Accident Analysis and Prevention 56 (2013) 51–58

Contents lists available at SciVerse ScienceDirect

Accident Analysis and Prevention

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

Investigating different approaches to develop informative priors in hierarchical Bayesian safety performance functions

Rongjie Yu[a,b,*], Mohamed Abdel-Aty[a]

[a] *Department of Civil, Environmental and Construction Engineering, University of Central Florida, Orlando, FL 32816-2450, United States*
[b] *School of Transportation Engineering, Tongji University, 4800 Cao'an Road, 201804 Shanghai, China*

ARTICLE INFO

Article history:
Received 9 October 2012
Received in revised form 25 February 2013
Accepted 21 March 2013

Keywords:
Informative prior
Bayesian inference
Hierarchical Bayesian model
Safety performance functions
Weather and traffic data

ABSTRACT

The Bayesian inference method has been frequently adopted to develop safety performance functions. One advantage of the Bayesian inference is that prior information for the independent variables can be included in the inference procedures. However, there are few studies that discussed how to formulate informative priors for the independent variables and evaluated the effects of incorporating informative priors in developing safety performance functions. This paper addresses this deficiency by introducing four approaches of developing informative priors for the independent variables based on historical data and expert experience. Merits of these informative priors have been tested along with two types of Bayesian hierarchical models (Poisson-gamma and Poisson-lognormal models). Deviance information criterion (DIC), *R*-square values, and coefficients of variance for the estimations were utilized as evaluation measures to select the best model(s). Comparison across the models indicated that the Poisson-gamma model is superior with a better model fit and it is much more robust with the informative priors. Moreover, the two-stage Bayesian updating informative priors provided the best goodness-of-fit and coefficient estimation accuracies. Furthermore, informative priors for the inverse dispersion parameter have also been introduced and tested. Different types of informative priors' effects on the model estimations and goodness-of-fit have been compared and concluded. Finally, based on the results, recommendations for future research topics and study applications have been made.

1. Introduction

Considerable research have been conducted to develop safety performance functions (SPFs) using the Bayesian inference technique. In order to gain better understanding of the hazardous factors and achieve better model fit, various types of data have been employed in crash frequency studies. In addition to the basic geometric characteristics and annual average daily traffic (AADT) data, extended weather and traffic related data have also been utilized in the analyses. In order to deal with the multi-type data, hierarchical Bayesian (HB) models were introduced to the traffic safety analyses. The introduction of more data resources would yield clearer understanding of the crash occurrence mechanisms, however, these kinds of studies commonly suffered from small sample sizes and low sample means problems due to data limitations. Previous studies have concluded that this kind of data would affect the accuracy of model estimation and potential SPF applications (Lord, 2006; Lord and Miranda-Moreno, 2008).

One key advantage of the Bayesian inference method compared to the conventional frequentist inference approach is that extra knowledge and experience about the data can be used as prior information in the analyses. The Bayesian framework provides a complete and coherent way to balance the empirical data and prior expectations, which would be very promising to be applied in traffic safety analyses. However, most current crash frequency studies that have utilized Bayesian inference methods employed non-informative priors (sometimes referred to as vague priors) to let the data "speak for itself," which ignored the merit of the Bayesian inference approach.

Recently, researchers started to incorporate prior information from previous studies and experts' judgment (Washington and Oh, 2006; Lord and Miranda-Moreno, 2008); nonetheless, none of them focused on how to develop prior distributions for the independent variables. This study fills the gap in formulating informative priors for the independent variables in traffic safety studies. Four different ways to formulate the informative priors have been introduced and their effects on hierarchical Bayesian models were examined. Inference results of the models with informative priors were compared to models with non-informative priors. One year of crash

* Corresponding author at: Department of Civil, Environmental and Construction Engineering, University of Central Florida, Orlando, FL 32816-2450, United States. Tel.: +1 407 823 0300.
E-mail addresses: rongjie.yu@gmail.com, rongjie.yu@knights.ucf.edu (R. Yu).

0001-4575/$ – see front matter © 2013 Elsevier Ltd. All rights reserved.
http://dx.doi.org/10.1016/j.aap.2013.03.023

data for a 15-mile freeway section on I-70 in Colorado was used in this study along with the roadway geometric characteristics, weather information, and real-time traffic data. The most commonly used Poisson-gamma and Poisson-lognormal models have been employed to examine the effects of informative priors. Moreover, in addition to the informative priors for the independent variables, effects of informative priors for the inverse dispersion parameter have also been examined. Based on the results, the most effective informative priors and the more robust HB model were suggested for future applications.

2. Background

The Bayesian inference method has been commonly utilized in traffic safety analyses, however, only few past studies discussed how to formulate informative priors and their superiority. Washington and Oh (2006) incorporated experts' judgment on different countermeasures and used the Bayesian methodology to assess the prior information's effects on accident modification factors (AMF). The authors surveyed 11 transportation professionals independently to rate the effectiveness of chosen countermeasures. Then the results of this survey were transformed into prior distributions for the AMFs with Beta distributions. Within the Bayesian framework, these prior distributions were combined with results from the literature regarding these countermeasures. It was concluded that after the utilization of prior information, the ranking of the countermeasures did not change too much while the most significant impact is the increased certainty in the estimations.

