Journal of the Korean Data & Information Science Society
2012, **23**(4), 851–858
http://dx.doi.org/10.7465/jkdi.2012.23.4.851
한국데이터정보과학회지

Bayesian analysis for the bivariate Poisson regression model: Applications to road safety countermeasures[†]

Hyeong-Gu Choe[1] · Joon-Beom Lim[2] · Yongho Won[3] · Soobeom Lee[4] · Seong W. Kim[5]

[135]Division of Applied Mathematics, Hanyang University
[24]Department of Transportation Engineering, University of Seoul
Received 20 June 2012, revised 18 July 2012, accepted 23 July 2012

Abstract

We consider a bivariate Poisson regression model to analyze discrete count data when two dependent variables are present. We estimate the regression coefficients associated with several safety countermeasures. We use Markov chain and Monte Carlo techniques to execute some computations. A simulation and real data analysis are performed to demonstrate model fitting performances of the proposed model.

Keywords: Accident prediction model, bivariate Poisson distribution, Gibbs sampler, Metropolis-Hastings algorithm, safety countermeasure.

1. Introduction

Analysis of discrete count data has been done in various fields, including applied sciences, economics, and transportation engineering. Several models have been proposed in recent decades to analyze these count data, especially traffic crash data. Modeling these data begins with the multiple linear regression model (MLR) in conjunction with normality assumptions and homoscedasticity on error terms. The MLR is conceptually sound and computationally feasible in estimating regression coefficients and explaining correlations between variables. However, these conditions are not compatible with characteristics of crash data. Jovanis and Chang (1986) initially propose the Poisson regression model (PRM) to circumvent these limitations. However, because the Poisson distribution has the same mean and variance, it does not often reflect typical characteristics such as substantial overdispersion contained in count data. A great deal of work regarding model building for crash data has been based on the negative binomial regression model (NBRM). See Persaud (1991, 1994), Persaud and Dzbik (1993), Lord and Persaud (2004) for the related work based on the NBRM.

---

† This work was supported by the National Research Foundation of Korea (NRF) grant funded by the Korea government (MEST) (No. 2011-0028933).
1 Undergraduate student, Division of Applied Mathematics, Hanyang University, Ansan 426-791, Korea.
2 Graduate student, Department of Transportation Engineering, University of Seoul, Seoul 130-743, Korea.
3 Undergraduate student, Division of Applied Mathematics, Hanyang University, Ansan 426-791, Korea.
4 Professor, Department of Transportation Engineering, University of Seoul, Seoul 130-743, Korea.
5 Corresponding author: Professor, Division of Applied Mathematics, Hanyang University, Ansan 426-791, Korea. E-mail: seong@hanyang.ac.kr

There are two types of crash prediction models in transportation safety fields. One is the conventional prediction model in which the total number accidents is simply the sum of the frequencies contained in different severity: "property damage only," "injury," and "fatal and injury." So, the PRM or NBRM can be proceeded with this total number of accidents as a dependent variable. The other model is called a severity model in a sense that one can make an adjustment by imposing some weights in each severity category. Consequently, the severity model can be more focused on severe accidents like fatal injury. However, both models use only one dependent variable, which cannot incorporate the correlations between severities.

Multivariate Poisson distributions (MPD) are useful tools in applications involving multivariate discrete data. A decent number of techniques has been proposed to deal with the computational problems of the multivariate Poisson distribution. Kim *et al.* (2006) and Kim and Jeong (2006) investigate some theoretical properties for MPD inclduing random number generation algorithms and solutions to linear equations. There have been some papers regarding the multivariate Poisson (MVP) regression models developed by Tsionas (2001). Tunaru (2002) utilizes a hierarchical model to analyze multiple count data. Bijleveld (2005) investigates the correlation between the number of accidents and victims with a MVP model. Ma and Kockelman (2006) consider an objective Bayesian approach in analyzing injury count data by severity based on trivariate Poisson regression model. Our analysis is quite similar to that of Ma and Kockelman (2006), though, there are some differences. Specifically, we use a bivariate Poisson regression model with some informative proper priors. To the best of our knowledge, the estimation of the model parameters with this bivariate model in analyzing Korean crash data has not been given in the literature. Moreover, it is a bit meaningful to estimate the regression coefficients and interpret the effects associated with safety countermeasures. The detailed results of analysis will be presented extensively in Section 4. In conjunction with bivariate Poisson distributions, Papageorgiou and Loukas (1988) compute the MLEs by considering the conditional distributions. Adamids and Loukas (1994) utilize the EM algorithm to estimate parameters when some observations are missing. See Tsionas (2001) for more references.

