Bayesian Inference for the Negative Binomial-Sushila Linear Model

Darika Yamrubboon[1*], Ampai Thongteeraparp[1], Winai Bodhisuwan[1***], Katechan Jampachaisri[2****], and Andrei Volodin[3*****]**

(Submitted by A. M. Elizarov)

[1]*Department of Statistics, Kasetsart University, Bangkok, 10903 Thailand*
[2]*Department of Mathematics, Naresuan University, Phitsanulok, 65000 Thailand*
[3]*Department of Mathematics and Statistics, University of Regina, Regina, SK. S4S 0A2 Canada*
Received August 14, 2018

Abstract—The aim of this article is to develop a new linear model for count data. The main idea is in an application of a new generalized linear model framework, which we call the Negative Binomial-Sushila linear model. The Negative Binomial-Sushila distribution has been proposed recently and applied to count data. This distribution is constructed as a mixture of the Negative Binomial and Sushila distributions. The mixed distribution is a flexible alternative to the Poisson distribution when over-dispersed count data is analyzed. The parameters of this distribution are estimated using a Bayesian approach with R2jags package of the R language. The Negative Binomial-Sushila linear model is applied to fit two real data sets with an over-dispersion and its performance is compared with the performance of some traditional models. The results show that the Negative Binomial-Sushila generalized linear model fits the data sets better than the traditional generalized models for these data sets.

1. INTRODUCTION

The Poisson distribution is the most popular model for count data analysis. However, the Poisson distribution has one of the essential property: the equality of its mean and variance. If such a characteristic is not conformed by the data we are planning to model, then the test statistics derived based on the Poisson distribution are not suitable [1]. Obviously, the difference between the variance and the mean occurs when the variance of data is significantly larger or smaller than its expectation. This phenomenon is called *over-dispersion* or *under-dispersion*, respectively. In practical applications, the problem of over-dispersion has commonly occurred, thus, many researchers attempt to resolve the problem by using a mixture of some distributions, for example, to mix the Poisson and Negative Binomial. Some recent studies about mixed distributions for modeling over-dispersed data are the Poisson-Gamma or Negative Binomial (NB) [2], Poisson-Lognormal (PL) [3], Negative Binomial-Inverse Gaussian (NB-IG) [4], Negative Binomial-Lindley (NB-L) [5] and Negative Binomial-Generalized Exponential (NB-GE) [6] distributions.

The statistical analysis of count data in the discrete univariate parametric distribution framework has a long and rich history, see, for example [7]. One of the most commonly used statistical analyses tools is a regression model and obviously, for count data, it has been considered in literature. In regression

*[*]E-mail: `darika.y@ku.th`
[**]E-mail: `fsciamu@ku.ac.th`
[***]E-mail: `fsciwnb@ku.ac.th`
[****]E-mail: `katechanj@nu.ac.th`
[*****]E-mail: `andrei.volodin@uregina.ca`

analysis of event counts, the response (dependent) variable of interest is a non-negative integer (*count data*), which is explained or analyzed in terms of a set of covariates (independent variables). This linear regression model has been applied to count data and is known as the Poisson regression, NB regression, mixed Poisson regression, and mixed NB regression in literature. See, for example [8]. The linear regression model described above is also called generalized linear model (GLM).

The GLM generalizes ordinary regression models into two ways: It allows the response variable to have a specific distribution other than the normal one, and it allows to model some function of the mean. Both advantages of GLM are important for categorical data [9]. The GLMs have become very popular because of their generality in a wide variety of applications and its flexibility for a distribution of the response variable or link function. In addition, the GML theory, as a part of the Regression Analysis, is one of the most important parts of the contemporary Statistical Inference and is of crucial importance for an Analysis of Variance, and Categorical Data modeling [10].

Recently, several mixed distributions have been proposed to describe the behavior of count response variable with some explanatory variables in a linear form. Some of these include the mixed NB and mixed Poisson regression [11], the Generalized Waring regression [12], mixed Poisson-Inverse Gaussian regression, Poisson-Weighted Exponential regression [13], and Generalized Poisson-Lindley linear model [14]. These works are based on the GLM with the application of the frequentist approach, namely with the maximum likelihood estimation of the model parameters. Moreover, some the new mixed NB models are based on the Bayesian estimation of parameters. For example, NB-L generalized linear model was presented in [15] and NB-GE generalized linear model were proposed in [16]. The main difference between the frequentist statistical theory and the Bayesian approach is that the latter considers parameters as random variables that are characterized by a prior distribution. The advantages of the Bayesian approach are well known and include a choice of prior beliefs, avoidance of asymptotic approximation and practical estimation of function of parameters [9, 10].

In this article we suggest a new mixed NB model based on so-called the Negative Binomial-Sushila (NB-S) distribution. The NB-S distribution has been recently introduced in [17]. It is a new mixed NB distribution obtained by mixing the Negative Binomial with the Sushila distributions [19]. In some cases, the NB-S distribution can be regarded as an alternative model for count data, especially for the over-dispersion phenomena. Since, the NB-S distribution has not been considered in the regression model, we are interested in extending the NB-S distribution to GLM in order to obtain the NB-S linear model.

The objective of this research is to propose the NB-S distribution to the GLM for counts. Furthermore, the model parameters are estimated using a Bayesian approach. Alternatively, Markov Chain Monte Carlo (MCMC) technique can be implemented for the model parameter estimation by using Bayesian software, such as JAGS. Moreover, the comparison of GLMs based on Poisson, NB, and NB-S distributions on two real data sets is presented. The NB-S GLM is an alternative model to study the relationship between a count response and a set of covariates.

The content of the aritcle is as follows. In Section 2, we describe the Negative Binomial-Sushila distribution. The derivation of NB-S linear model is discussed in Section 3. The Bayesian implementation and criteria for model comparison are presented in Section 4. The applications of Poisson, NB and NB-S models for two real data is shown in Section 5. Afterward, the summary of model performance is discussed in Section 6, and some of conclusions are drawn in Section 7.

2. THE NEGATIVE BINOMIAL-SUSHILA DISTRIBUTION

The Negative Binomial-Sushila distribution is defined as a mixture of the Negative Binomial and Sushila distributions. It was first proposed and presented in [17], where also some its basis properties including special cases and the factorial moments are discussed. The Negative Binomial-Sushila distribution is one of the mixed Negative Binomial distributions which provides a flexible model for over-dispersed count data analysis. Before discussing the Negative Binomial-Sushila distribution, we will refer to some characteristics of the Negative Binomial and Sushila distributions themselves.