Lord and Miranda-Moreno (2008) conducted a study to verify the effects of small sample sizes and low sample means on the estimation of dispersion parameters. Informative and non-informative priors for the inverse dispersion parameter have been tested under different datasets with a variety of sample means and sample sizes. Results of the study demonstrated that small sample sizes and low sample means would seriously affect the estimation of posterior means for the inverse dispersion parameter. Finally, an informative prior was suggested to be adopted to reduce the likelihood of mis-estimation. Schlüter et al. (1997) utilized a HB model to rank hazardous traffic accident locations. Based on the practitioners' prior knowledge about the mean accident rates and their bounds for a group of sites, informative priors have been developed. The authors indicated that hierarchical models provide a natural mechanism whereby prior information can be elicited and incorporated in the analysis. Jang et al. (2010) analyzed count data containing a large amount of zero observations by zero-inflated Poisson and zero-inflated negative binomial regression models. A power prior was introduced and calculated by the likelihood function with the historical data, raised to power α_0. Results indicated that the zero-inflated models with the power prior performed better than the conventional inference approach. Haleem et al. (2010) introduced a reliability process to reduce the prediction uncertainty on crash occurrence. Two-stage updated prior information was used with log-gamma and negative binomial likelihood functions. After comparing the results of models with informative and non-informative priors, the authors concluded that the full Bayesian updating framework using log-gamma likelihood function has the most promising results.

Outside the traffic safety area, informative priors have been widely adopted in political studies, reliability estimations, economic research, and psychological experimental analyses. Jaynes (1985) discussed the role of prior information in statistical inference and analyzed a seasonal adjustment economic issue. Results showed that prior information for the seasonal parameters has a large effect on the estimations; prior information made a noticeable improvement in seasonal adjustment. The author concluded that with the help of prior knowledge, estimations of some irregular functions can be much more accurate when compared with estimations from the sampling theory results. Western and Jackman (1994) documented a Bayesian approach to analyze the political comparative research. The data for the comparative research suffered from repeated and collinearly data, which were not suitable for use in the conventional inference technique. A Bayesian regression approach had been used to analyze the data and priors were assumed by the researchers. Bayesian regression solved the multicollinearity problem by using extra prior information other than the ordinary least squares, and it generated smaller standard errors for the regression coefficients. Guikema (2007) stated that priors played an important role in employing Bayesian methods in risk analyses. Informative priors formulated from existing information can lead to more accurate posterior inferences. Five different methods to formulate the informative priors were introduced: method of moments; maximum likelihood estimation; maximum entropy estimation; starting from a non-informative 'pre-prior'; and fitting a prior based on confidence/credible interval matching. Vanpaemel (2011) introduced a hierarchical way to construct informative priors and the method had been tested by an example dataset. Results of the models indicated that psychological intuition about the relative plausibility of the models can be formally captured by an informative prior distribution. The informative priors kept a fair balance between avoiding the random fluctuations in the data caused by sampling variability and provide theoretically based expectations without losing flexibility.

Conclusions from different study areas as mentioned above indicate that incorporating the informative priors into the Bayesian analysis would result in a balanced posterior distribution from both the information of the experts and the data itself (Bedrick et al., 1996). Also the informative priors could solve the problems caused by data limitations.

3. Data preparation

The datasets that were utilized in this study were constructed from four original datasets: (1) one year of crash data provided by Colorado Department of Transportation (CDOT), (2) roadway geometric characteristics data captured from the Roadway Characteristics Inventory (RCI), (3) real-time weather data recorded by 6 weather stations along the study's roadway section and (4) real-time traffic data detected by 30 remote traffic microwave sensor (RTMS) radars.

A total of 251 crashes were documented and included in the analysis. Due to lack of historical information and knowledge about the model results, the original dataset has been split into two datasets. One training dataset is created to play the role of historical data while the other test dataset is intended for implementing and evaluating the effects of the different informative priors. With the consideration of the significant differences between the two seasons' crash occurrence mechanisms (Yu et al., 2013); a binary seasonal index has been created based on the crash season (dry season from May to September and snow season from October to April). Random selection was performed for the original crash data. In order to consider the seasonal bias, a stratified partitioning based on the binary seasonal index was performed. The training and test datasets ended up with 118 crashes and 122 crashes, respectively. The stratified partitioning approach resulted in two comparable datasets while keeping the original seasonal biased data's characteristics. From Table 1, it can be seen that the summary statistics for the two datasets are very close for all the variables.

Table 1
Summary of variables descriptive statistics for the training data and test data.

Variables	Description	Training dataset		Test dataset	
		Mean	Std dev.	Mean	Std dev.
Crash frequency	Crash frequency counts for the segment	0.54	1.1	0.55	1.1
Length	Segment length (mile)	0.24	0.16	0.24	0.16
VMT	Daily vehicle miles traveled	6582	4419	6582	4419
Av_temp	Average temperature during the crashes (°F)	38.8	15.4	39.1	14.6
Av_visibility	Average visibility during the crashes (mile)	4.0	1.6	4.1	1.6
Cov (precipitation)	Coefficient of variance for the 1 h precipitation (rain/snow) before the crash	5.3	3.8	5.1	3.4
AS	Average speed for the crash segment (mph)	55.8	8.4	55.2	9.5
Grade	Longitudinal grade, eight categories: upgrade: 0–2% = 1, 2–4% = 2, 4–6% = 3, 6–8% = 4; downgrade: 0–(−2)% = 5, (−2)–(−4)% = 6, (−4)–(−6)% = 7, (−6)–(−8)% = 8	4.7	2.4	4.7	2.4