The rest of this article is organized as follows. In Section 2, we explain the model. In Section 3, we present Bayesian estimation procedures along with computation schemes such as the Gibbs sampler. In Section 4, numerical results are provided based on both simulated and real data.

2. The model

Consider three independent random variables Z_0, Z_1, and Z_2, where Z_k follows a Poisson distribution with mean λ_k for $k = 0, 1, 2$. Define random variables as

$$X = Z_0 + Z_1 \text{ and } Y = Z_0 + Z_2.$$

Then the random vector (X, Y) follows a bivariate Poisson distribution and it will be denoted by $(X, Y) \sim \mathrm{BP}(\lambda_0, \lambda_1, \lambda_2)$. Note that X and Y marginally follow Poisson distribution with means $\lambda_0 + \lambda_1$ and $\lambda_0 + \lambda_2$ respectively. Further, the correlation coefficient between X and Y is

$$\rho_{XY} = \frac{\lambda_0}{\sqrt{(\lambda_0 + \lambda_1)(\lambda_0 + \lambda_2)}}.$$

The joint probability mass function for (X, Y) is then

$$p(x,y) = e^{-(\lambda_0+\lambda_1+\lambda_2)} \sum_{j=0}^{\min(x,y)} \frac{\lambda_0^j \lambda_1^{x-j} \lambda_2^{y-j}}{(x-j)!(y-j)!}.$$

We refer the reader Tsionas (2001) for the extension to the multivariate Poisson distribution.

In order to proceed for the analysis, suppose we have a sample $(X_i, Y_i) \overset{\text{indep}}{\sim} \text{BP}(\lambda_0, \lambda_{1i}, \lambda_{2i})$ for $i = 1, \ldots, n$. We use the canonical log link function by letting $\lambda_{1i} = \exp\{w_i'\beta\}$ and $\lambda_{2i} = \exp\{w_i'\gamma\}$, where β and γ are $k \times 1$ vectors of unknown regression coefficients and w_i' is the ith row of the $n \times k$ design matrix with $k-1$ covariates when the intercept term is included in the model. We note that the common parameter λ_0 remains as it is without incorporating regressors.

Remark 2.1 Although we use the same covariates on the links for our analysis, one can use different covariates. However, interpretations could be more complicated with this setting.

3. Bayesian inference

3.1. Prior specifications

First, we use a gamma prior for λ_0. That is, $p(\lambda_0) \propto \lambda_0^{\delta-1} e^{-\epsilon\lambda_o}$. Second, we use a proper, but considerably vague prior with a large variance for the regression coefficients β_l and γ_l for $l = 1, \ldots, k$. In particular, these priors have forms with high spread such as a normal density with extremely large variance. Vague or weakly informative priors have been extensively used in several papers (cf. Smith and Spiegelhalter, 1980; Spiegelhalter and Smith, 1982; Akman and Raftery, 1986). So, we assume that the priors on β_l and γ_l are normally distributed with means μ_{1l} and μ_{2l}, and variances σ_{1l}^2 and σ_{2l}^2 respectively, for $l = 1, \ldots, q$. Finally, we assume $(\lambda_0, \beta_1, \ldots, \beta_k, \gamma_1, \ldots, \gamma_k)$ are independent a *priori*.

3.2. Estimation procedures

Estimation of the proposed model can be done in a Bayesian framework in conjunction with data augmentation. The data augmentation algorithm can be used to estimate the parameters with sampling based approaches like the Gibbs sampler. The Gibbs sampler is proposed by Geman and Geman (1984), and is widely used as a general method in Bayesian computation (Gelfand and Smith, 1990; Zeger and Karim, 1991).

