Safety Science 62 (2014) 495–498

Contents lists available at ScienceDirect

Safety Science

journal homepage: www.elsevier.com/locate/ssci

A caution about using deviance information criterion while modeling traffic crashes

Srinivas Reddy Geedipally [a,*], Dominique Lord [b,1], Soma Sekhar Dhavala [c]

[a] Texas A&M Transportation Institute, Texas A&M University System, 110 N. Davis Dr. Suite 101, Arlington, TX 76013-1877, United States
[b] Zachry Department of Civil Engineering, Texas A&M University, 3136 TAMU, College Station, TX 77843-3136, United States
[c] Dow AgroSciences, Indianapolis, IN, United States

ARTICLE INFO

Article history:
Received 18 June 2013
Accepted 6 October 2013
Available online 30 October 2013

Keywords:
Deviance information criterion
Poisson-Gamma
Negative binomial
Traffic crashes

ABSTRACT

The Poisson-Gamma (PG) or negative binomial (NB) model still remains the most popular method used for analyzing count data. In the software WinBUGS (or any other software used for Bayesian analyses), there are different ways to parameterize the NB model. In general, either a PG (based on the Poisson-mixture) or a NB (based on the Pascal distribution) modeling framework can be used to relate traffic crashes to the explanatory variables. However, it is important to note that the way the model is parameterized will influence the output of the Deviance Information Criterion (DIC) values. The objective of this short study is to document the difference between the PG and NB models in the estimation of the DIC. This is especially important given that the NB/PG model is still the most frequently used model in highway safety research and applications. To accomplish the study objective, PG and NB models were developed using the crash data collected at 4-legged signalized intersections in Toronto, Ont. The study results showed that there is a considerable difference in the estimation of the DIC values between the two models. It is thus recommended not to consider the DIC as the sole model selection criterion and the comparison should be done only between the models that have similar parameterization. Other alternatives such as Bayes Factors, Posterior predictive performance criterion, Bayesian Information Criterion (BIC), among others need to be considered in addition to the DIC in the model selection.

1. Introduction

Despite the recent developments in regression modeling and analysis techniques, the Poisson-Gamma (PG) or negative binomial (NB) model still remains the most popular method used for analyzing count data (Hilbe, 2011). Its popularity in highway safety is no different than with other fields of research (Lord and Mannering, 2010). The extensive application of the PG/NB model is explained by its ability to capture (moderate) over-dispersion (i.e., the variance is large than the mean), the simplicity in manipulating the relationship between the mean and the variance, and the fact the model is available in all commercially available statistical programs.

The PG/NB model can be derived using several approaches (Hilbe, 2011). The most common approach is based on the PG mixture distribution (Lawless, 1987; Cameron and Trivedi, 1998). The PG model has properties that are very similar to the Poisson model in which the dependent variable Y_i is modeled as a Poisson variable with a mean μ_i where the model error is assumed to follow a Gamma distribution. As it names implies, the Poisson-Gamma is a mixture of two distributions and was first derived by Greenwood and Yule (1920). This mixture distribution was developed to account for over-dispersion that is commonly observed in discrete or count data (Lord et al., 2005). It became very popular because the conjugate distribution (same family of functions) has a closed form and leads to NB distribution. As discussed by Cook (2009), “the name of this distribution comes from applying the binomial theorem with a negative exponent”.

Recently, researchers in statistics and highway safety have been using an alternative parameterization of the NB distribution for analyzing count data (Zamani and Ismail, 2010; Lord and Geedipally, 2011). This parameterization is based on the probability of successes and failures in successive trials (Casella and Berger, 1990). This process is also referred to as the Pascal distribution. The proposed parameterization was needed for the development of the NB-Lindley model (Geedipally et al., 2012). In theory, the PG (based on the Poisson-mixture distribution) and the NB (based on the Pascal distribution) models will provide the same estimates. During the development of the NB-Lindley model (Geedipally et al. 2012), it was noted that, although the PG and NB models provided the same modeling output (i.e., coefficients, standard errors, etc.),

* Corresponding author. Tel.: +1 (817) 462 0519.
E-mail addresses: srinivas-g@ttimail.tamu.edu (S.R. Geedipally), d-lord@tamu.edu (D. Lord), somasd@gmail.com (S.S. Dhavala).
[1] Tel.: +1 (979) 458 3949.

0925-7535/$ - see front matter
http://dx.doi.org/10.1016/j.ssci.2013.10.007

differences were observed with the Deviance Information Criterion (DIC), a commonly used goodness-of-fit (GOF) measure used for assessing the performance of competitive Bayesian models. The researchers investigated this difference and this paper documents this effort in order to warn other researchers about potential issues and pitfalls with using the DIC for comparing Bayesian models.

2. Background

This section briefly describes the differences between the parameterization of the PG and the NB models.