The 15-mile freeway section, starting from Mile Marker (MM) 205 and ending at MM 220, has been divided into 120 homogenous segments (60 segments in each direction): with the major segmentation criterion of roadway alignment homogeneity. According to the Roadway Characteristics Inventory (RCI) data, both horizontal and vertical alignments were scrutinized; a minimum-length of 0.1 mile was used to avoid the low exposure problem and the large statistical uncertainty of the crash rate per short segment (Ahmed et al., 2011). The two datasets have been aggregated into segment based crash frequency datasets to be used in the analyses. Summary statistics regarding the data can be found in Table 1. The raw traffic data was first aggregated into 5-min intervals, then each crash was assigned to the nearest downstream radar detector, and the crash's traffic status is defined as the 5–10 min time period that precedes the crash time. For example, if a crash occurred at 15:25, at the Mile Marker 211.3. The corresponding traffic status for this crash is the traffic condition of time interval 15:15 and 15:20 recorded by RTMS radar at Mile Marker 211.8. To avoid errors in the reported crash-times, the average speed (AS) within the 5–10 min time interval that proceeds the reported crash time was extracted for each crash. For the weather variables, crashes were first assigned to the nearest weather station according to the MM. Then, the closest weather record prior to the crash reported time has been extracted and used as the crash occurrence weather conditions. Traffic and weather variables were aggregated based on the segment level: if more than one crash happened at the specific segment, mean values are calculated and used in the analysis; for zero crash segments, average values from the original data (one year raw data) for these variables are used. Moreover, the coefficient of variance for the 1 h precipitation has also been calculated based on the segment level.

4. Methodology

Crash occurrence has been assumed to follow a Poisson process, and the Poisson models have been frequently utilized in crash frequency studies. However, due to the lack of ability to handle the over-dispersion problem, Poisson mixture models such as Poisson-gamma and Poisson-lognormal models were introduced to compensate for the shortcoming of Poisson regression models (Lord and Mannering, 2010).

The Poisson-lognormal model was formulated by introducing multiplicative gamma distributed random effects into the log-linear Poisson model, which implies a negative binomial marginal sampling distribution. The hierarchical model can be setup as follows:

$$
\begin{aligned}
&Y_{it} \sim Poission(\lambda_{it}) \quad \text{for } t = 1, 2\\
&\log \lambda_{it} = \log e_{it} + \boldsymbol{X_{it}\beta} + \gamma_1 b_i\\
&b_t \sim N(0, \sigma_b^2)
\end{aligned}
$$

The Poisson-gamma model is a commonly used model for count data with over-dispersion problem. The model can be setup as:

$$
\begin{aligned}
&Y_{it} \sim Poission(\lambda_{it}\mu_{it}) \quad \text{for } t = 1, 2\\
&\mu_{it} \sim Gamma(r_{it}, r_{it})\\
&\log \lambda_{it} = \boldsymbol{X_{it}\beta}\\
&r_{it} \sim Gamma(\varphi, \varphi)
\end{aligned}
$$

where Y_{it} is the crash count at segment i (i = 1,. . ., 120 (60 segments in each direction)) during season t (t = 1 for dry season, 2 for snow season). $\boldsymbol{X_{it}}$ represent the risk factors and $\boldsymbol{\beta}$ is the vector of regression parameters. For the Poisson-lognormal model, b_i is assumed to follow a normal distribution with the variance parameter that has been specified a gamma prior as Gamma (0.001, 0.001). For the Poisson-gamma model, μ_{it} is a multiplicative random effect which is usually being assumed to follow a gamma distribution with mean of 1 and variance of $1/\varphi$; where φ is regarded as the inverse dispersion parameter and usually set up to follow a gamma prior as Gamma (0.001, 0.001).

According to Ntzoufras (2009), Bayesian models have an inherently hierarchical structure. The prior distribution $f(\boldsymbol{\beta}|\,\boldsymbol{a})$ of the model parameters $\boldsymbol{\beta}$ with prior parameters $\boldsymbol{a}$ can be considered as the first level of hierarchy. The likelihood of the Bayesian models has the posterior distribution $f(\boldsymbol{\beta}|\,\boldsymbol{y}) \propto f(\,\boldsymbol{y}|\boldsymbol{\beta})f(\boldsymbol{\beta}\,;\,\boldsymbol{a})$ via the Bayes theorem. Moreover, the Bayesian hierarchical model is defined when a prior distribution is also assigned on the prior parameters $\boldsymbol{a}$ associated with the likelihood parameters $\boldsymbol{\beta}$. The posterior distribution can be written as

$$
\begin{aligned}
&f(\boldsymbol{\beta}|\boldsymbol{y}) \propto f(\boldsymbol{y}|\boldsymbol{\beta})f(\boldsymbol{\beta};\boldsymbol{a})f(\boldsymbol{a};\boldsymbol{b})\\
&\propto f(\boldsymbol{y}|\boldsymbol{\beta})f(\boldsymbol{\beta}|\boldsymbol{a})f(\boldsymbol{a}|\boldsymbol{b})
\end{aligned}
$$

The Bayesian inference method involves prior information with the data sample to formulate posterior probability statements for the parameters of a model (Western and Jackman, 1994). The inherent structure of Bayesian inference distinguishes itself from the convention reference from two aspects. First, Bayesian inference is built on a subjective probability concept. Second, the permissibility of the introduction of the prior information enhances the sample information in making the statistical inference. With the rare information of how to formulate the prior distributions in traffic safety analysis, this paper generally investigates four possible approaches (two-stage Bayesian updating, maximum likelihood estimation, method of moments and expert experience) to generate the informative priors for the explanatory variables. Normal distribution is adopted as informative prior distributions since the normal prior distributions are straightforward to develop.

4.1. Two-stage Bayesian updating

This is a promising way to formulate informative priors based on historical data. First, historical data are treated separately to perform the Bayesian inference with uniform non-informative priors for all the parameters. After the estimation, a set of posterior distributions for the independent variables can be achieved. The means and variances of the posterior distributions for the independent variables from the inference would perform as informative priors incorporated in the later inference procedure. By this approach, new dataset is treated as an extension of the historical data, since the posterior distributions for the first stage Bayesian inference have been updated with the pure information from the sample data with non-informative priors. The Bayesian inference for the new dataset with informative priors would perform the so-called "two-stage Bayesian updating" and provides more precise results. This approach would make efficient use of historical data if similar model results are available in previous studies and it would work even if the data is highly skewed (Guikema, 2007).

