Journal of Safety Research 46 (2013) 31–40

Contents lists available at SciVerse ScienceDirect

Journal of Safety Research

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

Bayesian road safety analysis: Incorporation of past evidence and effect of hyper-prior choice

Luis F. Miranda-Moreno [a,*], Shahram Heydari [a,1], Dominique Lord [b,2], Liping Fu [c,3]

[a] Department of Civil Engineering and Applied Mechanics, McGill University, Macdonald Engineering Building, 817 Sherbrooke St. W., Montreal, Quebec H3A 2 K6, Canada
[b] Zachary Department of Civil Engineering, Texas A&M University, College Station, 3136 TAMU, College Station, TX, USA
[c] Department of Civil Engineering, University of Waterloo, Waterloo, 200 University Avenue West, N2L 3G1, ON, Canada

ARTICLE INFO

Article history:
Received 9 November 2012
Received in revised form 11 March 2013
Accepted 11 March 2013
Available online 24 March 2013

Keywords:
road safety
Bayesian approach
past evidence
hyper-prior assumptions

ABSTRACT

Problem: This paper aims to address two related issues when applying hierarchical Bayesian models for road safety analysis, namely: (a) how to incorporate available information from previous studies or past experiences in the (hyper) prior distributions for model parameters and (b) what are the potential benefits of incorporating past evidence on the results of a road safety analysis when working with scarce accident data (i.e., when calibrating models with crash datasets characterized by a very low average number of accidents and a small number of sites). *Method:* A simulation framework was developed to evaluate the performance of alternative hyper-priors including informative and non-informative Gamma, Pareto, as well as Uniform distributions. Based on this simulation framework, different data scenarios (i.e., number of observations and years of data) were defined and tested using crash data collected at 3-legged rural intersections in California and crash data collected for rural 4-lane highway segments in Texas. *Results:* This study shows how the accuracy of model parameter estimates (inverse dispersion parameter) is considerably improved when incorporating past evidence, in particular when working with the small number of observations and crash data with low mean. The results also illustrates that when the sample size (more than 100 sites) and the number of years of crash data is relatively large, neither the incorporation of past experience nor the choice of the hyper-prior distribution may affect the final results of a traffic safety analysis. *Conclusions:* As a potential solution to the problem of low sample mean and small sample size, this paper suggests some practical guidance on how to incorporate past evidence into informative hyper-priors. By combining evidence from past studies and data available, the model parameter estimates can significantly be improved. The effect of prior choice seems to be less important on the hotspot identification. *Impact on Industry:* The results show the benefits of incorporating prior information when working with limited crash data in road safety studies.

1. Introduction

In traffic safety studies, Bayesian methods have been widely applied for both identifying hotspot locations and evaluating the effectiveness of countermeasures (Hauer, 1997; Heydecker & Wu, 2001; Higle & Witkowski, 1988; Lan & Persaud, 2012; Miaou & Song, 2005; Miranda-Moreno, Fu, Saccomano, & Labbe, 2005; Nathan & Gary, 2006; Park, Park, & Lomax, 2010; Persaud & Lyon, 2007; Persaud, Lyon, & Nguyen, 1999; Schluter, Deely, & Nicholson, 1997; Song, Ghosh, Miaou, & Mallick, 2006). Within the class of Bayesian methods, we can distinguish two main approaches commonly used in road safety studies: the empirical (EB) approach and the hierarchical Bayes (HB) approach (also referred as full Bayes analysis). One important difference between these two approaches is in the way the prior parameters are determined. In the EB approach, the prior parameters are estimated using a maximum likelihood technique or other techniques involving the use of the accident data, such as the method of moments (Hauer, 1997; Lord, 2006; Lord & Park, 2008; Miranda-Moreno et al., 2005). Alternatively, in a hierarchical Bayesian analysis, the parameters of the prior distributions depend in turn on additional parameters with their own priors, also referred as hyper-priors (e.g., Berger, 1985; Carlin & Louis, 2008; Rao, 2003. In traffic safety, the hierarchical Bayes approach is usually utilized when working with hierarchical Poisson models (Miaou & Song, 2005). In these types of models, the model parameters at the second level (e.g., regression or dispersion parameters in the hierarchical Poisson/Gamma model) are also supposed to follow some underlying distributions (or hyper-priors), adding another level of randomness.