2.1. Poisson-Gamma model

For modeling traffic crash data, researchers have been using the following model structure for the PG model. The crash frequency 'y_i' for a particular ith site when conditional on its mean μ_i is Poisson distributed and independent over all sites and time periods (Miaou and Lord, 2003):

$$y_i \mid \mu_i \sim \text{Poisson } (\mu_i) \quad i = 1, 2, \ldots, I \tag{1}$$

The crash mean μ_i is structured as:

$$\mu_i = f(X; \beta) \exp(e_i) \tag{2}$$

where $f(X; \beta)$ is a function of the explanatory variables (X), β is a vector of coefficients that are estimated from the data; and, e_i is the model error independent of all the covariates, which follows a gamma distribution with same shape and location parameters.

From the above equations, it can be shown that y_i, conditional on μ_i and ϕ, is distributed as a PG random variable with a mean μ_i and a variance $\mu_i + \mu_i^2/\phi$, respectively. The probability mass function (PMF) of the PG structure described above is given by the following equation:

$$f(y_i; \phi, \mu_i) = \frac{\Gamma(y_i + \phi)}{\Gamma(\phi) y_i!} \left[\frac{\phi}{\mu_i + \phi}\right]^{\phi} \left[\frac{\mu_i}{\mu_i + \phi}\right]^{y_i} \tag{3}$$

where y_i, response variable for site i; μ_i, mean response for site i; and, ϕ, inverse dispersion parameter of the PG distribution.In the software WinBUGS (Spiegelhalter et al., 2003), for example, the coefficients of the PG regression model will be estimated using the following parameterization:

$$y(i) \sim dpois(\mu[i])$$

$$\mu[i] = f(X; \beta) \exp(e[i])$$

$$\exp(e[i]) \sim dgamma(\phi, \phi)$$

It can be can recognized that the PG model is a hierarchical model, where the Poisson component constitutes the likelihood at the data level and the gamma distribution appears in the next hierarchy at random effects level.

2.2. Negative binomial model

As discussed above, the NB can be derived using the probability of successes and failures in successive trials (Benjamin and Cornell, 1970). It can be shown that the PMF of the NB distribution can be given as:

$$P(Y = y_i; \phi, p_i) = \frac{\Gamma(\phi + y_i)}{\Gamma(\phi) \times y_i!} (1 - p_i)^{\phi} (p_i)^{y_i}; \quad \phi > 0, \; 0 < p_i < 1 \tag{4}$$

The parameter 'p' is defined as the probability of success in each trial and is given as:

$$p_i = \frac{\mu_i}{\mu_i + \phi} \tag{5}$$

where μ_i, mean response for observation i; and, ϕ, inverse dispersion parameter of the NB distribution.

In the software WinBUGS (Spiegelhalter et al., 2003), the coefficients of the NB regression model will be estimated using the following parameterization:

$$y(i) \sim dnegbin(p[i], \phi)$$

$$p[i] = \frac{\phi}{\phi + \mu[i]}$$

$$\mu[i] = f(X; \beta)$$

3. Methodology

This section describes the functional form used for estimating the models and the deviance information criterion.

3.1. Functional form

The functional form used for models is as follows:

$$\mu_i = \beta_0 F_{Maj\_i}^{\beta_1} F_{Min\_i}^{\beta_2} \tag{6}$$

where μ_i, the mean number of crashes per year for intersection i; F_{Maj_i}, entering flow for the major approach (average annual daily traffic or AADT) for intersection i; and, F_{Min_i}, entering flow for the minor approach for intersection i.

3.2. Deviance Information Criterion (DIC)

The DIC is a widely used GOF statistic for comparing models in a Bayesian framework (Spiegelhalter et al. 2002). DIC is a hierarchical modeling generalization of the Akaike Information Criterion (AIC) and Bayesian Information Criterion (BIC), defined as:

$$DIC = \overline{D(\theta)} + P_D \tag{7}$$

and

$$P_D = \overline{D(\theta)} - D(\bar{\theta}) \tag{8}$$

where θ represents the collection of parameters, P_D is a measure of model complexity and is interpreted as the effective number of parameters. The larger the P_D, the easier it is to fit the model to the data. $\overline{D(\theta)} = E[-2\log L]$ is the expectation of the deviance under the posterior of the un-standardized model, where L is the model likelihood. Larger $\overline{D(\theta)}$ values correspond to a worst fit. $D(\bar{\theta})$ is the deviance evaluated at a posterior summary of θ, which is typically the mean but the median or mode can also be considered when appropriate. Models with smaller DIC should be preferred to models with larger DIC. Models are penalized by the value of $\overline{D(\theta)}$, which will decrease as the number of parameters in a model increases, and P_D, which compensates for this effect by favoring models with a smaller number of parameters.

One of the drawbacks of the DIC is that it is not invariant to re-parameterization, and therefore, parameterization of the models must be carefully chosen. Formal justification for the DIC requires that the posterior be approximately normal, which may or may not be true in practice. Despite these drawbacks, the DIC is very popular and widely used due its simplicity and that it is readily available as built-in tool in software like WinBUGS. Such easy access to this tool has one important ramification: DIC can be misused and misinterpreted. Below, the researchers explore the reasons that are well documented but often overlooked (Millar, 2009).

The definition of the DIC presented above is not unique in the context of multi-level models or hierarchical models and in fact depends very much on what part of the model is being considered as