4.2. Maximum likelihood estimation

For the conventional inference, maximum likelihood estimation (MLE) is developed with fitting a probability density function (PDF) to the data. Estimation results from the MLE approach can be used as informative priors in the latter Bayesian analysis. In order to avoid multicollinearity influences on coefficient estimations, previously chosen significant variables would be entered into the model one at a time. Results of the means and variances for the estimated coefficients will be recorded and used as prior information in the Bayesian analysis. This approach does not need historical dataset; existing dataset would first be estimated by multi-MLE inference with one independent variable each time and then information about the coefficients would be recorded and utilized later.

4.3. Method of moments

The method of moments' way to formulate informative priors is the simplest approach. The basic idea is to match the moments of the data to the moments of the informative prior distributions in the Bayesian analysis. Results of the summary statistics would provide the mean and variance values for each independent variable. The informative priors will be assigned as normal distributions with information from the summary statistics. This approach is easy to implement since no extra models and data are needed.

4.4. Expert experience

This approach is also referred to as inclusion of general information since it does not come from any models or studies; it is purely based on expert experience on the relationship between explanatory variables and crash occurrence. Information is achieved from surveyed expert opinions about the independent variables; for example: AADT would have positive effects on crash occurrence. This approach is not data based and sometimes may not have the ability to provide sufficient information to avoid the implausible inferences due to subjectivity.

5. Modeling results and discussion

5.1. Informative priors for the explanatory variables

To test the effects of informative priors developed from different approaches, Poisson-lognormal and Poisson-gamma models were developed using WinBUGS (Lunn et al., 2000). Three chains of 15,000 iterations were set up in the software and the first 5000 iterations were discarded and regarded as burn-in. Convergence of the models were checked by conducting the BGR (Gelman and Rubin, 1992) statistics and Geweke diagnostic through the R package *boa* (Smith, 2007).

Informative priors were formulated based on the four abovementioned methods. For the two-stage Bayesian updating approach, the prepared training dataset was used as historical data to perform the first updating step. Posterior distributions of the independent variables were used as informative priors for the test dataset. The MLE was performed by using SAS (SAS Institute Inc., 2004) with a negative binomial model performed on the test dataset; each selected variable was evaluated individually. Means and variances of the independent variables were calculated and transformed into normally distributed prior distributions. The method of moments approach is comparatively easy, whereas informative priors were elicited from the summary statistics (Table 1). Furthermore, regarding the expert experience method, informative priors for variables such as visibility, temperature, and precipitation were formulated based on common meteorology knowledge; while for the average speed variable, prior information was provided by the expert experience as mean traveling speed would likely be around the speed limits with a large variance. Detailed information for the informative and non-informative priors used in the Poisson-gamma model can be found in Table 2. Priors from the two-stage Bayesian updating and maximum likelihood approaches reflect the relationship between the independent variables and the response variable (for example, temperature has a positive influence on crash occurrence likelihood in the dry seasons); while the method of moments priors represent only the information about the independent variables themselves (for example, mean average temperature during the dry season). Expert experience priors are a combination of those two types of information (priors for LogVMT and Cov(precipitation) reflect their relationship with the crash occurrence while the other variables' priors are based on general knowledge only).

Modeling results for the Poisson-lognormal and Poisson-gamma models are presented in Tables 3 and 4, respectively. Generally, all the models provided similar estimated coefficients and consistent signs for the independent variables. The Log VMT variable shows a positive effect on crash frequency, which means large VMT would increase crash occurrence likelihood due to the larger exposure. The geometric characteristic parameter (Grade index) has a consistent result in the models. Although some of the grade groups are not significant, the trend of the coefficients indicates that the steeper slopes experienced a higher crash frequency while the upgrade are relatively safer than the downgrade segments. Besides, traffic speed shows a consistent influence on crash occurrence for both seasons. The results demonstrate that the crash occurrence likelihood increases as the average speed 5–10 min prior to the crash occurrence decreases; which demonstrates that crashes are more likely to occur during congestion and this result has also been found in previous real-time crash risk evaluation models (Ahmed et al., 2013).

For the dry season, the temperature variable has a positive coefficient demonstrating that extreme temperatures may lead to a less safe state as identified in Malyshkina et al. (2009). While for the snow season, average visibility has negative effect on crash occurrence which means that good visibility conditions would decrease the crash occurrence likelihood. Moreover, the coefficient of variance for the 1 h precipitation amount is significant with a negative sign. This indicates that crashes are more likely to happen at segments suffering from sudden rain or snow than those suffering continuous precipitation. Drivers are driving more carefully through those consistently high precipitation areas, which may be due to warning signs implemented along these frequent precipitation segments, or drivers adjust and are better prepared and cautious.

From model selection point of view, Poisson-lognormal and Poisson-gamma models are compared based on DIC values. DIC,

Table 2
Examples of prior distributions for the Poisson-gamma model.