Initially, the Negative Binomial statistical model can be derived as a mixture of the Poisson and Gamma distributions or as a series of Bernoulli trails. The probability mass function (pmf) of the NB

distribution for the number of observed y failures before the rth success in a sequence of Bernoulli trails is expressed as follows:

$$P\{Y = y_i\} = g(Y = y_i; p, r) = \binom{r + y_i - 1}{y_i} p^r (1 - p)^{y_i},$$

where $y_i = 0, 1, 2, \ldots, 0 < p < 1, r > 0$, and $i = 1, \ldots, n$. Now consider a sequence of random variables $\{Y_i, 1 \leq i \leq n\}$ with the NB distribution with possibly different parameters p_i. That is,

$$P\{Y_i = y_i\} = g(Y = y_i; p_i, r) = \binom{r + y_i - 1}{y_i} p_i^r (1 - p_i)^{y_i}, \tag{1}$$

where $y_i = 0, 1, 2, \ldots, 0 < p_i < 1, r > 0$, and $i = 1, \ldots, n$.

Now, we re-parameterize parameters p_i and r as $p_i = \dfrac{\phi}{\mu_i + \phi}$ and $r = \phi$. Then we obtain the most common implementation of the NB distribution as the NB2 model [24]. It has a pmf

$$g(Y_i = y_i; \mu_i, \phi) = \frac{\Gamma(\phi + y_i)}{\Gamma(\phi)\Gamma(y_i + 1)} \left(\frac{\phi}{\mu_i + \phi}\right)^\phi \left(\frac{\mu_i}{\mu_i + \phi}\right)^{y_i}, \tag{2}$$

where $y_i = 0, 1, 2, \ldots, \mu_i \geq 0, \phi > 0$ and $\Gamma(t) = \int\limits_0^\infty y^{t-1} e^{-y} dy, \ t > 0$ is the complete gamma function.

Note that sometimes parameter ϕ is considered as the inverse of the dispersion parameter ω, that is, $\omega = 1/\phi$. It is simple to derive from (2), the mean and variance are given by $E(Y_i) = \mu_i$ and $Var(Y_i) = \mu_i + \mu_i^2/\phi$. Let Y_i be the NB random variable with parameter μ_i and inverse dispersion parameter ϕ_i, written as $Y_i \sim NB(\mu_i, \phi)$.

In 2013, the Sushila distribution as a mixture of the Exponential$\left(\frac{\theta}{\alpha}\right)$ and Gamma$\left(2, \frac{\theta}{\alpha}\right)$ distributions and ϕ is the inverse of the dispersion parameter ω, that is $\omega = \phi^{-1}$ with $\phi > 0$ is introduced in [19] and some of its properties are discussed. The Sushila distribution has two parameters α and θ and is denoted as Sushila(α, θ). It can be seen that the Lindley distribution is a special case of this distribution when $\alpha = 1$. The probability density function (pdf) of the Sushila distribution is expressed as

$$f(Z_i = z_i; \alpha, \theta) = \frac{\theta^2}{\alpha(\theta + 1)} \left(1 + \frac{z_i}{\alpha}\right) \exp\left(-\frac{\theta}{\alpha} z_i\right), \tag{3}$$

where $z_i > 0, \ \theta > 0$ and $\alpha > 0$. If a random variable Z has the Sushila(α, θ) distribution, then its moment generating function is given by (see [17])

$$M_Z(t) = \frac{\theta^2}{(\theta + 1)} \frac{(\theta - \alpha t + 1)}{(\theta - \alpha t)^2},$$

where $t < (\theta + 1)/\alpha$; its mean and variance of Sushila distribution (see [19]) are

$$E(Z) = \frac{\alpha(\theta + 2)}{\theta(\theta + 1)} \quad \text{and} \quad Var(Z) = \frac{\alpha^2(\theta^2 + 4\theta + 2)}{\theta^2(\theta + 1)^2}. \tag{4}$$

Article [17] has recently proposed a compound NB distribution by re-parametrization of p_i in (1) as $p_i = e^{-\lambda_i}$, where λ_i have Sushila distribution. This distribution is called *Negative Binomial-Sushila* (NB-S) *distribution* and its pmf is obtained in the following form:

$$P\{Y_i = y_i\} = p(Y_i = y_i; r, \alpha, \theta) = \frac{\theta^2}{\theta + 1} \binom{r + y_i - 1}{y_i} \sum_{j=0}^{y_i} \binom{y_i}{j} (-1)^j \frac{\theta + \alpha(r + j) + 1}{(\theta + \alpha(r + j))^2},$$

where $y_i = 0, 1, 2, \ldots$, for $r > 0, \alpha > 0$, and $\theta > 0$. Some characteristics of the NB-S distribution were derived in [17] and, for example, its mean and variance are

$$E(Y_i) = \frac{r}{\delta_1}(\delta_2 - \delta_1), \quad Var(Y_i) = \frac{r^2\delta_3}{\delta_1} + \frac{r\delta_3}{\delta_1} - \frac{r^2\delta_2^2}{\delta_1^2} - \frac{2r\delta_2}{\delta_1} + r,$$

where $\delta_1 = \dfrac{\theta + 1}{\theta^2}, \delta_2 = \dfrac{-\alpha + \theta + 1}{(\alpha - \theta)^2}$ and $\delta_3 = \dfrac{-2\alpha + \theta + 1}{(\theta - 2\alpha)^2}, \theta \neq k\alpha; k = 1, 2$. See [17] for more details.

3. THE NEGATIVE BINOMIAL-SUSHILA LINEAR MODEL

In the previous section we discussed the NB-S distribution which has been recently introduced to analyze count data, especially in the presence of the over-dispersion. In some cases, the NB-S distribution can be more flexible than Poisson and NB distributions [17]. In this section, we extend the NB-S distribution to GLM to obtain the NB-S linear model for count data.

The GLM is a generalization and extension of the linear regression model and it has been widely discussed in the literature. It is a tool for establishing the relationship between the count variables and covariates. We consider the NB-S distribution as link function to connect linear combination of explanatory variables to the response variable and then specify a function of expected value of the response variable. Hence, the mean of count response, $\mu_i = E(Y_i)$, is actually positive. Therefore, the logarithm of the mean can be considered as the link function, $g(\mu_i) = \log \mu_i$, that relates μ_i to the predictors as a linear function (see [18]): $g(\mu_i) = \log \mu_i = \beta_1 + \beta_2 x_1 + \cdots + \beta_{k+1} x_k$. We can re-express the above relation as $\mu_i = \exp(\mathbf{x}_i^T \boldsymbol{\beta})$, where $\mathbf{x}_i^T$ is the ith row of a $n \times (k+1)$ design matrix $\mathbf{X}$, and $\boldsymbol{\beta} = [\beta_1, \ldots, \beta_{k+1}]^T$ is a $(k+1) \times 1$ vector of unknown regression coefficients to be estimated.

