - \Psi(m_i + \phi_i) - \Psi(m_i - y_i + \phi_i(1 - \mu_i)) + \Psi(m_i + \phi_i) - \Psi(\mu_i \phi_i) \\
&\quad + \Psi(\phi_i) + \Psi(\phi_i(1 - \mu_i)) - \Psi(\phi_i)] \left(\frac{\phi_i}{g'(\mu_i)} \right) x_{ji} \\
&= [\Psi(y_i + \mu_i \phi_i) - \Psi(m_i - y_i + \phi_i(1 - \mu_i)) - \Psi(\mu_i \phi_i) + \Psi(\phi_i(1 - \mu_i))] \left(\frac{\phi_i}{g'(\mu_i)} \right) x_{ji}
\end{aligned}
$$

$$
\frac{\partial l_i}{\partial \gamma_j} = [\Psi(y_i + \mu_i \phi_i) - \Psi(m_i - y_i + \phi_i(1 - \mu_i)) - \Psi(\mu_i \phi_i) + \Psi(\phi_i(1 - \mu_i))] \left(\frac{\mu_i}{h'(\phi_i)} \right) z_{ji}
$$

- In Section (7.2), some details of the derivatives of the logarithm of the $NB(\mu, \alpha)$ maximum likelihood function were dropped. These are:

$$
\begin{aligned}
\frac{\partial l_i}{\partial \beta_j} &= \left[\frac{-\alpha_i}{\mu_i} - (y_i + \alpha_i) \left(\frac{\mu_i}{\mu_i + \alpha_i} \right) \left(\frac{-\alpha_i}{\mu_i^2} \right) \right] \left(\frac{1}{g'(\mu_i)} \right) x_{ji} \\
&= \left[\frac{1}{\mu_i} - \left(\frac{y_i + \alpha_i}{\mu_i(\mu_i + \alpha_i)} \right) \right] \left(\frac{-\alpha_i}{g'(\mu_i)} \right) x_{ji}
\end{aligned}
$$

$$
\begin{aligned}
\frac{\partial l_i}{\partial \gamma_j} &= \left[1 + \log(\alpha_i) - \log(\mu_i) + \Psi(y_i + \alpha_i) - \Psi(\alpha_i) - \left(\frac{y_i + \alpha_i}{\mu_i + \alpha_i} \right) \right. \\
&\quad \left. - \log \left(1 + \frac{\alpha_i}{\mu_i} \right) \right] \left(\frac{1}{h'(\alpha_i)} \right) z_{ji} \\
&= \left[1 + \log(\alpha_i) + \Psi(y_i + \alpha_i) - \Psi(\alpha_i) - \left(\frac{y_i + \alpha_i}{\mu_i + \alpha_i} \right) - \log(\mu_i + \alpha_i) \right] \\
&\quad \left(\frac{1}{h'(\alpha_i)} \right) z_{ji}
\end{aligned}
$$

- Finally, regarding Section (7.3), we find the following first derivatives for the $(\mu, \sigma^2) - NB$ log-likelihood function:

$$
\frac{\partial l_i}{\partial \mu_i} = \left[\frac{\mu_i^2}{(\sigma_i^2 - \mu_i)^2} + \frac{2\mu_i}{\sigma_i^2 - \mu_i} \right] [\log(\mu_i) - \log(\sigma_i^2)] + \frac{\mu_i^2}{\sigma_i^2 - \mu_i} \left(\frac{1}{\mu_i} \right)
$$

$$
+ \frac{\Psi\left(y_i + \frac{\mu_i^2}{\sigma_i^2 - \mu_i} \right) \left[\frac{\mu_i^2}{(\sigma_i^2 - \mu_i)^2} + \frac{2\mu_i}{\sigma_i^2 - \mu_i} \right]}{\Gamma\left(y_i + \frac{\mu_i^2}{\sigma_i^2 - \mu_i} \right)} - \frac{\frac{\mu_i^2}{(\sigma_i^2 - \mu_i)^2} + \frac{2\mu_i}{\sigma_i^2 - \mu_i}}{\Gamma\left(y_i + \frac{\mu_i^2}{\sigma_i^2 - \mu_i} \right)} - y_i \left(\frac{1}{\sigma_i^2 - \mu_i} \right)
$$