Season	Variables	Prior distributions				
		Non-informative	Two-stage updating	Maximum likelihood	Moments method	Expert experience
Dry	LogVMT	(0.0, 1.0E−4)[a]	(0.9, 10)	(0.77, 2.6)	(8.66, 2)	(0.8,50)
	Av_temp		(0.06, 100)	(0.12, 50)	(47.77, 3.5E−3)	(52.6,0.08)
	AS		(0.06, 100)	(−0.12, 37)	(47.35, 5.8E−3)	(55,7.0E−3)
Snow	LogVMT		(0.5, 20)	(0.85, 5)	(8.66, 2)	(0.4,50)
	Av_visibility		(−0.3, 12.5)	(−0.37, 10)	(1.84,0.256)	(2.2,0.5)
	Cov		(−0.3, 25)	(−0.22, 25)	(5.1,0 .29)	(4.0,0.01)
	AS		(−0.01, 100)	(−0.068, 72)	(47.45,5.8E−3)	(60,0.014)

[a] Prior distributions are normal distribution with (mean, 1/variance).
** Prior distributions for the Grades are not shown in the table.

Table 3
Parameter estimates and model fitness of the Poisson-lognormal models.

Season	Variable	Non-informative	Two-stage updating	Maximum likelihood	Method of moments	Expert experience
Dry	LogVMT	1.3 (0.31)[a]	1.1 (0.25)	1.1 (0.26)	0.99 (0.21)	1.1 (0.23)
	Av_temp	0.092 (0.031)	0.055 (0.028)	0.046 (0.028)	0.054 (0.022)	0.056 (0.026)
	AS	−0.056 (0.021)	−0.061 (0.017)	−0.062 (0.019)	−0.056 (0.016)	−0.06 (0.018)
Snow	LogVMT	0.36 (0.14)	0.24 (0.12)	0.46 (0.074)	0.43 (0.076)	0.21 (0.11)
	Av_visibility	−0.3 (0.072)	−0.31 (0.071)	−0.29 (0.067)	−0.29 (0.071)	−0.32 (0.074)
	Cov	−0.27 (0.051)	−0.29 (0.05)	−0.27 (0.049)	−0.28 (0.049)	−0.29 (0.051)
	AS	−0.22 (0.0092)	−0.02 (0.0093)	−0.022 (0.0085)	−0.02 (0.0091)	−0.019 (0.0093)
Grade	Grade 1	−0.76 (0.53)	−1.3 (0.61)	−0.88 (0.41)	−0.9 (0.4)	−1.4 (0.61)
	Grade 2	−0.53 (0.31)	−0.52 (0.26)	−0.55 (0.26)	−0.6 (0.3)	−0.52 (0.26)
	Grade 3	−0.69 (0.26)	−0.81 (0.24)	−0.8 (0.23)	−0.8 (0.2)	−0.81 (0.24)
	Grade 4	−0.14 (0.34)	−0.36 (0.3)	−0.33 (0.3)	−0.3 (0.3)	−0.38 (0.3)
	Grade5	−0.95 (0.51)	−1 (0.38)	−0.94 (0.38)	−0.9 (0.4)	−1 (0.39)
	Grade 6	−0.66 (0.4)	−0.83 (0.34)	−0.77 (0.34)	−0.8 (0.3)	−0.84 (0.39)
	Grade 7	−0.2 (0.3)	−0.4 (0.28)	−0.41 (0.26)	−0.4 (0.3)	−0.41 (0.25)
Evaluation measures	*R*-square[b]	0.58 (0.053)	0.58 (0.061)	0.59 (0.052)	0.59 (0.049)	0.6 (0.058)
	DIC	344.448	346.329	344.318	343.29	346.016

[a] Standard errors for the coefficients are shown in parenthesis.
[b] Generalized *R*-square = 1 – (residual deviance/null deviance): the residual deviance is equivalent to the residual sum of squares, and null deviance is equivalent to the total sum of squares (Zuur et al., 2007).

recognized as Bayesian generalization of AIC (Akaike information criterion), is a widely used evaluation measure for the Bayesian models. As stated in El-Basyouny and Sayed (2010), according to Spiegelhalter et al. (2003), differences of more than 10 might definitely rule out the model with higher DIC. Differences between 5 and 10 are considered substantial. With the identical non-informative priors, Poisson-gamma model has the DIC of 337.281 which is slightly better compared to the Poisson-lognormal model with a DIC value of 344.448. Based on the abovementioned model selection standards, with a 7 difference in DIC between the two models, it is safe to say that the Poisson-gamma is relatively better to represent the crash occurrence distribution.

Furthermore, Poisson-lognormal models' goodness-of-fits do not seem to be affected by the informative priors, while the model fits of the Poisson-gamma models are significantly improved with the different kinds of informative priors. Therefore, more detailed comparisons have been conducted for the two types of models with different informative priors.

In addition, *R*-squared values were provided as an additional measurement beside the DICs to select superior models. Within the

Table 4
Parameter estimates and model fitness of the Poisson-gamma models.