Let $\beta_{-l} = (\beta_1, \beta_2, \cdots, \beta_{l-1}, \beta_{l+1}, \cdots, \beta_k)$ and $\gamma_{-l} = (\gamma_1, \gamma_2, \cdots, \gamma_{l-1}, \gamma_{l+1}, \cdots, \gamma_k)$. That is, β_{-l} and γ_{-l} are $(k-1) \times 1$ vectors with the lth component being excluded from β and

γ respectively. Then the full conditional distributions are given by

$$\begin{aligned} p(\beta_l|\beta_{-l},\lambda_0,\gamma,\theta,\eta,D) &\propto \left\{\prod_{i=1}^n e^{-\lambda_{1i}}\lambda_{1i}^{x_i}\left[\sum_{j=0}^{\min(x_i,y_i)}\frac{1}{(x_i-j)!(y_i-j)!}\left(\frac{\lambda_0}{\lambda_{1i}\lambda_{2i}}\right)^j\right]\right\} \\ &\cdot \exp\left\{-\frac{(\beta_l-\mu_{1l})^2}{2\sigma_{1l}^2}\right\}, \\ p(\gamma_l|\gamma_{-l},\lambda_0,\beta,\theta,\eta,D) &\propto \left\{\prod_{i=1}^n e^{-\lambda_{2i}}\lambda_{2i}^{y_i}\left[\sum_{j=0}^{\min(x_i,y_i)}\frac{1}{(x_i-j)!(y_i-j)!}\left(\frac{\lambda_0}{\lambda_{1i}\lambda_{2i}}\right)^j\right]\right\} \\ &\cdot \exp\left\{-\frac{(\gamma_l-\mu_{2l})^2}{2\sigma_{2l}^2}\right\}, \\ p(\lambda_0|\beta,\gamma,\theta,\eta,D) &\propto \left\{\prod_{i=1}^n \left[\sum_{j=0}^{\min(x_i,y_i)}\frac{1}{(x_i-j)!(y_i-j)!}\left(\frac{\lambda_0}{\lambda_{1i}\lambda_{2i}}\right)^j\right]\right\} \\ &\cdot \lambda_0^{\delta-1}e^{-(n+\epsilon)\lambda_0}. \end{aligned}$$

4. Experimental results

4.1. A simulation study

We conducted a simulation study to examine the performance of the proposed model. We use one covariate and the intercept with $n = 200$. Note that the parameters are $(\beta_1, \beta_2, \gamma_1, \gamma_2, \lambda_0)$ in this setting. First, we generate covariates $w_1, \ldots, w_n$ from a normal distribution with a mean of 5 and a variance of 1. Second, we compute $\lambda_{1i} = \exp\{\beta_1 + w_i\beta_2\}$ and $\lambda_{2i} = \exp\{\gamma_1 + w_i\gamma_2\}$ for $i = 1, \ldots, n$. Finally, we can generate dependent variables (x_i, y_i) from $\mathrm{BP}(\lambda_0, \lambda_{1i}, \lambda_{2i})$ for $i = 1, \ldots, n$. We have used the following parameter values: $\beta_1 = -1, \beta_2 = 1, \gamma_1 = 2, \gamma_2 = -2, \lambda_0 = 2$.

As mentioned in Section 3.1, We use a normal prior for β_l and γ_l centered at zero with a large variance. The hyperparameters are $\mu_{1j} = \mu_{2j} = 0$ and $\sigma_{1j} = \sigma_{2j} = 1,000$ for $j = 1, 2$. We use $(\delta, \epsilon) = (0.2, 0.1)$. Note that we use these hyperparameter values for real data analysis in the following subsection. We use the Metropolis-Hastings algorithm within the Gibbs sampler to obtain samples from the full conditional distributions. The Gibbs sampler ran for 200,000 iterations after the initial 10,000 iterates were discarded as a burn-in. We compute posterior means and 95% highest posterior density (HPD) intervals for each parameter. The numerical values are reported in Table 4.1. It seems that all estimates of the parameters based on the posterior means are close to the true values except γ_2. However, all of 95% HPD intervals include the true values. Figure 4.1 gives trace plots. Convergence was established after 50,000 iterations using Gelman and Rubin's R-statistic (Gelman and Rubin, 1992). The values of R were close to 1 for all parameters in the Gibbs sampler.