$$
= \frac{\mu_i}{\sigma_i^2 - \mu_i} \{ \left(\frac{\mu_i}{\sigma_i^2 - \mu_i} + 2 \right) [\log(\mu_i) - \log(\sigma_i^2)] + 1 + \frac{\Psi\left(y_i + \frac{\mu_i^2}{\sigma_i^2 - \mu_i} \right) \left[\frac{\mu_i}{(\sigma_i^2 - \mu_i)} + 2 \right]}{\Gamma\left(y_i + \frac{\mu_i^2}{\sigma_i^2 - \mu_i} \right)}
$$

$$
- \frac{\frac{\mu_i}{\sigma_i^2 - \mu_i} + 2}{\Gamma\left(y_i + \frac{\mu_i^2}{\sigma_i^2 - \mu_i} \right)} - \frac{y_i}{\mu_i} \}
$$

$$
= \frac{\mu_i}{\sigma_i^2 - \mu_i} \left[\frac{2\sigma_i^2 - \mu_i}{\sigma_i^2 - \mu_i} \left(\log(\mu_i) - \log(\sigma_i^2) + \frac{\Psi\left(y_i + \frac{\mu_i^2}{\sigma_i^2 - \mu_i} \right) - 1}{\Gamma\left(y_i + \frac{\mu_i^2}{\sigma_i^2 - \mu_i} \right)} \right) - \frac{y_i}{\mu_i} + 1 \right]
$$

$$
\frac{\partial l_i}{\partial \sigma_i^2} = \frac{-\mu_i^2}{(\sigma_i^2 - \mu_i)^2} [\log(\mu_i) - \log(\sigma_i^2)] + \frac{\mu_i^2}{\sigma_i^2 - \mu_i} \left(\frac{-1}{\sigma_i^2} \right) + \frac{\Psi\left(y_i + \frac{\mu_i^2}{\sigma_i^2 - \mu_i} \right) \left[\frac{-\mu_i^2}{(\sigma_i^2 - \mu_i)^2} \right]}{\Gamma\left(y_i + \frac{\mu_i^2}{\sigma_i^2 - \mu_i} \right)}
$$

$$
- \frac{\frac{-\mu_i^2}{(\sigma_i^2 - \mu_i)^2}}{\Gamma\left(y_i + \frac{\mu_i^2}{\sigma_i^2 - \mu_i} \right)} - y_i \left(\frac{1}{\sigma_i^2} - \frac{1}{\sigma_i^2 - \mu_i} \right)
$$

$$
= \frac{-\mu_i^2}{(\sigma_i^2 - \mu_i)^2} \left[\log(\mu_i) - \log(\sigma_i^2) + \frac{\sigma_i^2 - \mu_i}{\sigma_i^2} + \frac{\Psi\left(y_i + \frac{\mu_i^2}{\sigma_i^2 - \mu_i} \right)}{\Gamma\left(y_i + \frac{\mu_i^2}{\sigma_i^2 - \mu_i} \right)} - \frac{1}{\Gamma\left(y_i + \frac{\mu_i^2}{\sigma_i^2 - \mu_i} \right)} \right.
$$

$$
\left. - \frac{y_i(-\mu_i)}{\mu_i \sigma_i^2} \right]
$$

$$
= \frac{-\mu_i^2}{(\sigma_i^2 - \mu_i)^2} \left[\log(\mu_i) - \log(\sigma_i^2) + \frac{\sigma_i^2 - \mu_i}{\sigma_i^2} + \frac{\Psi\left(y_i + \frac{\mu_i^2}{\sigma_i^2 - \mu_i} \right) - 1}{\Gamma\left(y_i + \frac{\mu_i^2}{\sigma_i^2 - \mu_i} \right)} + \frac{y_i(\mu_i)}{\mu_i \sigma_i^2} \right]
$$