The aforementioned framework of GLM can be applied for deriving the NB-S linear model. The mixture of NB and Sushila distributions is given as

$$p(Y_i = y_i; \mu_i, \phi, \alpha, \theta) = \int\limits_0^\infty \mathrm{NB}(y; z\mu_i, \phi)\mathrm{Sushila}(z; \alpha, \theta)dz.$$

The mean response μ_i is similar to the one described in (2) and z follows the Sushila distribution. Based on (2), and (3), the pmf of the NB-S distribution can be re-written as

$$p(Y_i = y_i; \mu_i, \phi, \alpha, \theta) = \frac{\Gamma(\phi + y_i)}{\Gamma(\phi)\Gamma(y_i + 1)} \frac{\theta^2}{\alpha(\theta + 1)} \int\limits_0^\infty \left(\frac{\phi}{z\mu_i + \phi}\right)^\phi \left(\frac{z\mu_i}{z\mu_i + \phi}\right)^{y_i} \left(1 + \frac{z}{\alpha}\right) e^{-\frac{\theta}{\alpha}z}dz,$$

where $y_i = 0, 1, 2, \ldots, \mu_i \geq 0$ and $\phi > 0$ is inverse of the dispersion parameter.

The response variable Y_i is assumed to follow the NB-S distribution, represented as $Y_i \sim NB - S(\mu_i, \phi, \alpha, \theta)$. The mean of response can then be calculated using the conditional expectation as follows:

$$E(Y_i) = \mu_i \times E(z) = \mu_i \times \frac{\alpha(\theta + 2)}{\theta(\theta + 1)},$$

where $E(z)$ is the mean of the Sushila distribution given in (4). For the NB-S GLM, let Y_i be a response variable and $\mathbf{x}_i^T$ be the set of covariates. The conditional distribution of Y_i given $\mathbf{x}_i^T$ is assumed to follow the NB-S$(\mu_i, \phi, \alpha, \theta)$ distribution, express as $Y_i|\mathbf{x}_i^T \sim \mathrm{NB}\text{-}\mathrm{S}(\mu_i, \phi, \alpha, \theta)$, $\mu_i = e^{\mathbf{x}_i^T \boldsymbol{\beta}}$. The pmf of $Y_i|\mathbf{x}_i^T$ can be written in the linear model form with a log-link function as

$$f(y_i|\mathbf{x}_i^T) = \frac{\Gamma(\phi + y_i)}{\Gamma(\phi)\Gamma(y_i + 1)} \frac{\theta^2}{\alpha(\theta + 1)} \int\limits_0^\infty \left(\frac{\phi}{ze^{\mathbf{x}_i^T \boldsymbol{\beta}} + \phi}\right)^\phi \left(\frac{ze^{\mathbf{x}_i^T \boldsymbol{\beta}}}{ze^{\mathbf{x}_i^T \boldsymbol{\beta}} + \phi}\right)^{y_i} \left(1 + \frac{z}{\alpha}\right) e^{-\frac{\theta}{\alpha}z}dz. \quad (5)$$

Therefore, the mean response of observation Y_i is dependent on vector of independent variables $\mathbf{x}_i^T$ and can be expressed in the form of exponential link function

$$E(Y_i|\mathbf{x}_i^T) = \mu_i \times \frac{\alpha(\theta + 2)}{\theta(\theta + 1)} = e^{\mathbf{x}_i^T \boldsymbol{\beta}} \times \frac{\alpha(\theta + 2)}{\theta(\theta + 1)}, \quad i = 1, \ldots, n. \quad (6)$$

Let $\mathbf{y} = [y_1, \ldots, y_n]^T$ be the $n \times 1$ vector of response, and $\boldsymbol{\gamma} = [\phi, \alpha, \theta, \boldsymbol{\beta}^T]^T$ be the vector of parameters. Therefore, the likelihood function of $\boldsymbol{\gamma}$ associated with $(\mathbf{y}, \mathbf{X})$ is given by

$$L(\boldsymbol{\gamma}|\mathbf{y}, \mathbf{X}) = \prod_{i=1}^n f(y_i|\mathbf{x}_i^T, \boldsymbol{\gamma}) = \prod_{i=1}^n \frac{\Gamma(\phi + y_i)}{\Gamma(\phi)\Gamma(y_i + 1)} \frac{\theta^2}{\alpha(\theta + 1)}$$

$$\times \int_0^\infty \left(\frac{\phi}{z e^{\mathbf{x}_i^T \boldsymbol{\beta}} + \phi} \right)^\phi \left(\frac{z e^{\mathbf{x}_i^T \boldsymbol{\beta}}}{z e^{\mathbf{x}_i^T \boldsymbol{\beta}} + \phi} \right)^{y_i} \left(1 + \frac{z}{\alpha} \right) e^{-\frac{\theta}{\alpha} z} dz. \tag{7}$$

The vector of unknown parameters $\boldsymbol{\gamma}$ can be customarily estimated using the maximum likelihood method. Alternatively, all parameters can also be estimated using Bayesian approach, which allows the consideration of prior information into estimation. The practical advantages of the Bayesian approach are its flexibility and generality as this allows it to cope with complex problems [20].

4. BAYESIAN APPROACH

In this part, we implement the Bayesian approach using MCMC technique for NB-S linear model, which involves in hierarchical model of NB-S, prior distribution, joint posterior density, and MCMC algorithm. The details are provided as follows.

4.1. Hierarchical Model of NB-S

As shown in previous section, the likelihood function for NB-S model does not have a closed form. The parametrization based on (5) can be determined by using the hierarchical representation to implicit both in the integrand and in the definition of the Sushila distribution itself. Thus, the NB-S distribution is conditional upon the unobserved site-specific frailty term z, that explains additional heterogeneity, [15, 16] can be re-written as $p(Y_i = y_i; \mu_i, \phi, |z) = \mathrm{NB}(y_i; \phi, z\mu_i)$, where $z \sim \mathrm{Sushila}(z; \theta, \alpha)$. Consequently, the above form can be thought as the generalized linear mixed model where the mixed effect or the frailty term z follows the Sushila distribution and is mixture of the exponential and gamma distributions. The hierarchical representation of this distribution can then be applied. The Sushila distribution, a two-parameter continuous distribution, was proposed by [19] denoted as

$$z \sim \frac{1}{1+\theta} \mathrm{Gamma}(2, \theta/\alpha) + \frac{\theta}{\theta+1} \mathrm{Gamma}(1, \theta/\alpha).$$

In the hierarchical representation concept, Sushila distribution would be

$$z \sim \mathrm{Gamma}(1 + \epsilon, \theta/\alpha) \quad \text{and} \quad \epsilon \sim \mathrm{Bernoulli}(1/(1+\theta)).$$