Despite its modeling flexibility, the HB approach requires the specification of priors for the parameters included in the last level of the model

* Corresponding author. Tel.: +1 514 398 6589.
E-mail addresses: luis.miranda-moreno@mcgill.ca (L.F. Miranda-Moreno), shahram.heydari@mcgill.ca (S. Heydari), d-lord@tamu.edu (D. Lord), lfu@uwaterloo.ca (L. Fu).
[1] Tel.: +1 514 398 6589.
[2] Tel.: +1 979 458 3949.
[3] Tel.: +1 519 888 4567.

0022-4375/$ – see front matter
http://dx.doi.org/10.1016/j.jsr.2013.03.003

hierarchy, which are referred here as "hyper-priors" (Aul & Davis, 2006; Carlin & Louis, 2008; El-Basyouny & Sayed, 2012; Gelman, Carlin, Stern, & Rubin, 2003; Lord & Miranda-Moreno, 2008; Park et al., 2010; Rao, 2003). In practice, the so-called "vague" or "non-informative" hyper-priors are commonly used with the idea to let the data "speak for itself." In spite of this common practice (i.e., using a non-informative hyper-prior), the impact of hyper-prior specification may not be trivial when modeling accident data characterized by a low sample mean (LSM) and a small sample size (SSS) (Heydari, 2012; Lord & Miranda-Moreno, 2008; Nathan & Gary, 2006; Song et al., 2006).

Under these conditions, the use of vague hyper-priors can be problematic leading to inaccurate posterior estimates (e.g., Lambert, Sutton, Burton, Abrams, & Jones, 2005; Lord & Miranda-Moreno, 2008). In addition, the results may be sensitive to the distribution choice.

In the road safety literature, much of the attention has focused on the development of alternative model settings and safety measures, the selection of functional forms, and the estimation of accident modification factors among others (e.g., Cheng & Washington, 2005; Miaou & Lord, 2003; Miaou & Song, 2005; Miranda-Moreno, 2006; Persaud, 1986; Persaud & Lyon, 2007). Little work has so far investigated the effect of the hyper-prior choice adopted on model parameters when modeling accident data. Moreover, there is a lack of practical guidance for building informative hyper-priors using available evidence from past traffic studies and expert's opinions, as a potential solution to the LSM and SSS problem. In other fields such as epidemiology, biology and reliability engineering, both the effect of distribution choice and the incorporation of prior knowledge have attracted a lot of attention (Ashby & Smith, 2002; Guikema, 2007; Lambert et al., 2005; McMahon et al., 2006; Van Dongen, 2006). However, in road safety, very little research has been done in this respect. The work of Oh and Washington (2006) is one of the few attempts to incorporate experts' opinions (expert knowledge elicitation) for safety analyses.

According to these observations, the aim of this paper is two-fold:

- To propose a simple framework to incorporate available evidence from similar past studies to formulate informative hyper-priors for the Inverse dispersion parameter.
- To investigate the performance of alternative hyper-prior assumptions for modeling the inverse dispersion parameter in a hierarchical Bayes Poisson model under LSM and SSS.

To achieve our objectives, a simulation study is carried out to compare alternative distribution choices (e.g., Gamma versus Uniform) or hyper-prior specifications (e.g., vague versus informative priors) for the inverse dispersion parameter in a hierarchical Poisson/Gamma model (also known as the Negative Binomial model). The performance of alternative hyper-prior specifications is also evaluated according to the model capacity to detect the "true" hotspots. A comparative performance is then made in terms of parameter estimation accuracy.

2. Hierarchical poisson models for accident data

In the traffic safety literature, different model settings have been proposed for analyzing crash data, ranging from the standard Negative Binomial (NB) to more complex models such as the zero-inflated Poisson, hierarchical mixed Poisson, latent-class Poisson, Poisson mixture and Conway-Maxwell-Poisson models (e.g., Hauer, 1997; Miranda-Moreno, 2006; Lord et al., 2007; 2008b; Song et al., 2006; Park & Lord, 2009). Lord and Mannering (2010) provide a good summary of the latest models that have been introduced for analyzing crash data. Among them, the hierarchical Poisson/Gamma is perhaps the most popular. A simple version of this model is defined above.