Season	Variable	Non-informative	Two-stage updating	Maximum likelihood	Method of moments	Expert experience
Dry	LogVMT	0.58 (0.078)	0.88 (0.095)	0.92 (0.24)	0.87 (0.11)	0.57 (0.067)
	Av_temp	0.024 (0.017)	0.07 (0.014)	0.068 (0.018)	0.063 (0.014)	0.064 (0.013)
	AS	−0.06 (0.011)	−0.05 (0.01)	−0.05 (0.018)	−0.05 (0.015)	−0.047 (0.011)
Snow	LogVMT	0.57 (0.082)	0.46 (0.057)	0.4 (0.068)	0.48 (0.07)	0.34 (0.042)
	Av_visibility	−0.32 (0.073)	−0.31 (0.077)	−0.31 (0.086)	−0.3 (0.075)	−0.3 (0.085)
	Cov	−0.26 (0.045)	−0.27 (0.047)	−0.27 (0.047)	−0.24 (0.052)	−0.25 (0.046)
	AS	−0.02 (0.0078)	−0.024 (0.0062)	−0.023 (0.0066)	−0.025 (0.0093)	−0.023 (0.0057)
Grade	Grade 1	−0.94 (0.63)	−0.91 (0.47)	−0.82 (0.45)	−0.95 (0.45)	−0.68 (0.35)
	Grade 2	−0.56 (0.33)	−0.53 (0.28)	−0.49 (0.29)	−0.49 (0.26)	−0.14 (0.2)
	Grade 3	−0.76 (0.32)	−0.78 (0.23)	−0.75 (0.26)	−0.75 (0.24)	−0.46 (0.21)
	Grade 4	−0.0084 (0.35)	−0.29 (0.32)	−0.31 (0.32)	−0.28 (0.28)	0.054 (0.27)
	Grade5	−0.85 (0.54)	−0.87 (0.37)	−0.85 (0.4)	−0.97 (0.38)	−0.83 (0.39)
	Grade 6	−0.59 (0.37)	−0.67 (0.36)	−0.79 (0.38)	−0.71 (0.347)	−0.37 (0.3)
	Grade 7	−0.24 (0.057)	−0.35 (0.28)	−0.34 (0.28)	−0.31 (0.27)	0.052 (0.22)
Evaluation measures	*R*-square	0.61 (0.057)	0.66 (0.057)	0.63 (0.055)	0.64 (0.065)	0.62 (0.066)
	DIC	337.281	312.937	323.693	322.357	324.215

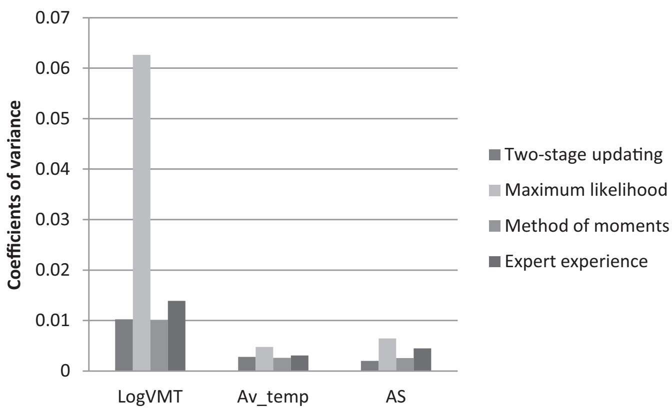

Fig. 1. Coefficients of variance in the Poisson-gamma model (snow season).

Poisson-lognormal models, the model with expert experience priors has the best *R*-squared value while the model that incorporated the method of moments' informative priors has the most promising DIC value. The two evaluation indexes are consistent through all the models.

The Poisson-gamma models are more robust with the informative priors: the two-stage Bayesian updating informative prior model reaches a low DIC value of 312.937, which is 25 lower than the non-informative prior model. All the models with informative priors have substantially lower DIC values than the base non-informative prior model. Similar results can be found by looking at R-squared values whereas the two-stage updating model has the highest value.

More attention has been paid to the coefficients of variance for the posterior estimations. These values represent the variability of the estimated coefficients; whereas large values indicate lesser confidence for the estimated effects (larger credible interval) and smaller numbers demonstrate greater confidence (smaller credible interval). Figs. 1 and 2 have been provided to compare the coefficients of variance for those significant variables in the snow and dry seasons, respectively. From the figures, it can be seen that the two-stage updating prior model provides the smallest coefficient of variance; which indicate that the two-stage Bayesian updating model provides the most precise estimations. Moreover, the method of moments' informative priors model has slightly larger coefficients of variance than the two-stage Bayesian updating approach model. Furthermore, the maximum likelihood prior model provided large coefficients of variance for the independent variables, especially for the snow season's parameters. Similar investigation has also been done for the Poisson-lognormal models and consistent results were concluded.

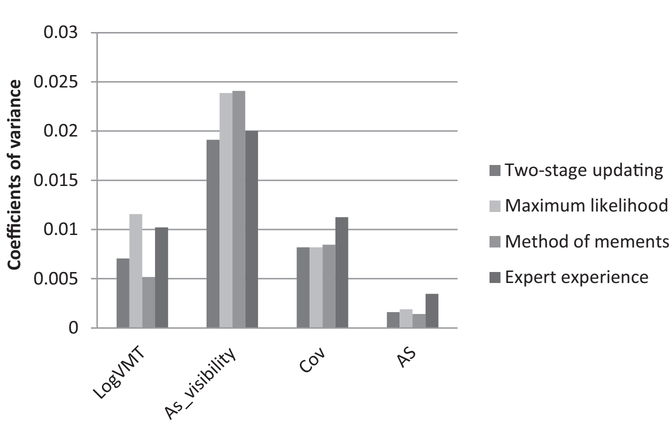

Fig. 2. Coefficients of variance in the Poisson-gamma model (dry season).

In addition, one point that needs to be mentioned is that the two-stage updating informative prior provided by historical data (Table 2) for the AS (average speed) variable was not significant in the training dataset. However, the AS variable was assigned a negative sign after the Bayesian updating, and it is consistent with all the other inferences' results. The misestimated prior information did not lead to implausible inference results and the posterior estimations were more accurate compared to the informative priors' (smaller estimation variance was achieved). This interesting fact would need further investigation in future research.

With informative priors' superiority for the Poisson-gamma models, the informative priors resulted in a more accurate and confident inference. It can be concluded that the proposed four ways of formulating informative priors can certainly improve the model fit and contribute to more precise estimated coefficients.

5.2. *Informative priors for the inverse dispersion parameter*

The section above discussed and evaluated the effects of informative priors for the independent variables. For the Poisson-gamma model, informative priors can also be formulated and utilized for the inverse dispersion parameter in addition to the independent variables. This section introduces informative prior for the inverse dispersion parameter and compares the effects of different priors on model estimations and goodness-of-fit.