The entire multilevel models can be written hierarchically as

$$p(Y_i = y_i; \mu_i, \phi, |z) = \mathrm{NB}(y; \phi; z\mu_i); \quad \mu_i = \exp(\mathbf{x}_i^T \boldsymbol{\beta}), \quad z \sim \mathrm{Gamma}(z; 1 + \epsilon, \theta/\alpha),$$
$$\epsilon \sim \mathrm{Bernoulli}(\epsilon; 1/(1+\theta)).$$

4.2. Prior Distributions and joint posterior density

The prior distributions of all unknown parameters $(\boldsymbol{\beta}, \phi, \alpha, \text{and} \theta)$ are considered in Bayesian approach. In this study, we assume the prior distributions on $\boldsymbol{\beta}$ to have normal distribution. The gamma distribution is assumed as the priors for ϕ, α and θ, given as

$$\boldsymbol{\beta} \sim \mathrm{N}(\mathbf{b}_0, \mathbf{S}_\beta), \quad \phi \sim \mathrm{Gamma}(a_\phi, b_\phi), \quad \alpha \sim \mathrm{Gamma}(a_\alpha, b_\alpha), \quad \theta \sim \mathrm{Gamma}(a_\theta, b_\theta),$$

where $a_\phi > 0$, $b_\phi > 0$, $a_\alpha > 0$, $b_\alpha > 0$, $a_\theta > 0$, and $b_\theta > 0$ are known.

Suppose that $\mathbf{b}_0$ is a $p \times 1$ fixed hyper-parameter vector and $\mathbf{S}_\beta$ is $p \times p$; $p = k + 1$ known nonnegative-specific matrix. Each parameter is supposed to be independently distributed, and the joint prior distribution of all unknown parameters can be written as

$$\pi(\boldsymbol{\gamma}) = \pi(\boldsymbol{\beta})\pi(\phi)\pi(\alpha)\pi(\theta). \tag{8}$$

Now, we combine the likelihood function in (7) and the prior distribution in (8) to derive for the posterior distribution as follows:

$$\pi(\boldsymbol{\gamma}|\mathbf{y}, \mathbf{X}) \propto \prod_{i=1}^n f(y_i|\mathbf{x}_i^T, \boldsymbol{\gamma})\pi(\boldsymbol{\beta})\pi(\phi)\pi(\alpha)\pi(\theta). \tag{9}$$

For the NB-S model, the full conditional posterior distributions for each parameter of $\boldsymbol{\gamma}$ derived from (9) are obtained as

$$\pi(\boldsymbol{\beta}|\mathbf{y}, \mathbf{X}, \phi, \alpha, \theta) \propto \prod_{i=1}^{n} f(y_i|\mathbf{x}_i^T, \boldsymbol{\gamma})\pi(\boldsymbol{\beta}), \quad \pi(\phi|\mathbf{y}, \mathbf{X}, \boldsymbol{\beta}, \alpha, \theta) \propto \prod_{i=1}^{n} f(y_i|\mathbf{x}_i^T, \boldsymbol{\gamma})\pi(\phi),$$

$$\pi(\alpha|\mathbf{y}, \mathbf{X}, \boldsymbol{\beta}, \phi, \theta) \propto \prod_{i=1}^{n} f(y_i|\mathbf{x}_i^T, \boldsymbol{\gamma})\pi(\alpha), \quad \pi(\theta|\mathbf{y}, \mathbf{X}, \boldsymbol{\beta}, \phi, \alpha) \propto \prod_{i=1}^{n} f(y_i|\mathbf{x}_i^T, \boldsymbol{\gamma})\pi(\theta).$$

In this study, the model parameters can be easily obtained from Bayesian method using MCMC algorithm to produce the posterior inference for each parameter. Based on these prior densities, we generated three parallel independent MCMC chains for 30,000 iterations in each parameter, discarding the first 15,000 iterations as a burn-in for computation. One of the foremost packages used in Bayesian modeling is JAGS. It is an acronym for Just Another Gibbs Sampler. It can be run through the R environment [8]. In this paper, the expected posterior of parameters are calculated using function `jags` in `R2jags` package on CRAN of the R language [21].

In addition, three criteria which are the deviance, the deviance information criterion (DIC) and the effective number of parameters (p_D), are considered for model comparison. The DIC is regarded as generalization of the Akaike's information criterion and Bayesian information criterion. It is often used as a goodness-of-fit (GOF) measure when the Bayesian approach is used. The DIC is beneficial to Bayesian model comparison problems where the posterior distributions of the model have been obtained by MCMC simulation [22]. This criterion is based on deviance, $D(\boldsymbol{\gamma})$, defined as $D(\boldsymbol{\gamma}) = -2 \log f(\mathbf{y}|\boldsymbol{\gamma})$, where $f(\mathbf{y}|\boldsymbol{\gamma})$ is the likelihood function, the conditional joint probability density function of the observations is given the unknown parameters.

The DIC, proposed by [22], consists of two model components: a Bayesian measure of model fit and a measure of model complexity. The DIC is defined as DIC $= \bar{D}(\boldsymbol{\gamma}) + p_D$, where the first term is the posterior mean of the deviance, $\bar{D}(\boldsymbol{\gamma})$, given by $\bar{D}(\boldsymbol{\gamma}) = E[-2 \log f(\mathbf{y}|\boldsymbol{\gamma})]$, and the second term is an alternative measure of effective number of parameters [23], p_D, defined as half of the variance of deviance: $p_D = Var(D(\boldsymbol{\gamma}))/2$. This is the definition of p_D used in the `R2jags` package [21]. DIC and p_D are especially useful for comparison of hierarchical models [23].

The p_D measure should be close to the actual number of parameters [23]. Also, the p_D can indicate the variance of the deviance in the model [20]. For other measures, the model with the smallest values would be the best.

5. APPLICATION

In this section, two real data sets are utilized to check the performance of the NB-S GLM. The results of GLM based on NB-S as an alternative to Poisson and NB are compared using three criteria: deviance, DIC and p_D.

5.1. Doctor Visits Data Set

We consider the first data set, doctor visits [24], which provides the number of doctor visits during two weeks time period for a single-adult from the Australian Health Survey 1977 to 1978 (5190 observations). This data set were discussed by [25] in the demand for health services. For the first application, the response is the number of doctor visits, (**visits**), which is modeled by the covariates, including gender (**gender**), age in years (**age**), annual income in tens of thousands of dollars (**income**), the number of illness in 2 past weeks (**illness**), the number of days of reduced activity in 2 weeks due to illness or injury (**reduced**), the general health questionnaire score (**health**), private insurance indicator (**private**), government health insurance due to low income (**freepoor**), government health insurance due to old age, disability or veteran status (**freerepat**), a chronic condition not limiting (**ncronic**) and a chronic condition limiting activity (**lchron**).