Table 4.1 Parameter estimates based on simulated data

parameter	true value	posterior mean	95% HPD
β_1	-1	-1.1042	(-1.2913, -0.8908)
β_2	1	1.0155	(0.9619, 1.0622)
γ_1	2	2.2060	(1.4015, 3.0151)
γ_2	-2	-1.6022	(-2.1390, -1.1349)
λ_0	2	2.2474	(1.9075, 2.4667)

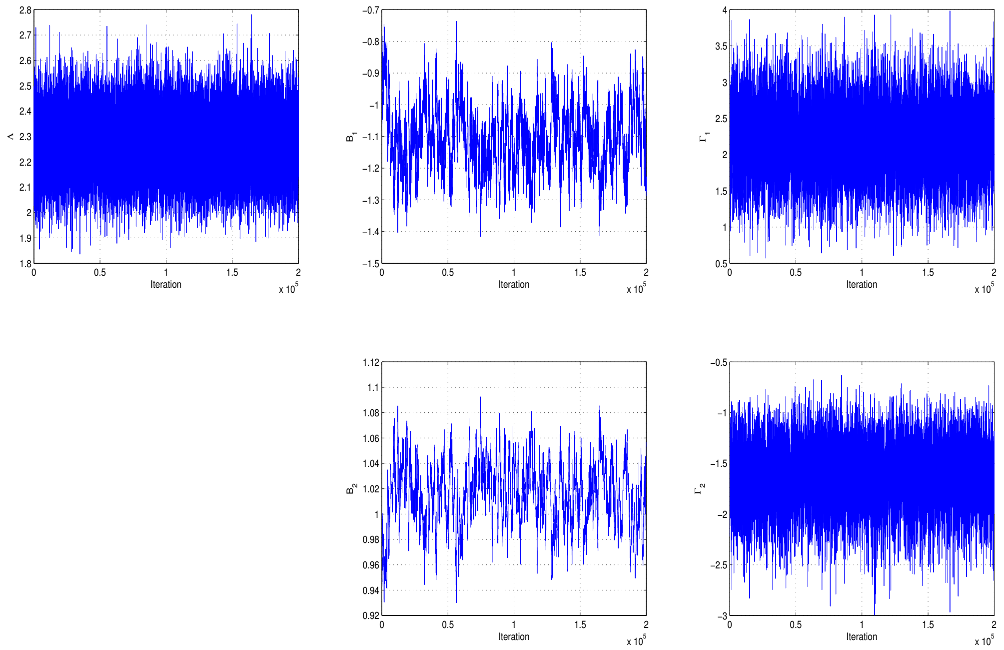

Figure 4.1 Plots of sampled values for a bivariate Possion model based on 200,000 iterations

4.2. Real data analysis

We analyze the accident data for National Highway Number 15 (NH 15) in Korea. The data were recorded in the year 2006 through the year 2011 for a six-year period. The data are gathered at straight (tangent) 137 segments from Mokpo to Geumcheon. The data set consists of five independent variables. They are the annual average daily traffic (AADT), the length of tangent, grade, the length of curve, and the radius of curve. We note that the length and radius of curve are calculated just before each straight segment. The dependent variables are the number of injuries (x) and the number of fatal accidents (y).

In this analysis, the Gibbs sampler ran for 10,000 iterations after the initial 1,000 iterates were discarded as a burn-in. We compute the posterior means and 95% HPD intervals for the parameters. They are reported in Table 4.2.