As stated in the methodology section, the inverse dispersion parameter φ in the Poisson-gamma model follows a gamma distribution:

$$\varphi \sim Gamma(a, b)$$

where a and b are the shape and scale parameters, respectively (Lord and Miranda-Moreno, 2008). With non-informative priors, a and b are set equal to 0.001 which provide φ a mean of 1 and large variance of 1000. Prior information about the inverse dispersion parameter can only be achieved from the two-stage Bayesian updating approach with the historical data. By fitting the training dataset with the Poisson-gamma model, the inverse dispersion parameter has a mean and variance equal to 1.64 and 0.317 respectively. Then the informative prior for the inverse dispersion parameter was $\varphi \sim Gamma(8.48, 5.17)$.

Table 5 compares the estimation results and model goodness-of-fits of the Poisson-gamma models with informative priors for (1) the inverse dispersion parameter, (2) the explanatory variables, (3) both the inverse dispersion parameter and explanatory variables, and also (4) the non-informative model.

From the table, it can be seen that Poisson-gamma model with informative prior for the inverse dispersion parameter has substantially lower DIC than the non-informative prior model does. Investigation has also been done for the coefficients of variance for the independent variables; instead of improving the estimation accuracies, the informative prior for the inverse dispersion parameter provided larger coefficients of variance for the independent variables as compared to the non-informative prior model. Moreover, informative priors for the independent variables showed having more effect on improving model fit and estimation accuracy than the inverse dispersion parameter prior information. Furthermore, combination usage of informative priors for both the independent variables and inverse dispersion parameter had also been conducted: less accuracy posterior estimations were obtained compared to the model with only one type of informative priors; model goodness-of-fit is comparable to the model with only prior information for the inverse dispersion parameter. The use of informative priors may cause the model structure to lose its flexibility of fitting the data.

Table 5
Model comparisons for the Poisson-gamma models with different informative priors.

Season	Variable	Non-informative	Informative prior for the inverse dispersion parameter	Informative prior for the independent variables	Informative prior for both independent and dispersion parameters
Dry	LogVMT	0.58 (0.078)	0.95 (0.151)	0.88 (0.095)	1.01 (0.19)
	Av_temp	0.024 (0.017)	0.065 (0.016)	0.07 (0.014)	0.053 (0.022)
	AS	−0.06 (0.011)	−0.058 (0.011)	−0.05 (0.01)	−0.06 (0.016)
Snow	LogVMT	0.57 (0.082)	0.32 (0.06)	0.46 (0.057)	0.43 (0.07)
	Av_visibility	−0.32 (0.073)	−0.33 (0.084)	−0.31 (0.077)	−0.29 (0.07)
	Cov	−0.26 (0.045)	−0.27 (0.063)	−0.27 (0.047)	−0.28 (0.05)
	AS	−0.02 (0.0078)	−0.019 (0.0078)	−0.024 (0.0062)	−0.02 (0.008)
Grade	Grade 1	−0.94 (0.63)	−0.98 (0.6)	−0.91 (0.47)	−0.87 (0.41)
	Grade 2	−0.56 (0.33)	−0.55 (0.31)	−0.53 (0.28)	−0.51 (0.26)
	Grade 3	−0.76 (0.32)	−0.67 (0.26)	−0.78 (0.23)	−0.78 (0.24)
	Grade 4	−0.0084 (0.35)	−0.16 (0.36)	−0.29 (0.32)	−0.33 (0.30)
	Grade5	−0.85 (0.54)	−0.96 (0.52)	−0.87 (0.37)	−0.95 (0.38)
	Grade 6	−0.59 (0.37)	−0.68 (0.37)	−0.67 (0.36)	−0.77 (0.34)
	Grade 7	−0.24 (0.057)	−0.24 (0.30)	−0.35 (0.28)	−0.39 (0.27)
Evaluation measures	*R*-square	0.61 (0.057)	0.62 (0.051)	0.66 (0.057)	0.62 (0.077)
	DIC	337.281	329.595	312.937	333.176

6. Conclusion

In past studies, Lord and Miranda-Moreno (2008) only investigated the priors for the inverse dispersion parameter and Jang et al. (2010) simply looked at the prior information for the power parameter. To the best of our knowledge, this paper is the first to introduce a variety of methods to formulate prior distributions for the independent variables in traffic safety studies. Informative priors are inherent and key part of the Bayesian inference theory. With the use of informative priors, researchers can achieve better model fit and have more confident inference results. More importantly, the informative prior can help to avoid implausible and inaccurate conclusions by utilizing extra information beyond the data sample. This is especially important to those studies with small sample size and low sample mean datasets.

Due to the ability to solve the over-dispersion issue, hierarchal Bayesian models became popular in recent crash frequency studies. In this study, the Poisson-lognormal model and Poisson-gamma model provided competitive performance with the non-informative priors. However, comparisons of the model goodness-of-fit measures demonstrated that the Poisson-gamma is more robust than the Poisson-lognormal model to incorporate informative priors. The same result was also being identified in an economic study where the purchase frequency was modeled (Jen et al., 2003).

The proposed four approaches of formulating informative priors showed strong evidence of their ability to enhance model fit and inference accuracies for the specific data used in this study. From the application easiness perspective, the two-stage updating and MLE methods require extra calculation and the two-stage updating even needs historical studies to provide informative priors. In the meanwhile, the method of moments priors need only simple statistical analyses which are likely to be conducted anyway during the preliminary analyses procedures. For the expert experience priors, surveys from the experts with experience in the safety analysis field are needed (Washington and Oh, 2006).