The percentage of zeros in observed response, **visits**, is 93.98%. Also, this data set indicates over-dispersion problem with mean 0.302 and variance 0.637 (index of dispersion 2.111).

5.2. Squirrel Data Set

We consider the second data set, Squirrel (`nuts`), which included 52 observations were obtained from `COUNT` package in R language [21]. The data set comes from [26]. As originally reported by [27], this study recorded information about red squirrel behavior and forest attributes across various plots in Scotland's Abernathy Forest. The response variable is the number of cones (**cones**) stripped by red squirrels per plot. The explanatory variables consist of standardized number of trees per plot (**sntrees**), standardized mean tree height per plot (**sheight**) and standardized percentage of canopy cover per plot (**scover**). The stripped cone count was only considered when the mean diameter of trees was under 0.6 meters.

The percentage of zero in the number of cones is 96.2% and the index of dispersion is 23.153, indicating that there is a high percentage of zeros and over-dispersion with mean 17.923 and variance 414.974.

The histograms of response variables are evident with right-skewed distributions from the first and the second data sets and are illustrated in Figure 1, accompanying by summary statistics as shown in Table 1 and Table 2, respectively.

The distribution of responses in two data sets are examined by the Poisson, NB, and NB-S distributions. The GOF analysis based on Kolmogorov−Smirnov (KS) test statistic for discrete random variable [28] is shown in Table 3. It is shown that the p-value of the NB-S distribution is clearly the largest compared to two competitive distributions. Thus, the NB-S distribution is assumed for two data sets.

5.3. Results

This section presents the modeling results of GLM based on NB-S, which is compared to the Poisson and NB counterparts, applied to two data sets that described in previous section. The efficiency of the three models are investigated based on deviance, DIC and p_D. The closeness of p_D to actual number parameters and the smallest values of deviance and DIC indicate the best fitted performance.

The NB-S model needs to be adjusted with the mean of Sushila distribution, which is presented in (6). This is the same as the adjustment of the intercept of the NB-GE model as [16].

For the Bayesian approach, monitoring and reporting the convergence of the algorithm is essential for constructing results from the posterior distribution of interest. The convergent monitoring of this study is described by the posterior density plots, which illustrate smoothed distributions of the parameter values sampled in each chain [10].

By considering the Poisson, NB, and NB-S models for doctor visits and squirrel data sets, the posterior means (mean), standard error (s.e.) and 95% credible intervals (Cr.I.) of the model parameters after fitting the NB-S and the competing Poisson and NB models are presented in Tables 4 and 5 respectively.
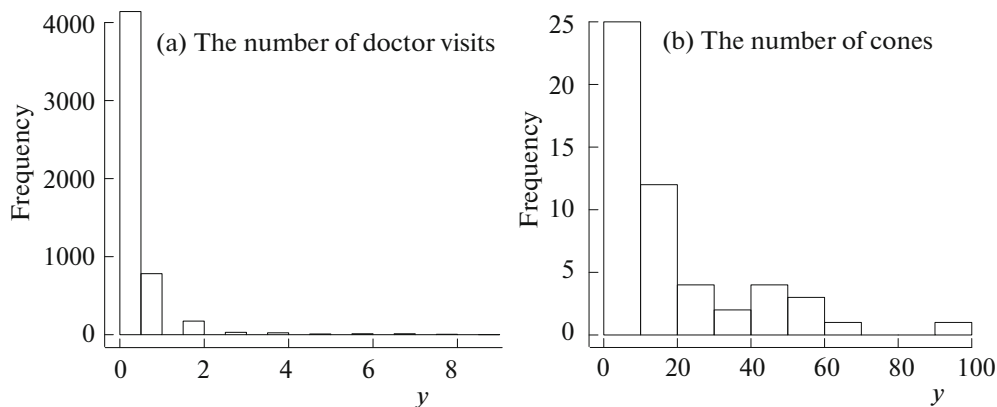
Fig. 1. (a) Histogram of the number of doctor visits in the doctor visits data set and (b) Histogram of the number of cones in the squirrel data set.
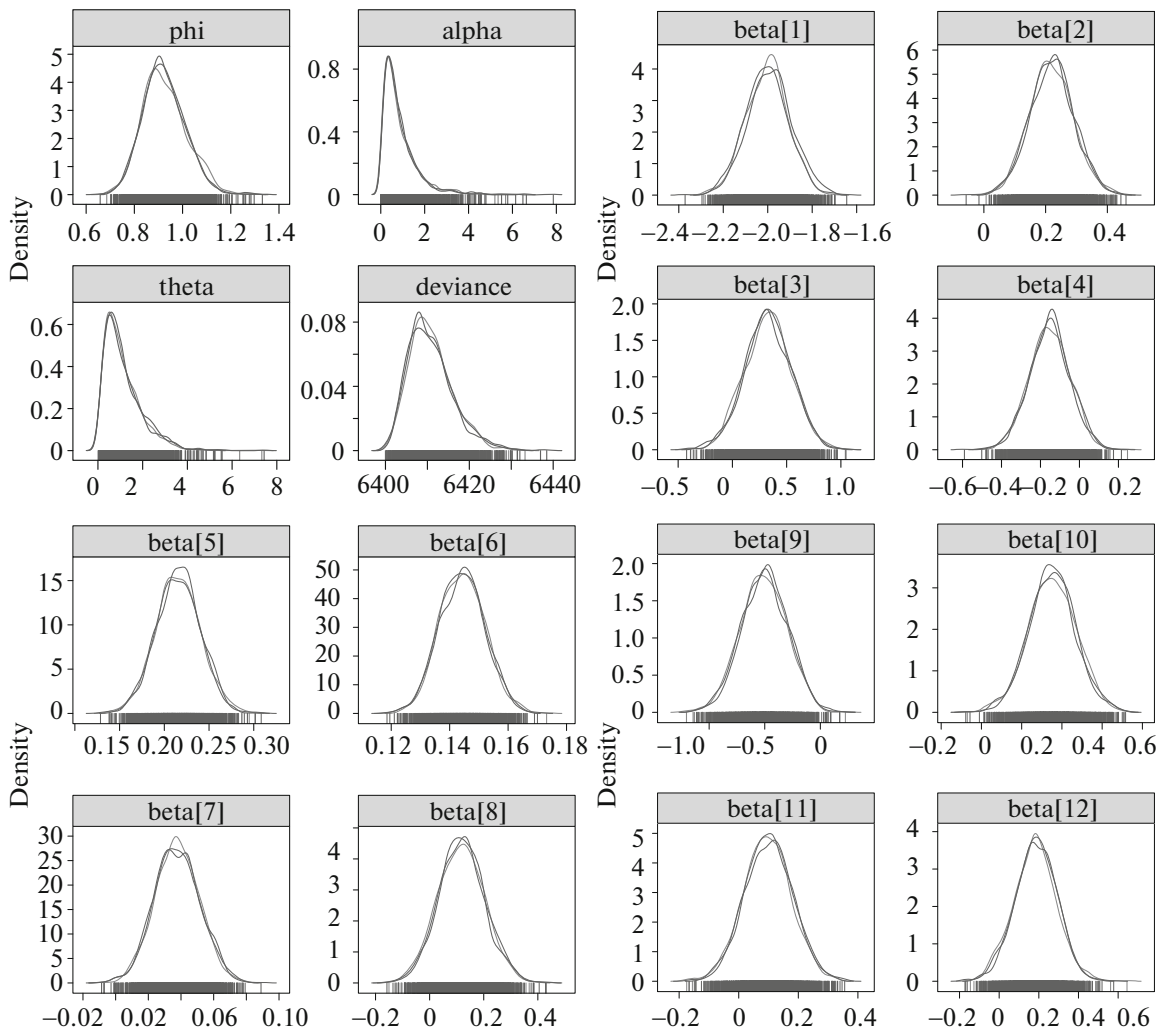
Fig. 2. Density plots for parameters ϕ, α, θ and $\beta_1, \beta_2, \ldots, \beta_{12}$, respectively from the NB-S linear model of doctor visits data set.
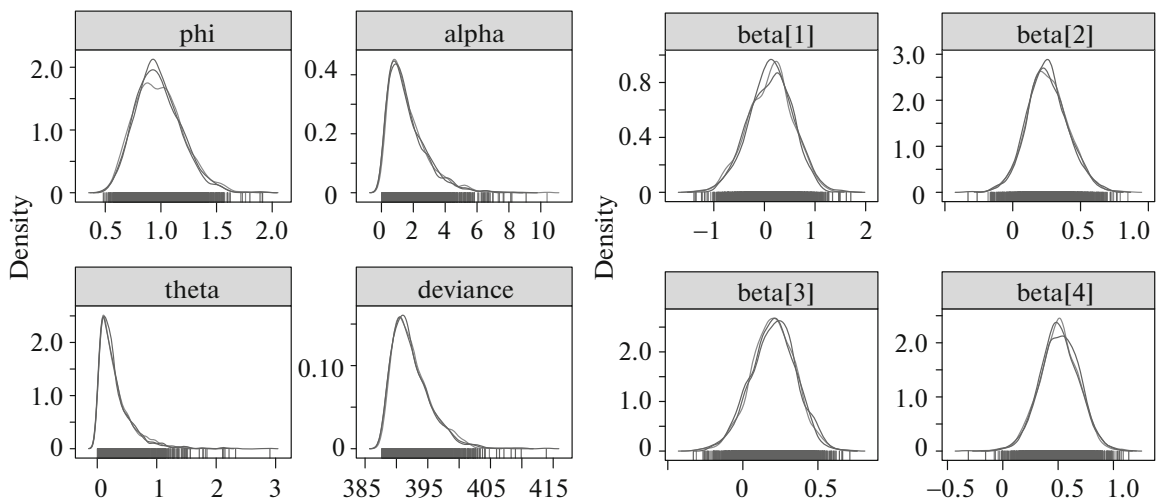

