Analytic Methods in Accident Research 9 (2016) 44–53

Contents lists available at ScienceDirect

Analytic Methods in Accident Research

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

Fast Bayesian inference for modeling multivariate crash counts

Volodymyr Serhiyenko [a], Sha A. Mamun [b,*], John N. Ivan [b], Nalini Ravishanker [a]

[a] University of Connecticut, Department of Statistics, Storrs, CT, USA
[b] University of Connecticut, Department of Civil and Environmental Engineering, Storrs, CT, USA

ARTICLE INFO

Article history:
Received 25 March 2015
Received in revised form
17 February 2016
Accepted 17 February 2016
Available online 5 March 2016

Keywords:
Multivariate Poisson Lognormal regression
Integrated Nested Laplace Approximations
Multivariate count modeling
Freeway
Divided limited access highway segments
Crash type count analysis

ABSTRACT

This paper investigates the multivariate Poisson Lognormal modeling of counts for different types of crashes. This multivariate model can account for the overdispersion as well as positive and/or negative association between counts. Approximate Bayesian inference via the Integrated Nested Laplace Approximations significantly decreases computational time which makes it attractive for researchers. The models are developed for single vehicle, same direction and opposite direction crash types using three years (2009–2011) of crash data on Connecticut divided limited access highway segments. Annual average daily traffic, segment length, and road specific covariates (median type, shoulder width, area type, and on-ramp indicator) are used as predictor variables. The results from the multivariate Poisson Lognormal model suggest that an increase in the annual average daily traffic, segment length, and shoulder width together with urban area type and presence of an on-ramp are associated with in an increase in crashes. The median type covariate has a mixed effect for different median types on different type of crashes. The multivariate Poisson Lognormal model results are compared with the results obtained from two univariate regression models, univariate Poisson Lognormal and univariate negative binomial, with respect to model implications and precision on analysis of crash counts. The results show that the coefficient estimates of predictors have almost similar effects across all three crash type count models; however, standard errors in the multivariate Poisson Lognormal model are smaller than standard errors from other two univariate models in most cases. Results on posterior means for the correlation coefficients between crash types indicate that there are significant correlations exist between the crash count vectors, which indicate that ignoring such a correlation could possibly lead to incorrect variance estimation for the parameters. Results on predicted mean absolute error (PMAE) indicate that Bayesian multivariate Poisson Lognormal model provides up to 33% less prediction error compared to the univariate negative binomial model, although there are no significant difference of PMAE values between multivariate and univariate Poisson Lognormal models results. The analysis results demonstrated that the Bayesian multivariate Poisson Lognormal model provides correct estimates for parameters in predicting crash counts by accounting for correlations in the multivariate crash counts.

Published by Elsevier Ltd.

* Corresponding author.

http://dx.doi.org/10.1016/j.amar.2016.02.002
2213-6657/Published by Elsevier Ltd.

1. Introduction

The economic and societal losses due to motor vehicle crashes in the USA exceed US$870 billion, nearly $900 per capita based on calendar year 2010 data (NHTSA, 2014a). In 2013, 32,719 people died and 2.3 million people were injured in motor vehicle crashes in the United States (NHTSA, 2014b). The situation is of particular interest on limited access highways, which experience significantly higher fatality rates than local and collector roads. About 60% of total 33,561 fatalities during 2012 occur on limited access highways (IIHS, 2014).