With the consideration of models' performance results, it can be concluded that the two-stage updating approach is the best way to develop informative priors. If past data or study results are available, this method should be the first choice to develop the informative priors. Nevertheless, if there are limited historical data and results, the method of moments approach can be used since this method performed comparably to the MLE method and provided smaller credible intervals. In addition, for new data with a small sample (usually faced in the AMF studies) the expert experience method can be utilized, but with care. Moreover, it is worthwhile to note that these conclusions were drawn based on the specific dataset used in this study; further investigation with other datasets would be needed to confirm the results.

In addition to the informative priors for the independent variables, informative priors for the inverse dispersion parameter in the Poisson-gamma model have also been examined in this study. Investigation regarding different types of informative priors indicated that: (1) both types of informative priors improved the model goodness-of-fit compared to the non-informative priors; (2) informative priors for the inverse dispersion parameter resulted in larger credible intervals for the estimated coefficients; (3) informative priors for the independent variables were more effective than the informative priors for the inverse dispersion parameter in improving the goodness-of-fit and estimation accuracies; (4) overuse of the informative priors would limit the flexibility of the model since combining usage of the two prior types did not further improve the model.

Acknowledgement

The authors wish to thank the Colorado Department of Transportation for funding this work and providing the data.

References

Ahmed, M., Huang, H., Abdel-Aty, M., Guevara, B., 2011. Exploring a Bayesian hierarchical approach for developing safety performance functions for a mountainous freeway. Accident Analysis and Prevention 43, 1581–1589.
Ahmed, M., Abdel-Aty, M., Yu, R., 2013. Assessment of the interaction between crash occurrence, mountainous freeway geometry, real-time weather and AVI traffic data. Journal of Transportation Research Board 2280, 51–59.
Bedrick, E., Christensen, R., Johnson, W., 1996. A new perspective on priors for generalized linear models. Journal of the American Statistical Association 91, 1450–1460.
El-Basyouny, K., Sayed, T., 2010. Safety performance functions with measurement errors in traffic volume. Safety Science 48, 1339–1344.
Gelman, A., Rubin, D., 1992. Inference from iterative simulation using multiple sequences. Statistical Science 7, 457–511.
Guikema, S., 2007. Formulating informative, data-based priors for failure probability estimation in reliability analysis. Reliability Engineering & System Safety 92, 490–502.
Haleem, K., Abdel-Aty, M., Mackie, K., 2010. Using a reliability process to reduce uncertainty in predicting crashes at unsignalized intersections. Accident Analysis & Prevention 42, 654–666.
Jaynes, E., 1985. Highly Informative Priors. Bayesian Statistics 2. Elsevier Science Publisher329–360.
Jang, H., Lee, S., Kim, S., 2010. Bayesian analysis for zero-inflated regression models with the power prior: applications to road safety countermeasures. Accident Analysis & Prevention 42, 540–547.

Jen, L., Chou, C., Allenby, G., 2003. A Bayesian approach to modeling purchase frequency. Marketing Letters 14, 5–20.
Lord, D., 2006. Modeling motor vehicle crashes using poisson-gamma models: examining the effects of low sample mean values and small sample size on the estimation of the fixed dispersion parameter. Accident Analysis & Prevention 38, 751–766.
Lord, D., Miranda-Moreno, L., 2008. Effects of low sample mean values and small sample size on the estimation of the fixed dispersion parameter of Poisson-gamma models for modeling motor vehicle crashes: a Bayesian perspective. Safety Science 46, 751–770.
Lord, D., Mannering, F., 2010. The statistical analysis of crash-frequency data: a review and assessment of methodological alternatives. Transportation Research Part A 44, 291–305.
Lunn, D., Thomas, A., Best, N., Spiegelhalter, D., 2000. WinBUGS—a Bayesian modelling framework: concepts, structure, and extensibility. Statistics and Computing 10, 325–337.
Malyshkina, N., Mannering, F., Tarko, A., 2009. Markov switching negative binomial models: an application to vehicle accident frequencies. Accident Analysis and Prevention 41, 217–226.
Ntzoufras, I., 2009. Bayesian Modeling Using WinBUGS. Wiley, New Jersey.
SAS Institute, 2004. SAS/graph (r) 9.1 reference.
Schlüter, P., Deely, J., Nicholson, A., 1997. Ranking and selecting motor vehicle accident sites by using a hierarchical bayesian model. Journal of the Royal Statistical Society: Series D (The Statistician) 46 (3), 293–316.
Smith, B., 2007. boa: an R package for MCMC output convergence assessment and posterior inference. Journal of Statistical Software 21, 1–37.
Spiegelhalter, D., Thomas, A., Best, N., Lunn, D., 2003. WinBUGS User Manual. MRC Biostatistics Unit, Cambridge.
Vanpaemel, W., 2011. Constructing informative model priors using hierarchical methods. Journal of Mathematical Psychology 55, 106–117.
Washington, S., Oh, J., 2006. Bayesian methodology incorporating expert judgment for ranking countermeasure effectiveness under uncertainty: example applied to at grade railroad crossings in Korea. Accident Analysis and Prevention 38, 234–247.
Western, B., Jackman, S., 1994. Bayesian inference for comparative research. American Political Science Review, 412–423.
Yu, R., Abdel-Aty, M., Ahmed, M., 2013. Bayesian random effect models incorporating realtime weather and traffic data to investigate mountainous freeway hazardous factors. Accident Analysis and Prevention 50, 371–376.
Zuur, A., Ieno, E., Smith, G., 2007. Analyzing Ecological Data. Statistics for Biology and Health. Springer, New York.