Considering that the number of accidents at site i (Y_i) over a given time period T_i, is Poisson distributed, a hierarchical Poisson/Gamma model may be defined as Eq. (1) (Miranda-Moreno, 2006; Rao, 2003):

$$
\begin{aligned}
&\text{i.} \quad && Y_i \big| T_i, \theta_i \sim \text{Poisson}(T_i \cdot \theta_i) \\
& && \quad \sim \text{Poisson}\left(T_i \cdot \mu_i e^{\varepsilon_i}\right) \\
&\text{ii.} && e^{\varepsilon_i} \sim \text{Gamma}(\phi, \phi) \\
&\text{iii.} && \phi \sim \pi(.) \ \text{and} \ f(\boldsymbol{\beta}) \propto 1
\end{aligned}
\tag{1}
$$

where, $\mu_i = f(F_{i1}, F_{i2}, \mathbf{x_i}; \boldsymbol{\beta})$ and $\boldsymbol{\beta} = (\beta_0, \ldots, \beta_k)'$ is a vector of regression coefficients to be estimated from the data. F_{i1} and F_{i2} are entering traffic flows in intersecting directions at intersections (Miaou & Lord, 2003). In addition, $\mathbf{x_i}$ is a vector of covariates representing site-specific attributes. T_i is the period of observation at site i, which is usually assumed to be the same for all locations in the hotspot detection activity (Note that T_i may also be different). In this case, the model error, e^{ε_i}, is assumed to follow a Gamma(.) distribution with both shape and scale parameters to be equal leading to $E\left[e^{\varepsilon_i}\right] = 1.0$ and $\text{Var}\left[e^{\varepsilon_i}\right] = 1/\phi$ (Winkelmann, 2003). This error assumption is defined in order to obtain a hierarchical version of the traditional Negative Binomial model, which has been widely used in traffic safety studies. Note again that in this hierarchical model, regression ($\boldsymbol{\beta}$) and dispersion (ϕ) parameters are assumed to be random. That is, a hyper-prior distribution, denoted by $\pi(.)$ is assumed on the dispersion parameter ϕ and, $f(\boldsymbol{\beta})$ denotes the hyper-prior on the regression coefficients $\boldsymbol{\beta}$, which is commonly assumed to be flat or non-informative; e.g., $\beta_j \sim N(0,1000)$.

One should take into account that with the introduction of random variations in the mean, the hierarchical modeling framework has the potential to address over-dispersion caused by unobserved or unmeasured heterogeneity. In the traffic safety literature, more complex hierarchical Poisson models with additive random effects have been implemented to account for spatial correlation (Aguero-Valverde & Jovanis, 2010; Miaou & Song, 2005; Song et al., 2006). To account for spatial dependency, Gaussian conditional autoregressive (CAR) models are very popular. However, for illustrative purpose, this research is based on the simple hierarchical model defined in Eq. (1) and more complex models are out of the scope of our study. To obtain parameter estimates for Eq. (1), posterior inferences are carried on by Markov Chain Monte Carlo (MCMC) simulation methods such as Gibbs sampling and Metropolis–Hastings algorithm (Carlin & Louis, 2008; Gelman et al., 2003; Rao, 2003), which are already implemented in the software package WinBUGS.

Note that in the hierarchical Bayes setting defined by Eq. (1), the dispersion parameter ϕ is assumed to be random, $\phi \sim \pi(.)$. In this study, we investigate what is the impact of alternative distributions on ϕ including the following:

$$
\begin{aligned}
&\text{(i)} \quad f(\phi; a_0, b_0) = \frac{b^{a_0}}{\Gamma(a_0)} \phi^{a_0 - 1} e^{-b_0 \phi}, \phi > 0, a_0 > 0, b > 0 \ \text{(Gamma distribution)} \\
&\text{(ii)} \quad f(\phi; a_0) = \frac{a_0}{(a_0 + \phi)^2}, \phi > 0, a_0 > 0, \ (\textit{Christiansen' distribution}) \\
&\text{(iii)} \quad f(\phi; a_0) = \frac{1}{b - a_0}, a_0 < \phi < b \ \text{(Uniform distribution)}
\end{aligned}
\tag{2}
$$

where, a_0 and b are the parameters of these alternative distributions. Remark that the distribution 2-(ii) was first suggested by Christiansen and Morris (1997), where a_0 is the hyper-prior guess for the median of ϕ. For this particular distribution, small values of a_0 are less informative. A conservative choice letting the data speak by itself would be to choose a_0 small enough so that a_0 is less than the median of ϕ (Nandram, Liu, & Choi, 2005). Later in section 2.5, we see how the prior 2-(ii) can be defined on the basis of a Pareto distribution.

Note that these alternative distributions on ϕ have been used in previous research (e.g., Miaou & Song, 2005; Nathan & Gary, 2006; Song et al., 2006). However, their impact on the final outcome of a safety analysis has not been investigated. As shown in previous studies (Lambert et al., 2005; Lord & Miranda-Moreno, 2008), hyper-prior