Given this context, there has been considerable interest in developing crash analysis models to assess the safety of limited access highway segments (Persaud and Dzbik, 1993; Kraus et al., 1993; Shankar et al., 1996; Golob et al., 2004; Lord et al., 2008; Bonneson et al., 2012). The most widely used crash count models in traffic safety studies consider the component responses as univariate (or scalar) count random variables, that is, traffic crash counts for each crash types are independent and estimated separately (King, 1989; Poch and Mannering, 1996; Ivan et al., 2000; Hauer, 1997; Miaou and Lord, 2003). Univariate Poisson regression and negative binomial (NB) regression models are two commonly used univariate distribution models to predict the total number of crash counts. Univariate Poisson regression is used under the assumption of nominal dispersion (Kim et al., 2002; Miranda-Moreno, 2006) and, if there is evidence of overdispersion, the negative binomial regression is used, since the negative binomial probability distribution is a Poisson-Gamma mixture distribution for which theoretical variance is greater than its expectation (Hauer, 2001). However, neither of these methods can accommodate the dependence between the components of the response vector. Both models ignore potential correlations due to latent factors present across crash types for a specific segment of roadway. Also the univariate model framework is not able to explain the effects of unobserved explanatory variables that might affect the crash frequencies (Mannering and Bhat, 2014).

In recent years, multivariate Poisson Lognormal (MVPLN) specification has been used to have better estimation of crash counts considering the dependence and likely correlation between response variables. However, current methods to deal with multivariate data in multivariate Poisson Lognormal model specification are relatively cumbersome and time consuming due to high dimensional dependencies (Mannering and Bhat, 2014). This paper investigates the application of Integrated Nested Laplace Approximations (INLA) in multivariate Poisson Lognormal crash count model to explain the occurrence of different types of crashes as a function of covariates, and conditional on a crash of a given type. Approximate Bayesian inference via the Integrated Nested Laplace Approximations significantly decreases computational time. It was shown that INLA is approximately 7 times faster than Markov Chain Monte Carlo (MCMC) in the case of multivariate Poisson Log-Normal models (Serhiyenko, 2015). For comparison purpose, univariate Poisson Lognormal and univariate negative binomial models also are estimated and the results are compared with the multivariate Poisson Lognormal model results.

2. Literature review

Application of multivariate models to investigate the likely correlations among crash counts have been proposed in the literatures. Serhiyenko et al. (2014) proposed a model that combines compositional modeling with the vector autoregression to account for the correlations between different severity of pedestrian crashes. Ma and Kockelman (2006) applied a multivariate Poisson approach to assess the effects of various covariates on the multivariate crash counts by severity. The multivariate Poisson regression models used by Ma and Kockelman (2006) assumed equal positive correlations across crash counts. In addition, the multivariate Poisson models do not allow for overdispersion that is often observed in the crash data. The assumption of equal covariance has been relaxed by Karlis and Meligkotsidou (2005); however, this extended multivariate Poisson model still could not explain the negative association between covariates. In order to account for overdispersion, Ladron de Guevara and Washington (2004) used a multivariate negative binomial model to investigate the simultaneity of crash counts and severity. However, the multivariate negative binomial model model with a proposed estimation method still has the shortcomings of inability to estimate the correlation between three or more variables.

In recent years, multivariate Poisson Lognormal regression approach developed by Chib and Winkelmann (2001) has been used as a good alternative to the multivariate Poisson regression approaches for analysis of multivariate crash count data (Aitchison and Ho, 1989; Park and Lord, 2007; Ma et al., 2008). The multivariate Poisson Lognormal approach can account for negative correlations and also allows modeling the dependence in the response vector for data that are possibly overdispersed. However, the implementation of multivariate Poisson Lognormal models is not straightforward and a major drawback of this approach is that the likelihood function is complicated, so that the computational time for modeling grows rapidly with an increase in the count vector dimension (Ravishanker et al., 2014). Therefore, it is necessary to adapt a simulation based method to cope with the complex likelihood function and to improve the computational time (Park and Lord, 2007).

Recently, the Bayesian framework using Markov Chain Monte Carlo (MCMC) simulation method for the multivariate Poisson Lognormal model estimation were undertaken to address the computational complexity. Park and Lord (2007) employed the MCMC simulation method in multivariate Poisson Lognormal models to assess the effects of covariates on crash counts. They found the multivariate Poisson Lognormal model can provide more accurate estimates by accounting for correlations in the multivariate crash counts. Ma et al. (2008) introduces an multivariate Poisson Lognormal approach to simultaneously model crash counts by injury severity. They have also used the MCMC simulation method and estimated the