Fig. 3. Density plots for parameters ϕ, α, θ, $\beta_1, \beta_2, \beta_3$ and β_4, respectively from the NB-S linear model of squirrel data set.

Table 1. Summary statistics for the doctor visits data set ($n = 5190$)

Variables	Min	Max	Median
visits (response)	0	9	0
age (divided by 100)	0.19	0.72	0.32
income	0	1.5	0.55
illness	0	5	1
reduced	0	14	0
health	0	12	0
gender (1=female, 0=male)	–	–	–
private (1=yes, 0=no)	–	–	–
freepoor (1=yes, 0=no)	–	–	–
freerepat (1=yes, 0=no)	–	–	–
ncronic (1=yes, 0=no)	–	–	–
lchron(1=yes, 0=no)	–	–	–

Table 2. Summary statistics for the squirrel data set ($n = 52$)

Variables	Min	Max	Median
cones (response)	0	91	11
sntrees	−1.087	4.033	−0.366
sheight	−3.031	1.665	0.150
scover	−3.987	0.942	0.356

Table 3. The GOF analysis of the number of doctor visits in Doctor visits data set and the number of cones in Squirrel Data

GOF	Doctor visits data			Squirrel data		
	Poisson	NB	NB–S	Poisson	NB	NB–S
KS test statistics	0.058	0.013	0.009	0.483	0.310	0.078
p-value	< 0.0001	0.343	0.812	< 0.0001	< 0.0001	0.909

For doctor visits data, we formulate predictive model of the number of doctor visits related to 11 covariates. The performance of the models displayed in Table 6. The results indicate that the deviance and DIC values of the NB-S linear model are the smallest and the estimate of p_D for the NB-S model is close to the implicit number of parameters in the model. Furthermore, the p_D is pointed out that the variance of deviance for NB-S model is the smallest. All models have the same sign of predicted values. Therefore, NB-S model can be the alternative model to predict the number of doctor visits data. The predictive model can be written as:

$$\hat{\mu}_i = \exp(0.859 + 0.220\mathbf{gender}_i + 0.327\mathbf{age}_i - 1.53\mathbf{income}_i + 0.216\mathbf{illness}_i$$
$$+ 0.14\mathbf{reduced}_i + 0.038\mathbf{health}_i + 0.118\mathbf{private}_i - 0.504\mathbf{freepoor}_i$$
$$+ 0.152\mathbf{freerepat}_i + 0.199\mathbf{Nchronic}_i + 0.185\mathbf{lchronic}_i) \times 1.1257.$$

For Squirrel data, The estimated parameters, s.e and Cr.I. are presented in Table 5. The performance

Table 4. Summary results with mean, s.e. and Cr.I. from the posterior distribution under Poisson, NB, and NB-S models for the doctor visits data set