We note that the signs of the estimates are congruent with what we would expect from typical characteristics in safety analysis perspectives. First, as the values of AADT increase so does the number of accidents due to simple exposures. Second, as the length of tangent increases so does the corresponding exposure. Moreover, as the length of tangent increases, the speed of drivers could go up. This does not necessarily increase the frequency of the accidents. However, at least, it could increase the number of fatal injury. In table 4.2 we can see that the signs of the estimates associated with AADT and the length of tangent are all positive for both components. We also note that the magnitude of estimates does not imply the degree of accident frequencies, which is well known in conventional regression settings. Third, the signs for grade are negative, which is a bit natural. In data collection, we put positive signs for up-hills and negative signs for down-hills. It is usual that the speeds for down-hills get higher resulting in more (fatal) accidents. When the length of curve gets larger, the corresponding length of curve also gets larger. Drivers have tendencies to increase the speed when they are entering to next straight lines. This results in the increase of accidents. Finally, the signs associated with the length of curve are negative. These results can be interpreted as follows: as the radius of curve gets smaller, many drivers will lower speed first and raise the speed when they enter next straight lines to restore a diminished portion between speeds. This speed difference could increase the number of accidents (Babkov, 1968; Garber, 1988).

Table 4.2 Parameter estimates based on NH 15 data

component	parameter	posterior mean	95% HPD
	λ_0	0.1361	(0.0369,0.2395)
x	intercept (β_1)	-1.0683	(-1.7422, -0.4053)
	AADT (β_2)	0.5233	(0.1496, 0.8848)
	length of tangent (β_3)	0.5828	(0.4164, 0.7471)
	grade (β_4)	-0.0811	(-0.2191,0.0640)
	length of curve (β_5)	0.0246	(-0.4714,0.4783)
	radius of curve (β_6)	-1.2933	(-2.5273,-0.2150)
y	intercept (γ_1)	-1.6166	(-2.7115,-0.6007)
	AADT (γ_2)	0.4239	(-0.2979,0.9906)
	length of tangent (γ_3)	0.5699	(0.3345,0.8319)
	grade (γ_4)	-0.1402	(-0.3722,0.0897)
	length of curve (γ_5)	0.0340	(-0.7187,0.8012)
	radius of curve (γ_6)	-1.9170	(-4.0492,-0.1110)

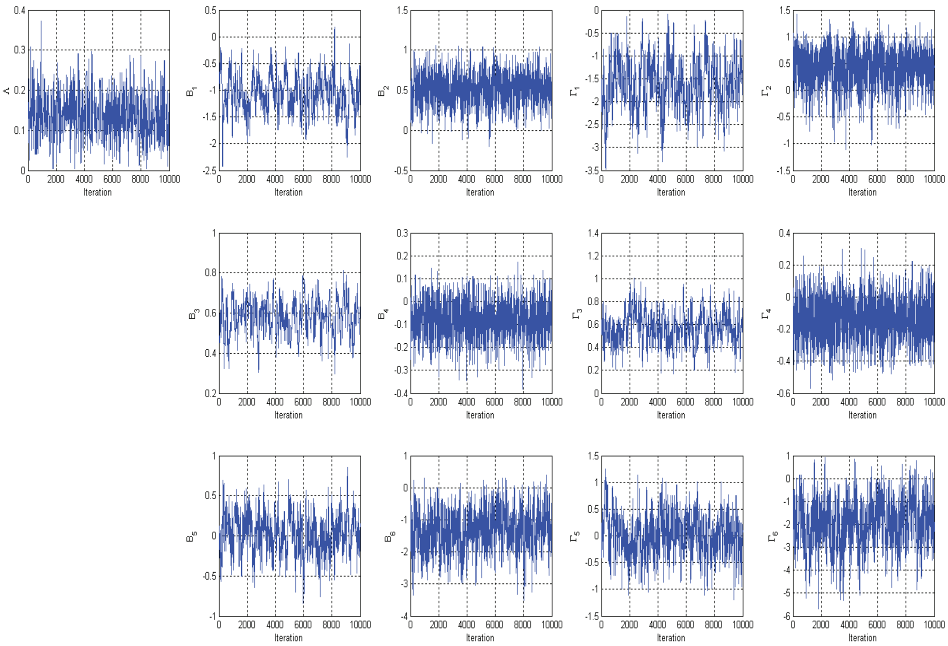

Figure 4.2 Plots of sampled values for a bivariate Possion model based on 10,000 iterations

References

Adamids, L. and Loukas, S. (1994). ML estimation in the bivariate Poisson distribution in the presence of missing values via the EM algorithm. *Journal of Statistical Computation and Simulation*, **50**, 163-172.