Parameter	Poisson		NB		NB-S	
	mean (s.e.)	Cr.I.	mean (s.e.)	Cr.I.	mean (s.e.)	Cr.I.
Intercept(β_1)	-2.014 (1.46×10^{-3})	$(-2.139, -1.891)$	-2.113 (1.06×10^{-3})	$(-2.266, -1.961)$	1.999 (1.34×10^{-3})	$(-2.189, -1.812)$
gender(β_2)	0.160 (6.53×10^{-4})	$(0.070, 0.249)$	0.191 (9.41×10^{-4})	$(-0.058, 0.319)$	-0.220 (9.83×10^{-4})	$(0.081, 0.360)$
age(β_3)	0.134 (1.15×10^{-3})	$(-0.032, 0.294)$	0.190 (2.74×10^{-3})	$(-0.201, 0.588)$	0.327 (2.95×10^{-3})	$(-0.090, 0.747)$
income(β_4)	-0.119 (8.96×10^{-3})	$(-0.239, 0.001)$	-0.249 (1.24×10^{-3})	$(-0.426, -0.075)$	-0.153 (1.44×10^{-3})	$(-0.357, 0.047)$
illness(β_5)	0.188 (2.56×10^{-4})	$(0.155, 0.223)$	0.208 (3.31×10^{-4})	$(-0.426, -0.075)$	0.216 (3.43×10^{-4})	$(0.167, 0.263)$
reduced(β_6)	0.130 (1.13×10^{-4})	$(0.121, 0.140)$	0.144 (1.09×10^{-4})	$(0.161, 0.256)$	0.144 (1.07×10^{-4})	$(0.130, 0.159)$
health(β_7)	0.027 (1.50×10^{-4})	$(0.007, 0.046)$	0.036 (1.89×10^{-4})	$(0.129, 0.159)$	0.038 (1.90×10^{-4})	$(0.011, 0.064)$
private(β_8)	0.088 (7.27×10^{-4})	$(-0.013, 0.189)$	0.105 (1.14×10^{-3})	$(0.009, 0.063)$	0.118 (1.12×10^{-3})	$(-0.044, 0.290)$
freepoor(β_9)	-0.138 (1.11×10^{-3})	$(-0.295, 0.019)$	-0.588 (2.82×10^{-3})	$(-1.007, -0.187)$	-0.504 (2.85×10^{-3})	$(-0.913, -0.111)$
freerepat(β_10)	0.097 (8.36×10^{-4})	$(-0.022, 0.216)$	0.137 (1.61×10^{-3})	$(-0.093, 0.369)$	0.152 (1.62×10^{-3})	$(-0.078, 0.386)$
nchronic(β_11)	0.063 (7.01×10^{-4})	$(-0.037, 0.161)$	0.098 (1.10×10^{-3})	$(-0.054, 0.252)$	0.099 (1.11×10^{-3})	$(-0.056, 0.254)$
lchronic(β_12)	0.090 (8.19×10^{-4})	$(-0.026, 0.207)$	0.185 (1.40×10^{-3})	$(-0.025, 0.382)$	0.185 (1.47×10^{-3})	$(-0.028, 0.389)$
ϕ	$-$	$-$	0.939 (1.34×10^{-3})	$(0.774, 1.146)$	0.924 (1.23×10^{-3})	$(0.763, 1.111)$
α	$-$	$-$	$-$	$-$	0.856 (1.11×10^{-2})	$(0.071, 3.113)$
θ	$-$	$-$	$-$	$-$	1.119 (1.12×10^{-2})	$(0.131, 3.227)$

of the model fittings is shown in Table 7 and the results show that the deviance and DIC clearly present the adequacy of NB-S model. The p_D of NB-S model is closer to the actual number of parameters in the model compared to other models. As a result, the comparison of the model fittings on the squirrel data indicate that the NB-S model can be used to predicted the number of cones in this data set. The

Table 5. Summary results with mean, s.e. and Cr.I. from the posterior distribution under Poisson, NB, and NB-S models for the squirrel data set

Parameter	Poisson		NB		NB-S	
	mean (s.e.)	Cr.I.	mean (s.e.)	C.I.	mean (s.e.)	Cr.I.
Intercept(β_1)	2.727 (7.46×10^{-3})	(2.641, 2.815)	2.454 (1.90×10^{-2})	(2.173, 2.714)	0.126 (6.05×10^{-2})	$(-0.744, 0.940)$
sntrees(β_2)	0.243 (4.09×10^{-3})	(0.187, 0.301)	0.264 (1.98×10^{-2})	$(-0.002, 0.551)$	0.258 (2.08×10^{-2})	$(-0.028, 0.572)$
snheight(β_3)	0.390 (6.40×10^{-3})	(0.046, 0.477)	0.208 (1.95×10^{-2})	$(-0.073, 0.478)$	0.203 (2.06×10^{-2})	$(-0.094, 0.487)$
scover(β_4)	0.775 (1.17×10^{-2})	(0.641, 0.904)	0.539 (2.32×10^{-2})	(0.206, 0.862)	0.511 (2.42×10^{-2})	(0.169, 0.850)
ϕ	–	–	0.937 (2.76×10^{-2})	(0.588, 1.376)	0.978 (2.86×10^{-2})	(0.628, 1.425)
α	–	–	–	–	1.635 (1.71×10^{-1})	(0.176, 4.758)
θ	–	–	–	–	0.297 (3.89×10^{-2})	(0.021, 1.049)

Table 6. The performance of the models for the doctor visits data set

Criterion	Poisson	NB	NB-S
Deviance	6741.97	6411.83	6410.79
DIC	66886.24	6426.08	6424.73
p_D	60144.26	14.25	13.93

Table 7. The performance of the models for the squirrel data set

Criterion	Poisson	NB	NB-S
Deviance	956.110	394.970	392.404
DIC	1673.275	405.061	397.063
p_D	681.165	10.091	4.659

predicted model can be given as:

$$\hat{\mu}_i = \exp(0.126 + 0.258\mathbf{sntrees}_i + 0.258\mathbf{sheight}_i + 0.203\mathbf{scover}_i) \times 9.753.$$

The posterior density plots of all parameters in the NB-S model for doctor visits and squirrel data sets are illustrated in Fig. 2 and 3, respectively.

The results reveal that the density plots of all parameters in three chains of both data sets overlap well after burn-in period. Therefore, the NB-S model can fit both data sets in context of Bayesian estimation framework.

6. DISCUSSION

Since, the NB-S model is a new model for GLM framework, it has never been implemented to predict the count phenomenon with related covariates. There are some interesting findings that are necessary to be discussed. First, the parametrization of the NB-S model in this study is slightly different from that proposed by [17]. However, it is related to the NB-L GLM in [15] and the NB-GE GLM in [16]. The authors discussed the problems of the parametrization including a non-linear of the mean response, non-invertible function of the covariates and the parameters. On the contrary, the parametrization suggested in this study is easily explained (See more details in [15] and [16]). [15] discussed some problems about this parametrization in GLM framework of NB-L model which are the main concepts of this study. The authors also suggested some methods to solve the problem, using the dispersion parameter (ω) instead of the inverse dispersion parameter (ϕ), the underlying hierarchical illustration of the model, and the adjusted mean of response. The adjustment is denoted by $E(Y_i) = \mu_i \times E(z)$ and then applied to the NB-S model. Furthermore, the GOF for the responses of two real data is performed prior to model fitting. Lastly, the processing time for MCMC of NB-S model is slightly longer than the NB and Poisson models because it entails additional parameters which will be supportive of the fitted model.