Akman, V. E. and Raftery, A. E. (1986). Bayes factors fornon-homogeneous Poisson processes with vague prior information. *Journal of the Royal Statistical Society Series B*, **48**, 322-329.

Babkov, V. F. (1968). Road design and traffic safety. *Traffic Engineering and Control*, 236-239.

Bijleveld, F. D. (2005). The covariance between the number of accidents and the number of victims in multivariate analysis of accident related outcomes. *Accident Analysis & Prevention*, **37**, 591-600.

Garber, N. J. and Gadirau, R. (1988). *Speed variance and its influence on accidents,* AAA Foundation for Traffic Safety, Washington, DC.

Gelfand, A. E. and Smith, A. F. M. (1990). Sampling based approaches to calculating marginal densities. *Journal of American Statistical Association*, **85**, 398-409.

Gelman, A. and Rubin, D. B. (1992). Inference from iterative simulation using multiple sequences. *Statistical Science*, **7**, 457-472.

Geman, S. and Geman, D. (1984). Stochastic relaxation, Gibbs distributions and the Bayesian restoration of images. *IEEE Transaction on Pattern Analysis and Machine Intelligence*, **6**, 721-741.

Jovanis, P. P. and Chang, H. L. (1986). Modelling the relationship of accidents to miles travelled. *Transportation Research Record*, **1068**, 42-51.

Kim, D. and Jeong, H. C. (2006). Multivariate Poisson distribution generated via reduction from independent Poisson variates. *Journal of the Korean Data & Information Science Society*, **17**, 953-961.

Kim, D., Jeong, H. C. and Jung, B. C. (2006). On the multivariate Poisson distribution with specific covariance matrix. *Journal of the Korean Data & Information Science Society*, **17**, 161-171.

Lord, D. and Persaud, B. N. (2004). Estimating the safety performance of urban road transportation networks. *Accident Analysis & Prevention*, **36**, 609-620.

Ma, J. and Kockelman, K. M. (2006). Bayesian multivariate Poisson regression for models of injury count, by severity. *Transportation Research Record*, **1950**, 24-34.

Papageorgiou, H. and Loukas, S. (1988). Conditional even point estimation for bivariate discrete distributions. *Communications in Statistics-Theory and Methods*, **17**, 3403-3412.

Persaud, B. N. (1991). Estimating accident potential of Ontario road sections. *Transportation Research Record*, **1327**, 47-53.

Persaud, B. N. (1994). Accident prediction models for rural roads. *Canadian Journal of Civil Engineering*, **21**, 547-554.

Persaud, B. N. and Dzbik, L. (1993). Accident prediction models for freeways. *Transportation Research Record*, **1401**, 55-60.

Smith, A. F. M. and Spiegelhalter, D. J. (1980). Bayes factors and choice criteria for linear models. *Journal of the Royal Statistical Society Series B*, **42**, 213-220.

Spiegelhalter, D. J. and Smith, A. F. M. (1982). Bayes factors for linear and log-linear models with vague prior information. *Journal of the Royal Statistical Society Series B*, **44**, 377-387.

Tsionas, E. G. (2001). Bayesian multivariate Poisson regression. *Communications in Statistics-Theory and Methods*, **30**, 243-255.

Tunaru, R. (2002). Hierarchical Bayesian models for multiple count data. *Austrian Journal of Statistics*, **31**, 221-229.

Zeger, S. L. and Karim, M. R. (1991). Generalized linear models with random effect: A Gibbs sampling. *Journal of the American Statistical Association*, **86**, 79-86.