7. CONCLUSION

Our article describes the utilization of GLM to develop the NB-S linear model. The parameter estimation is obtained via the Bayesian approach using MCMC method and its measurement can be easily carried out with the Bayesian Software package, namely JAGS, which can be run through R language.The prior and likelihood information are also discussed. We apply the NB-S linear model to two real data sets: doctor visits and squirrel data sets, and compare or results with traditional Poisson and NB models. This selection is supported by focus measurements: deviance DIC and p_D. These measurements are considered in the assessment and comparison of models in this article. As a result, the DIC and p_D are particularly useful for comparing hierarchical models, as discussed in [23]. For a comparison of the models, the deviances and DICs for models show that the best fit is the NB-S model, followed by the NB and finally the Poisson models. The p_D values of the NB-S model are close to the true numbers of parameters and imply that the variance of deviance of the NB-S model are the smallest variance for the two data sets. To sum up, the NB-S model seems to outperform other models in this study. It can be an alternative model to study a relationship between an over-dispersed count response variable and a set of covariates. We hope that the NB-S GLM may proffer a feasible alternative to the traditional Poisson and NB GLMs for analyzes of count data. For future work, we will study simulation of the proposed model to analyze the performance of the NB-S linear model.

REFERENCES

1. I. Ozmen and F. Famoye, "Count regression models with an application to zoological data containing structural zeros," J. Data Sci. **5** (4), 491−502 (2007).
2. J. Hilbe, *Negative Binomial Regression* (Cambridge Univ. Press, New York, 2011).
3. M. Bulmer, "On fitting the Poisson lognormal distribution to species-abundance data," Biometrics **30** (1), 101−110 (1974).
4. E. Gómez-Déniz, J. Sarabia, and E. Calderín-Ojeda "Univariate and multivariate versions of the negative binomial-inverse Gaussian distributions with applications," Insurance: Math. Econ. **42** (1), 39−49 (2008).
5. Z. Hossein and I. Noriszura, "Negative binomial-Lindley distribution and its application," J. Math. Stat. **6** (1), 4−9 (2010).
6. S. Aryuyuen and W. Bodhisuwan, "The negative binomial-generalized exponential (NB-GE) distribution," Appl. Math. Sci. **7** (22), 1093−1105 (2013).
7. N. Johnson, A. Kemp, and S. Kotz, *Univariate Discrete Distributions* (Wiley, Chichester, 2005).
8. J. Hilbe, *Modeling Count Data* (Cambridge Univ. Press, New York, 2014).
9. A. Agresti, *An Introduction to Categorical Data Analysis* (Wiley, New York, 1996).
10. J. Ntzoufras, *Bayesian Modeling Using WinBUGS* (Wiley, Hoboken, NJ, 2011).
11. J. Lawless, "Negative binomial and mixed Poisson regression," Canad. J. Stat. **15** (3), 209−225 (1987).
12. J. Rodríguez-Avi, A. Conde-Sánchez, A. Sáez-Castillo, M. Olmo-Jiménez, and A. Martínez-Rodríguez, "A generalized Waring regression model for count data," Comput. Stat. Data Anal. **53** (10), 3717−3725 (2009).
13. H. Zamani, N. Ismail, and P. Faroughi, "Poisson-weighted exponential univariate version and regression model with applications," J. Math. Stat. **10** (2), 148 (2014).

14. W. Wongrin and W. Bodhisuwan, "Generalized Poisson-Lindley linear model for count data," J. Appl. Stat. **44** (15), 2659−2671 (2016).
15. S. Geedipally, D. Lord, and S. Dhavala, "The negative binomial-Lindley generalized linear model: characteristics and application using crash data," Accident Anal. Prevent. **45**, 258−265 (2012).
16. P. Vangala, D. Lord, and S. Geedipally, "Exploring the application of the Negative Binomial-Generalized Exponential model for analyzing traffic crash data with excess zeros," Anal. Methods Accident Res. **7**, 29−36 (2015).
17. D. Yamrubboon, W. Bodhisuwan, C. Pudprommarat, and L. Saothayanun, "The Negative Binomial-Sushila distribution with application in count data analysis," Thailand Stat. **15** (1), 69−77 (2017).
18. J. Nelder and R. Wedderburn, "Generalized linear models," J. R. Stat. Soc., Ser. A: Gen. **135** (3), 370−384 (1972).
19. R. Shanker, S. Sharma, U. Shanker, and R. Shanker, "Sushila distribution and its application to waiting times data," Opinion—Int. J. Business Manage. **3** (2), 1−11 (2013).
20. A. Gelman, J. Carlin, H. Stern, D. Dunson, A. Vehtari, and D. Rubin, *Bayesian Data Analysis* (CRC, Boca Raton, FL, 2014).
21. *R: A Language and Environment for Statistical Computing, R Foundation for Statistical Computing* (Vienna, Austria, 2010).
22. D. Spiegelhalter, N. Best, B. Carlin, and A. van der Linde, "Bayesian measures of model complexity and fit," J. R. Stat. Soc., Ser. B: Stat. Methodol. **64** (4), 583−639 (2002).
23. D. Lunn, C. Jackson, N. Best, A. Thomas, and D. Spiegelhalter, *The BUGS Book: A Practical Introduction to Bayesian Analysis* (CRC, Boca Raton, FL, 2012).
24. A. Cameron and P. Trivedi, *Regression Analysis of Count Data* (Cambridge Univ. Press, New York, 2013).
25. A. Cameron, P. Trivedi, F. Milne, and J. Piggott, "A microeconometric model of the demand for health care and health insurance in Australia," Rev. Econ. Stud. **55** (1), 85−106 (1988).
26. A. Zuur, J. Hilbe, and E. Ieno, *A Beginner's Guide to GLM and GLMM with R: A Frequentist and Bayesian Perspective for Ecologists* (Highland Statistics, 2013).
27. S. Flaherty, G. Patenaude, A. Close, and P. Lurz, "The impact of forest stand structure on red squirrel habitat use," Forestry **85** (3), 437−444 (2012).
28. T. Arnold and J. Emerson, "Nonparametric goodness-of-fit tests for discrete null distributions," R J. **3** (2), 34−39 (2011).